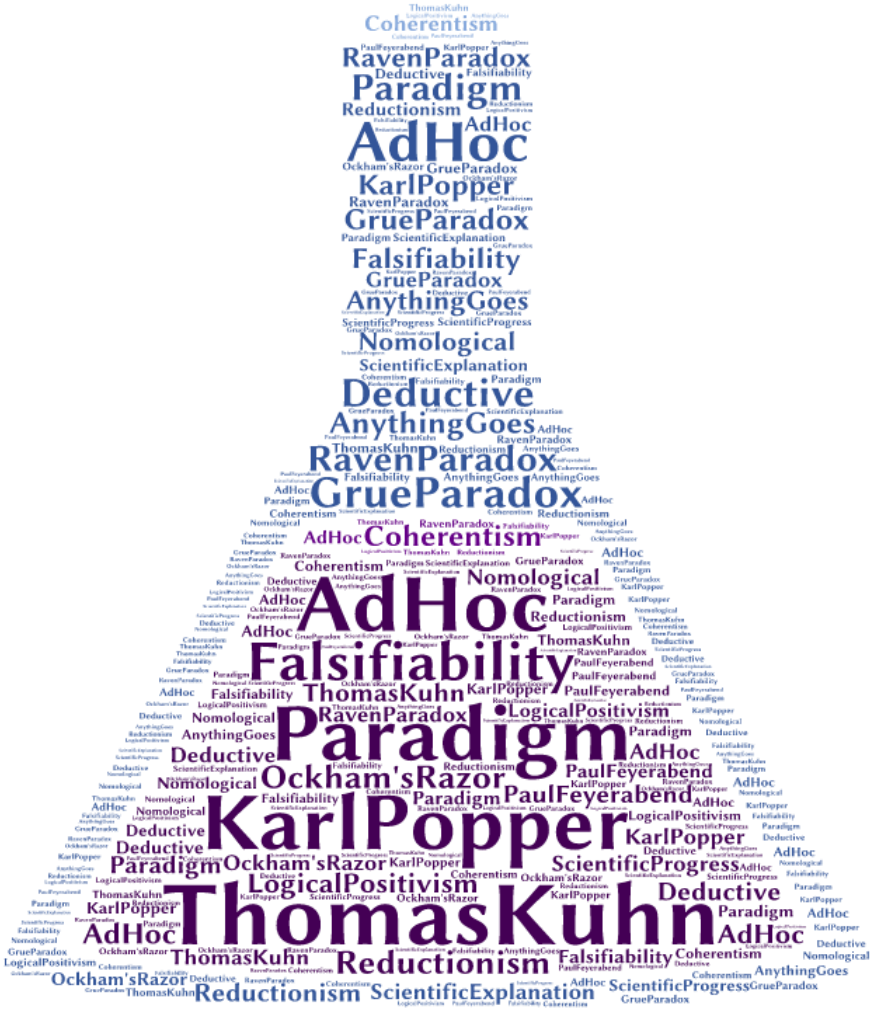
「第七屆亞太平洋科學哲學會議」會議手冊

The Booklet for the 7th Asian-Pacific Conference on Philosophy of Science



日期：2017 年（民國 106 年）12 月 15-16 日

Date: 15th-16th, December, 2017.

地點：台灣嘉義國立中正大學文學院 144 國際會議廳、143 研討室

Venue: R144 International Conference Hall and R143 Conference Room, College of Humanities, National Chung-Cheng University, Chiayi, Taiwan.

主辦單位：國立中正大學哲學系、亞太平洋科學哲學系列會議委員會

Hosted and Organized by the Department of Philosophy, National Chung-Cheng University, and the Committee of Asia-Pacific Series of Conference on Philosophy of Science.

協辦單位：中華民國科技部、國立中正大學國際處

Supported by the Ministry of Science and Technology (R.O.C), and the Office of International Affairs, National Chung-Cheng University.

目錄 Content

* This original file is PDF. And we intend to keep it as it is, in order to prevent from that the formatting is off.

** This original file is PPT.

Preface to the Special Section: Philosophy of Science in East Asia

(*Annals of the Japan Association for Philosophy of Science*, Volume 26 (2017) pp. 9-12)

Tetsuji Iseda (Graduate School of Letters, Kyoto University)

This special issue is a result of recent cooperation among philosophers of science in Asian countries, especially Japan, Korea and Taiwan. As a participant of this activity, let me first summarize the recent interactions among East Asian philosophers of science. To my knowledge, this international cooperation was fostered by Korean philosophers of science. Korean Society for Philosophy of Science (KSPS) is a relatively young organization established in 1995, but (or probably because of that) it has been very active in fostering international relationship among neighboring countries.[1]

Its fifth president, In-Rae Cho gave a presentation at the annual meeting of PSSJ (Philosophy of Science Society, Japan) in 2006. He invited back professors Nobuharu Tanji (the president of PSSJ at that time) in 2007. This exchange was developed into regular invitations from Korea; professor Fu Dawie of National Tsing-Hua University (2008) from Taiwan and professor Soshichi Uchii (2009) and myself (2010) from Japan were successively invited to annual meetings of KSPS.

The next step in the international interaction was joint conferences. A preliminary bilateral conference between Japanese and Korean philosophers of science was held in Kyoto in February 2011, organized by myself and sponsored by Department of Philosophy and History of Science, Kyoto University. As is mentioned I was invited to give a talk at the KSPS annual meeting in 2010, and found that there are many philosophers of science in Korea who share similar interests as Japanese colleagues. That is why I decided to invite five philosophers each from Korea and Japan to begin collective interactions.

The idea of multinational workshop on philosophy of science was taken up professor

[1] KSPS website
http://philsci.or.kr/eng/index.asp

Sang Wook Yi, who was the host of my talk at KSPS and an invited speaker of the Kyoto workshop. He decided to make it a conference series and also decided to expand it to include Taiwanese colleagues. The result was the first East Asian Philosophy of Science Workshop (EAPSW), held as a satellite event to KSPS annual meeting in July 2011. Professor Szu-Ting Chen, one of the contributors of current issue, was the participant of the first EAPSW and has been an active participant and organizer of the subsequent conferences. By the way, there is no organization specific to philosophy of science in Taiwan, but there are several internationally and interdisciplinarily active philosophers there, which makes Taiwan a kind of focal point in international interactions.

Professor Yi made it sure that this workshop is the official event of KSPS, and professor Young E. Rhee, the current president of KSPS and one of the contributors of current issue, took over the management of Korean side of the conference series.

At the time of writing this introduction, six conferences are held in this series, and one is scheduled. Professor Rueylin Chen and his former Ph.D. student Hsiao-Fan Yeh, authors of the third contribution to this special issue, took part in this conference series from the third event in 2013, and will going to host the 7th event in December 2017. The list of conferences is the following:[2]

1 The First East Asia Philosophy of Science Workshop
Hanyang University, Seoul, Republic of Korea, July 2, 2011
2 The Second East Asia Philosophy of Science Workshop
Miyazaki Station KITEN Building, Convention Room, Miyazaki-shi, Miyazaki, Japan, November 12, 2012 (a satellite event for the annual meeting of Philosophy of Science Society, Japan)
3 The Third East Asia Conference on the Philosophy of Science
National Tsing-Hua University, Hsinchu, Taiwan, October 3-4, 2013
4 The Fourth East Asia & Southeast Asia Conference on the Philosophy of Science 2014

[2] More detailed information on the first five conferences is available at East-Asian and Pacific Conference on Philosophy of Science (EAPCPS) website maintained by KSPS
http://philsci.or.kr/eng/html/sub04_01.asp
For more on the sixth event, see the following booklet.
http://www.cape.bun.kyoto-u.ac.jp/wp-content/uploads/2016/08/APCPS2016booklet.pdf

Institut Latihan Islam Malaysia (ILIM), Bangi, Selangor, Malaysia, November 5-6, 2014

Main theme “Indigenization of Knowledge and Intercivilizational Dialogue”

5 The Fifth East-Asian and Pacific Conference on Philosophy of Science

Seoul National University, Seoul, Republic of Korea, August 25-26, 2015

Main theme “The Philosophy of Science and the Science-Technology Civilization in the 21st Century”

6 The 6th Asia-Pacific Conference on Philosophy of Science

Kyoto University, Kyoto, Japan, September 10 - 11, 2016

7 The 7th Asian-Pacific Conference on Philosophy of Science

National Chung Cheng University, Chayi, Taiwan, December 2017 (scheduled)

The changing name of the conference series signifies the expansion of the circle of philosophers of science. Philosophers from Malaysia were invited to the third conference, and became the host of the fourth of the conference. In the sixth conference, philosophers of science from the People's Republic of China were invited for the first time in the series.

This conference series is not alone in terms of fostering international interactions among philosophers in East Asia. There is another conference series, called Conference on Contemporary Philosophy in East Asia (CCPEA), running alongside. This conference series has a wider scope including other fields of contemporary philosophy and especially so-called 'analytic Asian philosophy', i.e. attempts at integrating insights from Asian traditional philosophies into analytic philosophy. CCPEA conferences have been held in Taipei (2012), Kyoto (2014) and Seoul (2016).[3] Philosophers of science who got to know one another also take part in CCPEA conferences, which means that we have chance to meet in person more than once a year. This seems to be a quite close interaction, especially given that such interaction was virtually non-existent ten years ago.

What is the purpose of such international exchanges among Asian philosophers of

[3] The last two events of this series has their own website.
http://www.cape.bun.kyoto-u.ac.jp/ccpea2014/
http://www.ccpea2016.kr/

science? Why do we care about what colleagues in neighboring countries do? Naturally the answer is not simple, partly because participants have different motives in the first place. My own motives for organizing the prototype workshop between Japan and Korea were roughly the following. Philosophers of science in Japan and in Korea are in similar situations. They are relatively small in number; we are far away from the centers of research in this field. Maybe we can find someone with similar background and interest in the other country, i.e. potential research collaborators. I think my own prototype workshop and the subsequent conferences strongly confirmed my expectations, not just between Japan and Korea, but also among East Asian countries in general.

One thing I keep feeling sorry about this series of conferences is that the exchanges are entirely in English. This is a shame given that we share a lot of cultural backgrounds, especially Chinese characters (kanji). I hope that the conference series becomes the starting point for deeper interactions including learning one another's language. In the mean time, there is a bright side about this English environment. Because English is nobody's first language, younger researchers can practice English without feeling much pressure.

The three contributions to this special issue, one from Korea and two from Taiwan, will give you some idea as to what issues other philosophers in East Asia are interested in, and how they approach those issues. This is of course just a small sample of research in philosophy of science in this region. I hope the readers are stimulated to look for potential collaborators by themselves.

Campus Map of National Chung-Cheng University

*The no. 1 route is recommended, unless you drive by yourself.

Rules of Presentation

1. Every speaker will be allocated 40 minutes for presentation and discussion. In principle, a speaker can use 25 minutes to present and leave 15 minutes for Q&A. Presentations do not exceed 30 minutes.
2. Timekeepers will remind you of the used time by lifting a signboard when your presentation will have been going for 20, 25, and 30 minutes.
3. Please moderators control the timing.

The 7th Asia-Pacific Conference on Philosophy of Science

Date: December 15, 2017

Time	Event / Venue	
8:30-9:00	Registration	
	R144	R143
9:00-9:10	**Opening Remarks** Ruey-Lin Chen	N/A
Moderator	Jaemin Jung	Alan Hajek
9:10-10:30	**From Philosophy of Science to Philosophy of Inquiries** Tetsuji Iseda	**Modes of Convergence to the Truth** Han-ti Lin
	Case Study Method Revisited: Overgeneralization or a Straw in the Wind Wei Wang	**How and When Chances Guide Credences via the Principal Principle** Ilho Park
10:30-10:50	Refreshment Break	
Moderator	Kunihisa Morita	Linton Wang
10:50-12:10	**The Engagement of Kalam in Modern Science : A physicist view** Shahidan Radiman	**Self-knowledge and Objectivity in Participant Observation** Zhu Xu
	Is the standard model of cosmology built upon conventionalist stratagems? Man Ho Chan	**New data on the linguistic diversity of authorship in philosophy journals** Chun-Ping Yen and Tzu-Wei Hung
12:10-13:30	Lunch	
Moderator	Kai-Yuang Cheng	Kei Yoshida
13:30-14:50	**Science from perspective of Al-Qur'an** Mohd Yusof Hj Othman	**Methodological Individualism and Reductionism** Francesco Di Iorio
	Malay Values in Scientific Inquiry Abdul Latif Samian	**The Non-Identity problem and the social choice procedure based on**

		asymmetrical relationships Reiko Gotoh
14:50-15:10	Refreshment Break	
Moderator	Jiwon Shim	Iseda Tetsuji
15:10-16:30	**Jabir Ibn Hayyan: The Islamic Philosophy of The Father of Chemistry** Ibrahim N. Hassan, Mohd Yusof Hj Othman and Abdul Latif Samian	**Between Scylla and Charybdis: Investigating a Possibility of the Social Sciences** Kei Yoshida
	Chemical Decomposition and Analogical Reasoning in Humphry Davy's Electrochemistry Jonathon Hricko & Yafeng Shan	**Knowledge and Big Data** Paul Dumouchel
16:30-16:40	Break	
Moderator	Wei Wang	Young E. Rhee
16:40-18:00	**What the unsupervised learning could deliver us (or, what not)** Insok Ko	**Staying Regular?** Alan Hájek
	A taxonomy of experiments or modes of interventions? Hsiao-Fan Yeh and Ruey-Lin Chen	**Inference to the Hidden Factors** Linton Wang, Ming-Yuan Hsiao and Jhih-Hao Jhang

18:30-20:30 Welcome Dinner (invited)

Date: December 16, 2017

Time	Event / Venue	
8:30-9:00	Registration	
	R144	R143
Moderator	Zhu Xu	Min OuYang
9:00-10:20	**Big data, logic of scientific discovery, and abduction** Young E. Rhee	**Are there laws of evolution?** Jun Otsuka
	Two Visual Systems and Phenomenology of Visual Consciousness Feng Yu	**Individuating genes as types or individuals** Ruey-Lin Chen
10:20-10:40	Refreshment Break	
10:40-12:00	Keynote: **An Epistemology of Scientific Practice** C. Kenneth Waters (Chair: Ruey-Lin Chen)	N/A
12:00-13:00	Lunch	
13:00-13:30	**Communication Session (R143)** Kei Yosida (ANPOSS) C. Kenneth Waters (The Summer Institute) Ruey-Lin Chen (ISHPSSB)	
Moderator	Mohd Yusof Hj Othman	Jun Otsuka
13:30-14:50	**Cognitive Decision Theory and Permissive Rationality** Jaemin Jung	**Backtracking analysis and causal ascription of singular historicals** Richard W. T. Hou
	Problems of Intrinsic Time Direction within A-Theories Kunihisa Morita	**Practice-based Paradigms in Biological Sciences: Large-Scale Quantitative and Qualitative Analyses of a Case Study on Heart-Rate Variability**

		Karen Yan, Meng-Li Tsai and Tsung-Ren Huang
14:50-15:20	Refreshment Break	
Moderator	Insok Ko	Karen Yan
15:20-16:40	**On The Notion of Interaction in The Nature-nurture Debate** Qiao-Ying Lu	**Holobionts from an Immunological Perspective: the Problem of Pan-Homeostasism** Lynn Chiu
	Debate of Permitted Limit of a Prosthetic Limb when Boarding Jiwon Shim	**How to Characterize The Individuality of Holobionts?** Shi-Jian Yang
16:40-17:00	**Closing Remarks & Ceremony** (R144) (Group Photo)	
17:00-17:40	N/A	**APCPOS Committee Meeting** (R143)

18:00-20:00 Farewell Dinner (invited)

15th R144 Session I

From Philosophy of Science to Philosophy of Inquiries

Tetsuji Iseda
Kyoto University

Philosophy of science tends to become more narrowly focused nowadays, refraining from making a general theory that cover all branches of science. Even though philosophy of science provides many insights potentially useful for inquirers of various fields, the self-restraint of philosophy of science has been made such insights harder to see for them.

I use the concept 'inquiry' as a general notion that include wide variety of information gathering activities. In an abstract model, an inquiry is a three-term relationship: the target of inquiry, the inquiring subject, intermediaries. There are many different types of inquiries, some of which may be regarded as non-science depending on what definition of science we adopt, while being indispensable as our access to the information we need.

There are three major reasons why it is desirable to extend philosophy of science to include all those inquiries as its subject matter to give an unified account of inquiries. First is that the methodology of those inquiries is much less well-defined than ordinary science, and philosophy of science can find some guiding principles applicable to those other fields. Second, seemingly remote inquiry domains may actually have common features which make an inter-domain comparison fruitful. Third, there are many interdisciplinary studies nowadays in which practitioners of different types of inquiry should work together. To understand what is going on (epistemically, not socially) in such interdisciplinary studies, it is better to have a conceptual framework that can cover all inquiries that participate in the study.

Case Study Method Revisited: Overgeneralization or a Straw in the Wind

Wei Wang
Institute of Science, Technology, and Society, Tsinghua University
wangwei@tsinghua.edu.cn

The case study method is very common in the social researches. However, it meets several challenges in recent decades, especially generalization from a single case or a few cases to a social theory may commit the inductive fallacy of overgeneralization and general applicability of case study fails frequently. So some social scientists even claim that the only generalization is there is no generalization. The paper reviews the central debates on case study in social sciences. Using an analogy with laboratory tests in natural sciences, the author suggests that a well-constructed single case can be regarded as an experimental prototype, from which scientists build, confirm or disconfirm theories, but transferability of case study may often fail due to too complex situation and conditions.

Keywords: **case study method, generalizability, transferability**

15th R143 Session I

Modes of Convergence to the Truth

Hanti Lin
UC Davis

Those who engage in normative or evaluative studies of induction (such as formal epistemologists, statisticians, and theoretical computer scientists) have provided many positive results for justifying (to a certain extent) various kinds of inductive inferences. But they have said little about a very familiar kind: I call it full enumerative induction, which concerns inference to the *full* conclusion that *all* ravens are black (rather than the restricted conclusion that all the ravens observed in the future will be black). To remedy this, I develop a learning-theoretic solution in a way that can even be embraced by some Bayesians. The idea is to (i) define and study various modes of convergence to the truth, construed as epistemic ideals for an inquirer to achieve where possible, (ii) look for the the modes of convergence that can be achieved when tackling the problem of whether all ravens are black, (iii) of those modes, identify the one that corresponds to the highest achievable epistemic ideal, and (iv) see whether full enumerative induction can be justified as (that is, proved to be) a necessary means for achieving that epistemic ideal. The answer is positive, according to the main theorems of this paper. The Bayesian versions of those results are proved as well. The results are also extended for justification of a kind of Ockham's razor. The key to all these results is to introduce a mode of convergence slightly weaker than Gold's (1965) and Putnam's (1965) identification in the limit; I call it almost everywhere convergence to the truth, where the conception of "almost everywhere" is borrowed from geometry and topology. (This talk will not presuppose knowledge of topology.)

When Chances Guide Credences:
Another Non-commutativity of Bayesian Updating

Ilho Park
Chonbuk National University

This paper is intended to examine the relationship between the Principal Principle and Conditionalization. In particular, I will show that our credences are sensitive to when objective chances guide our rational credences. This result can be regarded as another kind of non-commutativity of Bayesian belief updating. For this purpose, I will first formulate several versions of the Principal Principle and Conditionalization. And then, I will provide two ways of feeding objective chances to what I will call 'chance-free credence functions': The evidence-first belief updating and the chance-first belief updating. With these two ways at hand, it will be proved that these two kinds of belief updating lead us to different posterior credence functions. Lastly, I will consider some possible responses to my result, and show that such responses are of no help to the non-commutativity in question.

15th R144 Session II

The Engagement of Kalam in Modern Science: a Physicist View

Shahidan Radiman
School of Applied Physics, Faculty of Science and Technology
Universiti Kebangsaan Malaysia
E-mail: shahidan@ukm.edu.my

Kalam or Islamic Rational Science was established during the Umayyad (8th C AD) and reached its height during the Abbasid Caliphate period (10th C AD) in an attempt to understand aspects of the Islamic faith by logical reasoning in main due to the influence of Greek rational thought on Muslim philosophers especially in Basra and Baghdad, both in Iraq . Two main branch of Kalam were developed namely Jalil al-Kalam (Divine attributes and actions) and Daqiq al-Kalam (rational science). Two main school of Kalam emerged namely the Muktazilites and the Asharites. Despite the different views by the Kalam philosophers (Mutakallimun) they all subscribed to some common basic principles in understanding nature. Here we discussed 5 of these principles which influenced the philosophy of modern science, especially reductionism. The dichotomy of jawhar (substance) and a'radh (accident) can be seen to perpetuate into the dichotomy of ontic and epistemic modalities found in modern quantum theory. Discreteness of natural structures including space and time which is one of the main Kalam principle played important role in quantum gravity theories either via causal sets or pregeometry and braneworld models. Whereas the Mutakallimun were once divided about the existence of the Universe into ex niliho vs pre-existing, current cosmologists are equally divided into this category as well with new understanding on Multiverse scenario. Some aspects of modern physics view had already been proposed by one of the most famous Mutakallimun namely Fakhr al-Din al Razi, some of which will be discussed in this paper.

Is the Standard Model of Cosmology Built upon Conventionalist Stratagems?

Man Ho Chan
The Education University of Hong Kong

Recently, some studies point out that our current standard model of cosmology is built upon a set of conventionalist stratagems. Some believe that the theories of dark matter and dark energy are ad hoc hypotheses to account for the observational anomalies. As Karl Popper argues, it is crucial to avoid conventionalist stratagems if falsifiability of a theory has to be preserved. In this presentation, from the perspective of history and philosophy of science, I will show that the current model of cosmology, the Lambda-Cold-Dark-Matter model, is not built upon any conventionalist stratagem. The concepts of dark matter and dark energy have good theoretical ground in physics so that they should not be regarded as ad hoc hypotheses.

15th R143 Session II

Self-knowledge and Objectivity in Participant Observation

Zhu Xu
Department of Philosophy, East China Normal University
zxu@philo.ecnu.edu.cn

Participant observation is mainly flourished in anthropology in 20th century. It requires the researcher to take part in activities in order to grasp the agent's experience and point of view on native life, as well as to maintain intellectual distance for the purpose of critical reflections upon what she participates. As a research technique, participant observation firstly connects with Bronislaw Malinowski, who carried out studies on native life in the Trobriand Islands, and published *Argonauts of the Western Pacific* in 1922.

Julie Zahle (2012; 2013; 2016) recently defenses the objectivity of participant observation in philosophy of social sciences. She argues that participant observation could be a reliable method for knowing native ways of life, especially knowing tacitly how to be an appropriate agent in practices.

Though "observations of her own actions" made by social scientists has been identified as one of the four types of observation, Zahle seems not to attribute particular significance upon that issue. Nevertheless, it is the researchers' self-knowledge that configures the objectivity of participant observation as a much more sophisticated issue, than most scientific observations. In order to defense an objective status, it is not enough merely to avoid "the observer's distortion of the situation", claimed by Zahle, but also to resolve particular concerns involved in self-knowledge. And I will argue for a layer-model upon non-observational and observational self-knowledge, which is supposed to guarantee the objectivity of participant observation.

New Data on the Linguistic Diversity of Authorship in Philosophy Journals

Chun-Ping Yen* and Tzu-Wei Hung**

This paper investigates the representation of authors with different linguistic backgrounds in academic publishing. We first review some common rebuttals of concerns about linguistic injustice. We then analyze 1,039 authors of philosophy journals, primarily selected from the 2015 Leiter Report. While our data show that Anglophones dominate the output of philosophy papers, this unequal distribution cannot be solely attributed to language capacities. We also discover that ethics journals have more Anglophone authors than logic journals and that most authors (73.40%) are affiliated with English-speaking universities, suggesting other factors (e.g. philosophical areas and academic resources) may also play significant roles. Moreover, some interesting results are revealed when we combine the factor of sex with place of affiliation and linguistic background. It indicates that while certain linguistic injustice is inevitable in academic publishing, it may be more complex than thought. We next introduce Broadbent's (2009a, 2009b, 2012, 2013, 2014) contrastive account of causation to give a causal explanation of our findings. Broadbent's account not only well characterizes the multifaceted causality in academic publishing but also provides a methodological guideline for further investigation.

Keywords: **Linguistic injustice; Lingua franca; Philosophy journals; Linguistic privilege, Causality**

* Graduate Institute of Philosophy, National Tsing Hua University; chunping.yen@gmail.com.
** Institutie of European and American Studies, Academia Sinica; htw@gate.sinica.edu.tw.

Science from Perspective of Al-Qur'an

Mohd. Yusof Hj. Othman
Institute of Islam Hadhari, Universiti Kebangsaan Malaysia
e-mail: myho@ukm.edu.my

According to Toby Huff (1995), from about eighth century till the end of the thirteenth century, the Arabic-Islamic world had the most advanced science to be found anywhere in the world. They established excellent centre of learning called House of Wisdom in Baghdad (786), the world's first university called al-Qarawiyin University in Fez, Morocco in 859, followed by University of al-Azhar in Cairo (972). They produced renowned scientist such as Jabir Ibn Hayyan (722) (the father of chemistry), al-Biruni (973) (in astronomy, mathematics, physics, geography and history), Ibn Sina (980) (in medicine), Ibn Haytham (965) (in optics) and many more. They also introduced terminologies in science which is being used in science until today such as algorithm, algebra, camera, music, chemistry, alkali, alcohol, cornea, and so on. They modified old theories and established new concepts in science such as the concept of light in optic and on how eyes see an object; heliocentric not geocentric concept of solar system as was mentioned in Syriac version of science.

Al-Qur'an is the holy script of Islam. Muslims in the past and present day regard al-Qur'an as the source of knowledge including the subject of science and technology. This paper discusses the concept of science from perspective of this holy script. Of course al-Qur'an is not the book of science, but al-Qur'an requests Muslims to observe and understand the entire scientific phenomenon around them.

Keywords: **al-Qur'an, observation, *Tawhidic* Science, scientific method, *quranic* epistemology.**

Malay Values in Scientific Inquiry

Abdul Latif Samian
Institute of Civilizational Islam, Universiti Kebangsaan Malaysia
abdlatif@ukm.edu.my

From the Malay perspective, as a general theory of ethics and values, God is *The Good,* i.e., *Yang Maha Baik*. Whatever that is good entails from The Good. There are a plethora of hierarchical goodness, either in the tangible or intangible forms and the dominant Malay worldview is founded on the Unity of God by way of the teachings of Islam. In this paper, the author examines the values of scientific inquiry espoused by scientists and scholars (particularly the Malay thinker Hamka) in the Malay world and civilization, taking into accounts both the esoteric and exoteric values of the Malays and the position of Islam in their worldview.

Keywords: **The Good, Ethics, Values, Truth**

15th R143 Session III

Methodological Individualism and Reductionism

Francesco Di Iorio
Nankai University

Methodological individualism (MI) does not have a good reputation in many sectors of the philosophy of social science because it is often regarded as committed to reductionism, where reductionism means an atomistic theory of society that is mistaken because it naively denies both the systemic nature of the social world and the structural constraints imposed on the individuals by sociocultural factors. This definition of MI is incorrect because there is no equivalence between MI and reductionism. Two variants of MI can be distinguished: a reductionist one, which has been theorized by the social contract theory and some atomistic economic approaches, and a nonreductionist one, which is rooted in the Scottish Enlightenment and includes individualist sociologists, members of the Austrian School of economics, Popper and his followers. The difference between these two variants of MI is often neglected in the MI literature. The entire individualist tradition is regarded as reductionist by many thinkers. The accusation of 'reductionism' levelled against the entire individualist tradition is expressed in two variants. The first, developed by Udehn and critical realists such as Archer and Bhaskar, interprets MI in terms of idealist reductionism; the second, developed by analytic philosophers such as Kincaid and Pettit, interprets MI in terms of semantic reductionism. I shall criticize both these interpretations of MI and demonstrate why they are historically and logically incorrect and cannot be applied to the nonreductionist variant of MI.

The Non-Identity Problem and the Social Choice Procedure Based on Asymmetrical Relationships

Reiko Gotoh
Hitotsubashi University

The problem of fetal Minamata disease has challenged accepted views of the medical science. Why an unborn child had to suffer severe physical and mental illnesses, even when his mother could escape from suffering? An absent-minded mother could mutter that her baby had absorbed all poison from her. The purpose of this paper is to reconsider the "Non-Identity problem" proposed by Derek Parfit (1984), which is summarized as follows. "Because we chose a Risky Policy, many people were later handicapped. If we had chosen an alternative Safe Policy, however, these handicapped people would never have existed." Can we accuse fetal Minamata disease without disvaluing lives of the babies born handicapped? Can any imaginary social choice procedure in which the future generation would choose policies solve this problem? Unfortunately, the answer may be No, since the future generationborn under the Risky Policy, may not choose the Safe Policy, which should result in denying their very existence. This paper responds to this difficult question, first, by restructuring the social choice procedure of the Rawlsian type, which assumes symmetrical relationships among individuals. Second, by constructing a *public reciprocal system*, which secures *basic well-being for all*, allowing for asymmetrical relationships among individuals, but which needs not make a complete ordering based on comparing individual values.

15th R144 Session IV

Jabir ibn Hayyan: the Islamic Philosophy of the Father of Chemistry

Ibrahim N. Hassan*, Mohd Yusof Hj Othman, Abdul Latif Samian
Institute of Islam Hadhari, The National University of Malaysia
*ibnhum@ukm.edu.my

Nearly 3000 cursive about chemistry, as well as several other sciences, was found belonging to the father of chemistry, Jabir ibn Hayyan. The foremost Muslim alchemist was born c. 721, Ṭūs, Khurasan and died c. 815, Al' Kūfah, Iraq. Jabir was Ja'far as-Sadiq's most noticeable student and a colleague of Imam Abu Hanifa, the founders of the Sunni Hanafi School of fiqh (Islamic jurisprudence). In addition to chemical and laboratory equipment and apparatuses, Jabir has developed a lot of chemical compounds, as well as medicines, aiming to help his people who suffer from diseases. The Jabirian corpus is renowned for its contributions to alchemy. It perfectly expresses the recognition of the importance of experimentation, "The first essential in chemistry is that thou shouldest perform practical work and conduct experiments, for he who performs not practical work nor makes experiments will never attain to the least degree of mastery.Therefore, in this paper, we will try to look at Jabir ibn Hayyan from Islamic point of view, attempting to discover his philosophy as a Muslim Chemist and how he developed Chemistry based on his viewpoint as a Muslim.

Chemical Decomposition and Analogical Reasoning in Humphry Davy's Electrochemistry

Jonathon Hricko, Yafeng Shan*
Education Center for Humanities and Social Sciences,
National Yang-Ming University
jonathon.hricko@gmail.com
*Durham University

How did chemists in the early nineteenth century know whether they had decomposed a substance and isolated one or more of its component substances? In order to answer this question, I focus on Davy's uses of electrolysis and analogical reasoning to draw conclusions regarding the composition of various substances. My goal is to show how analogy played a major role in driving progress in chemistry at a time when chemists lacked accurate theories about how such things as chemical bonding and electrolysis work. I highlight two uses of analogical reasoning in Davy's work. The first is found in Davy's 1806 Bakerian Lecture, which focuses on using a Voltaic pile to decompose substances whose composition was already more or less understood. The fact that Davy was able to obtain analogous results when using analogous substances and instruments gave him a better understanding of how to use a Voltaic pile to decompose substances. The second is found in Davy's 1807 Bakerian Lecture, which reports the use of this method to decompose two previously undecomposed substances, namely, potash and soda, and isolate potassium and sodium for the first time. Davy's success was largely due to the fact that he treated the decomposition of potash and soda as analogous to the decomposition of the substances he discussed in the previous year's Bakerian Lecture. More generally, Davy's conclusions regarding the electrochemical decomposition of various substances were grounded in relatively local analogies as opposed to general theories or principles.

15th R143 Session IV

Between Scylla and Charybdis: Investigating a Possibility of the Social Sciences

Kei Yoshida
Waseda University

The aim of this presentation is to investigate a possibility of the social sciences by overcoming both ethnocentrism and cultural relativism. As a matter of fact, there are many customs or habits in the world. True, we should welcome the plurality of cultures or worldviews, because there is no guarantee that only our view is correct. But what if some of the customs or habits are unacceptable to us because they look "brutal" or "irrational" from our point of view? According to cultural relativists, these customs or habits are relative to cultures, and all of our judgments are ethnocentric because they are based on our cultures or worldviews. Facing a choice between ethnocentrism and cultural relativism, anthropologists tend to opt for cultural relativism. But such a choice could be a serious obstacle to the social sciences. If cultural relativism is endorsed, then social scientific knowledge must be local and acceptable only to those who accept scientific thinking. Thus, some anthropologists such as Clifford Geertz abandon anthropology as a social science and argue that the social sciences should be merged with the humanities. In this presentation, I shall examine debates about cultural relativism and argue that we need to encourage mutual criticism while defending the plurality of cultures.

Knowledge and Big Data

Paul Dumouchel
Ritsumeikan University

This paper inquires into the relationship of big data to knowledge. The term "big data" is commonly used to refer to disparate cognitive systems, models and procedures that collect and exploit very large sets of data to make predictions about the behavior of financial and other markets, as well as individual consumers and collectives. Big data's claim range over a wide range of behaviors from crime rates to teacher's performance or the probability of defaulting on a loan. What is the relationship of these systems to knowledge? What type of knowledge are these predictions? Commercial, social, scientific? What is the epistemic value of the results of big data?

A central characteristic of these systems is that they are epistemically opaque. We do not know how they obtain their results. This opacity has two dimensions. First, the algorithms and model they rest on are unavailable for inspection because they constitute commercial secrets. Second, in science, especially astrophysics, molecular biology and climate studies many cognitive systems that use very large sets of data also are epistemically opaque. No one, even those who made them, knows exactly how they obtain their results. Is this also the case of big data?

Further, because big data systems are not pure cognitive systems, but active elements of social or commercial policies, they often include a feedback loop that transforms their predictions into self-fulfilling prophecies? This also raises the questions of their relationship to knowledge and of their claim to truth.

What the Unsupervised Learning Could Deliver Us (or, What Not)

Insok Ko
Inha University
insok@inha.ac.kr

The unsupervised learning, as a mode of machine learning, shall make a machine so clever that it distinguishes (the pictures of) cats from (those of) dogs, whatever specific breeds of cats and dogs were presented against whatever background. It is interesting that there is no need to feed the machine explicit information about the categories of cat and dog in order to get the machine to such level of classificatory competence. Though it is not yet clear, how far this kind of competence would reach, it gives us certain hope for an objective classification, i.e. for one that is free from prejudices or cultural biases. We might also overcome the problem of theoryladenness of observation. In this paper I will present an evaluation of this prospect, investigating the process and structure of unsupervised learning.

A Taxonomic Framework of Interventional Experiments in Biology

Hsiao-Fan Yeh*, Ruey-Lin Chen
*PhD, Department of Philosophy, National Chung Cheng University

Carl F. Craver and Lindley Darden (2013) build up a taxonomy of experiments in biology. They distinguish loosely three categories of experiments and reclassify every category into several subkinds. We think that this taxonomy of experiments is so complicated as to raise some problems. Our goal in this paper is to propose a new taxonomic framework of interventional experiments in biology. We argue for the four points: (1) This framework identifies two typical roles of experimentation: testing and discovering. (2) Intervention is used as essential means to realize the functional roles. (3) Interventions have two directions that are bottom-up and top-down and two effects that are excitatory and inhibitory. (4) A single experiment may perform two functions and use multiple patterns of intervention at one time.

15th R143 Session V

Staying Regular?

Alan Hájek
Australian National University
hajek.alan@anu.edu.au

'Regularity' conditions provide bridges between possibility and probability. They have the form:

If X is possible, then the probability of X is positive

(or equivalents). Especially interesting are the conditions we get when we understand 'possible' doxastically, and 'probability' subjectively. I characterize these senses of 'regularity' in terms of a certain internal harmony of an agent's probability space <omega, F, P>.

I review several arguments for regularity as a rationality norm. An agent could violate this norm in two ways: by assigning probability zero to some doxastic possibility, and by failing to assign probability altogether to some doxastic possibility. I argue for the rationality of each kind of violation.

Both kinds of violations of regularity have serious consequences for traditional Bayesian epistemology. I consider their ramifications for:

- conditional probability
- conditionalization
- probabilistic independence
- decision theory

Inference to the Hidden Factors

Linton Wang, Ming Yuan Hsiao*, Jhih Hao Jhang
Department of Philosophy, National Chung Cheng University
* Department of Philosophy, Soochow University

This paper aims at elaborating what we will call the inference to the hidden factors (IHF in short), and exploring its applications in the scientific and epistemic reasoning. IHF roughly takes the following form: (a) IHF entertains two kinds of premises including (i) theory-oriented but observation-related conditionals such as if P had been the case then Q would have been the case (i.e. $P > Q$), (ii) observation-related statements P but $\neg Q$, but (b) IHF entertains the conclusion that there is some R such that if P and R had been the case then Q would not have been the case (i.e. $\exists R((P \wedge R) > \neg Q)$). To elaborate IHF, we argue that the theory-oriented conditional must be some sort of defeasible conditional in that $P > Q \not\models P \supset Q$ (i.e. $P > Q$ does not logically entail $P \supset Q$), in which '$\supset$' stands for the material implication in the classical logic. To explore some examples of applying IHF, we show (a) how IHF is engaged in a scientific research agenda, (b) how engaging IHF makes a prediction failure give rise an anomaly to a scientific theory rather than falsify it, and (c) how engaging IHF can differentiate the epistemic status of probabilistic evidence from observational evidence.

Big Data, Logic of Scientific Discovery, and Abduction

Young E. Rhee
Kangwon National University

As information technology and artificial intelligence evolves, the role of big data is growing. Now, big data is a major driving force of the Fourth Industrial Revolution, and beyond industry it is affecting the fields such as science, engineering, humanities, social science, and arts. From the standpoint of philosophy of science, big data can offer insights and means for dealing both theoretical and methodological issues, and especially, it can contribute to philosophy of science by providing sophisticated methods and means to understanding scientific reasoning.

In this paper I examine the logic of discovery: deductive, inductive, and abductive method. The deductive method is the so-called theory-driven method that was typically exemplified in the discovery of Newton's laws of motion. The inductive method is the so-called data-driven method that was exemplified in the discovery of Kepler's laws of planetary motion. Big data has the capacity to do the process of inductive method as a logic of discovery, which is beyond the ability of a scientist or a scientific society. The abductive method is the third method of forming scientific theories but the process of its logic has not been properly explained so far. I suggest how big data can help us to explain the logic of discovery based on the abductive method.

Two Visual Systems and Phenomenology of Visual Consciousness

Feng Yu
Department of Philosophy, East China Normal University

Many neuroscientists and philosophers think that two visual systems (TVS) hypothesis (Milner & Goodale, 1992,1995,2006) is incompatible with the egocentric character of visual experience (Brogaard,2012; Wu,2014). According to their analysis, TVS argues for two claims, one of which is that the ventral stream of human visual systems, which contributes to our visual experience of the world, works in an allocentric frame of reference, whereas the dorsal stream, which the visual control of action, uses egocentric frames of reference. The other claim of TVS is that there is division of labor between the two visual streams, to wit, dorsal-stream processing for action does not contribute to the contents of visual experience and is largely isolated from ventral-stream processing. However, based on the following two premise, (1) Visual experience is egocentric, (2) Egocentric information is processed by the dorsal stream alone, the contradictive conclusion follows, namely visual experience is influenced by dorsal stream content. In this paper, I will explain there are three varieties of egocentric representations in the visual system and why there is no incompatibility between TVS and egocentricity of visual consciousness.

Are There Laws of Evolution?

Jun Otsuka
Kyoto University

Whether evolutionary theory has its own set of laws has been debated for decades, and most philosophers have answered to this question negatively. The skeptics emphasize that the inherent contingency of the living world rules out any possibility of non-trivial universal generalizations in biology. However, this conception of laws as universal generalizations can be questioned. In physics, laws are characterised not by universality but by invariance with respect to a certain transformation groups, such as Galilean or Lorentz transformations. I argue the question about lawfulness in biology should be reframed along this criterion, i.e. as a search for an evolutionary invariance/symmetry. I will introduce a few theoretical studies relevant to this research question and discuss their philosophical implications.

Individuating Genes as Types or Individuals

Ruey-Lin Chen
Department of Philosophy, National Chung Cheng University, Taiwan
pyrlc@ccu.edu.tw

In this paper, I argue that there are at least two kinds of individuation of genes. The transgenic technique can individuate “a gene” as an individual while the technique of gene mapping in classical genetics can only individuate “a gene” as a type or a kind. The two kinds of individuation involve different techniques, different objects that are individuated, and different references of the term “a gene”. Thus, I also discuss this semantic phenomenon in using “gene” and the problem about the relation between kinds and individuals in the individuation of genes.

Cognitive Decision Theory and Permissive Rationality

Jaemin Jung
Wonkwang University
gptp98@gmail.com

Cognitive Decision Theory (CDT) says that a rational credence is one with maximum expected *epistemic* value—one such that your expected *epistemic* value for how states of the world will turn out, given your adopting it, is at least as high as that of any alternative doxastic state you might adopt.

One of the central issues in epistemology is whether epistemic rationality is permissive or not: Some claim that (*Uniqueness*) for any total evidence, there is a unique doxastic state that any agent with that total evidence should take; others claim that (*Permissivism*) for some total evidence, there are multiple doxastic states that an agent with that total evidence can take.

How does *CDT* relate to the debate over permissive rationality? I present and assess an argument against *CDT* that goes as follows: On the assumption of *Epistemic Conservatism*, the correct theory of epistemic rationality will not endorse non-conservative doxastic state shifts from one credence function to another, in the absence of new evidence. In some cases, however, *CDT* endorses non-conservative doxastic state shifts, in the absence of new evidence. This seems unfortunate. However, I further show that when we clearly distinguish among several types of *Permissivism*, the argument is not a real threat to any CDTer. Depending on which version of *Permissivism* or *Uniqueness* a CDTer endorses, the argument may be avoided in one of two general ways. One response appeals to the stability of beliefs over time, while the other allows that the instability of beliefs over time fits naturally with epistemic rationality.

Problems of Intrinsic Time Direction within A-Theories

Kunihisa Morita, Dr.
Faculty of Arts and Science, Kyushu University
morita@artsci.kyushu-u.ac.jp

The philosophical community has long been concerned with the following questions concerning the world's temporal nature: (1) whether the world had a beginning, (2) whether the world is deterministic, and (3) whether time flows. Recently, a few physicists have also begun to study these issues, specifically question (1) and, since the advent of quantum mechanics, question (2). Relativity theory's concept of "space-time" seems to show that time does not flow independently, but instead that is related to space. In this presentation, I demonstrate that the world had no beginning, that it must be deterministic, and that if the world can be described by physics, then time does not flow. To arrive at these conclusions, I demonstrate the following two propositions: (i) the world had a beginning if and only if it is indeterministic and (ii) if time flows, the world must have had a beginning (and thus must be indeterministic).

Backtracking Analysis and Causal Ascription of Singular Historicals

Richard W. T. Hou
Department of Philosophy, National Chung Cheng University

One task of historians is to construct causal ascription of singular historicals between eminent historical events. For instance, the controversy resulting from the confusing butterfly ballot of Florida's year 2000 presidential election cost Gore his presidency. However, to research into these matters is inevitably to appeal to counterfactual deliberation in an epistemic fashion because the past is fixed. One standard idea is Max Weber's, Weber causation: "f was a cause of φ" is assertable iff "$\neg f \,\Box\!\!\rightarrow \neg\varphi$" is assertable. Reiss (2009) gives an exceptionally good analysis of this topic and outlines historians' reasoning, claiming that backtracking analyses of counterfactual conditionals employed in historical thought experiments is the signature of historical study of causal ascription of singular historicals. Nevertheless, he concludes that it is very difficult to reach an uncontroversial ascription for this sort in most cases. For this reason, he proposes to find difference-making relations that will suffice. The objective of this paper is to provide a more fine-grained, intervention-based, backtracking analysis of counterfactual conditionals upon which a more satisfactory account of causal ascription of singular historicals can be given. Reiss' account of difference-making relation will be shown to be unsatisfactory. Moreover, a formal ground of the epistemology of historical thought experiments can be given, along with the constraints of this account resultant from the semantic features of non-transitivity and strong centring of counterfactual conditionals. Finally, some epistemological points of causal ascription of singular historicals and historical thought experiments will be given.

Key words: **backtracking counterfactual analysis, historical thought experiment, intervention, causal ascription of singular historicals, Weber causation**

Practice-based Kuhnian Paradigms in Biological Sciences: Large-Scale Quantitative and Qualitative Analyses of a Case Study on Heart-Rate Variability

Karen Yan*, Meng-Li Tsai**, Tsung-Ren Huang+
* the first and corresponding author, Institute of Philosophy of Mind and Cognition, National Yang-Ming University
**Department of Biomechatronic Engineering, National Ilan University
+Department of Psychology, National Taiwan University

This paper aims to address the following two questions: Are Kuhnian paradigms applicable to biological sciences? If so, what information about biological sciences can we gain from applying these paradigms? We argue that Kuhnian paradigms are also applicable to biological sciences provided that we adopt Rouse's practice-based understanding of Kuhnian paradigm. We argue for this claim by analyzing 20,618 articles on heart-rate variability (HRV) from 1970 to 2016. We use three sets of tools to conduct our analyses: (1) a large-scale citation analysis, (2) a large-scale qualitative analysis, and (3) a large-scale text analysis. The resulting data supports the claim that there is a significant relationship between the quantitative and qualitative changes of HRV field and the publication of one highly cited review article. We then analyze the review article in detail and show that standardizing HRV practices is the core achievement of the article. We then suggest that a practice-based Kuhnian paradigm provides an understanding of how the standardizing work influences the quantitative and qualitative change of HRV field.

On the Notion of Interaction in the Nature-Nurture Debate

Qiaoying Lu
Sun Yat-sen University

The methods used in biometrics of ascribing causal responsibility to genotype has been criticized by interactionists over the years. And "gene-environment interaction" is the central notion regarding this version of nature-nurture debate. James Tabery in his recent book proposes that the controversy stems largely from the fact that biometricians and interactionists use the same term "interaction" to refer to different concepts. However, Tabery does not give a thorough investigation of how these two concepts of interaction raise disputes between two parties. This paper aims to distinguishing possible notions of interaction in biometric context, and identifying the one(s) that do posit a problem for biometricians. Firstly, three notions of interaction are defined in terms of biometric analysis. Namely, trivial interaction, vernacular interaction and statistical interaction, based on which three kinds of gene-environment interdependence will be given. By examining those notions in the interactionist context, I show that only the statistical interaction posits a challenge for biometric methods. Second, I revisit Tabery's own case study of the dispute between Hogben and Fisher, and show that it concerns exactly the statistical interaction, which ultimately amounts to an empirical question of whether this kind of interdependence is common in nature. Finally, I propose a potential challenge regarding the interpretation of statistical interaction: when there are changes of one or more unknown relevant factor(s) affecting the developmental processes of individuals, it would lead to a misrepresentation of interaction.

Debate of Permitted Limit of a Prosthetic Limb when Boarding

Jiwon Shim
the Institute of Human, Environment and Future at Inje Unversity in Southkorea
g1dmpkr@gmail.com

The human body is the most basic mediator connecting relationship the self and the world. Development of technology has been rearranged in a new relationship such intervention by the body. Robotics associated with the human spirit about the artificial intelligence and the human body, the human enhancement, have been actively studied in Korea. In contrast, the study of the implant body itself is an incomplete buildup problem. Body implants are subject to an ethical position somewhere between the object and the body. Body implant is to demonstrate that the things and the body, which is another subject of ethical values. And it presents the criteria for classification in accordance with artifacts, cybernetics, robotics.

I would like to discuss the following concrete examples. A South African disabled shot-putter was prevented from boarding with his prosthetic leg after 2016 Rio Paralympic. This led to the following discussion: the permissible range of the prosthetic arm or leg into the flight. Currently, the related airlines do not provide accurate judgment criteria for the allowance of bringing the prosthesis into the cabin. The battery, sensors, networks, biological and mechanical technologies are supporting the up-to-date prosthesis but the appropriate regulations are not provided yet. This study is intended to discuss the permissible range of a prosthesis into the flight and discuss the guidelines. The purpose of this paper is to provide a basis on which body implants can be recognized as part of the body.

Holobionts from an Immunological Perspective: the Problem of Pan-Homeostasism

Lynn Chiu
University of Bordeaux/CNRS

From an immunological point of the view, the problem of the holobiont is often framed as (a) the problem of maintaining the *homeostasis* between the host organism and its resident microbiota, (b) the problem of microbial-influence on immune *homeostasis*, (c) as well as the impact of the immune system on the *homeostasis* and dysbiosis of the microbiota.

These three aspects of the "immunological holobiont" have led some to argue that the host-microbiota homeostasis (the "superorganism") is an optimal adaptation (Eberl 2010), that adaptive immune system has been selected *for* its ability to manage the precarious balance between hosts and microbes (Lee & Mazmanian, 2010), or that immune "negotiations" enable the host and microbiota to (co)-evolve as a "team" (Gilbert et al. 2012).

Not all arguments on the physiological individuality of the holobiont rely on homeostasis theses (see Pradeu 2010), but to the extent that they do, it is important to first examine whether the homeostasis explanations are justified.

The goal of this paper is to uncover an ideology governing immunological theories and practice--"Pan-Homeostatism[4]"-- and explore to what extent it also frames immunological arguments about the holobiont. The project is modeled after similar criticisms and analyses of the "Adaptationism Program" in evolutionary biology (Gould and Lewontin 1979, Godfrey-Smith 2001, Lewens 2009).

Homeostasis is a foundational concept in immunology. The term is embedded in descriptions of the *function* of the immune system --to maintain the homeostasis of the organism (organism and tissue homeostasis), the *mechanisms* of immune responses -- the homeostatic regulation of immune responses (immune homeostasis), and the *necessary conditions* of immunity -- the homeostatic maintenance of T and B cell numbers and diversity (lymphocyte homeostasis).

First, I will identify four problematic explanatory strategies in immunology: (1) The all-homeostasis problem: presume that all (important) immunological phenomena are

[4] Maël Lemoine coined the term.

explained by homeostatic explanations, (2) The no-alternative problem: do not examine or test alternative hypotheses, (3) The proliferation problem: propose new subconcepts of homeostasis to explain discrepancies, (4) The heterogeneity problem: there is no unified definition of homeostasis. Second, I propose that there are empirical, disciplinary, explanatory, and methodological dimensions of Pan-Homeostasism. Last, I will characterize rival hypotheses within immunology, and consider their implications for the immunological holobiont.

How to Characterize the Individuality of Holobionts?

Shi-jian Yang
Department of Philosophy, Xiamen University

There is recently a heated debate upon the individuality of holobionts. A holobiont is a symbiotic collective formed by a multicellular animal/plant organism and the microbial community living inside its body. Godfrey-Smith insisted that holobionts with horizontal transmission do not have clear lineages, thus do not qualify as units of selection (2012). With the framework by Hull (1980), Booth raised a middle-way idea that a holobiont with horizontal transmission could be regarded as a unit of selection, only as far as an interactor, but not a replicator (2014). Pradeu showed that holobionts could be viewed as organisms, with the standard of immunological continuity (2012). On the other hand, Queller and Strassman argued that most holobionts do not qualified as organisms, using high cooperation and low conflict as the criteria of organismality (2016). Later, Doolittle and Booth raised a new idea that the holobiont construct an interaction patterns which could be viewed as units of selection. (2017). It seems that people have their own viewpoints and arguments, based on their conceptual frameworks respectively, but those systems could hardly be reconciled with each other. Here I propose that what has been missing in previous discussions is a careful examination on various levels of biological organization in the holobiont, which might serve as a common ground for reasonably analyzing the status of holobionts.

Keywords: **Holobiont; individuality; units of selection; organism**

Modes of Convergence to the Truth

Hanti Lin

UC Davis
ika@ucdavis.edu

December 1, 2017

Abstract

Those who engage in normative or evaluative studies of induction (such as formal epistemologists, statisticians, and theoretical computer scientists) have provided many positive results for justifying various kinds of inductive inferences. But they have said little about a very familiar kind: I call it *full enumerative induction*, which concerns inference to the *full* conclusion that *all* ravens are black (rather than the restricted conclusion that all the ravens observed in the future will be black). To remedy this, I develop a learning-theoretic solution in a way that can even be embraced by some Bayesians. The idea is to (i) define and study various modes of convergence to the truth, construed as epistemic ideals for an inquirer to achieve where possible, (ii) look for the the modes of convergence that can be achieved when tackling the problem of whether *all* ravens are black, (iii) of those modes, identify the one that corresponds to the highest achievable epistemic ideal, and (iv) see whether full enumerative induction can be justified as (that is, proved to be) a necessary means for achieving that epistemic ideal. The answer is positive, according to the main theorems of this paper. The Bayesian versions of those results are proved as well. The results are also extended for justification of a kind of Ockham's razor. The key to all these results is to introduce a mode of convergence slightly weaker than Gold's (1965) and Putnam's (1965) identification in the limit; I call it *almost everywhere convergence to the truth*, where the conception of "almost everywhere" is borrowed from geometry and topology.

Keywords: Enumerative Induction, Ockham's Razor, Convergence, Topology, Learning Theory, Bayesian Epistemology, Epistemic Norms, the Problem of Induction.

Contents

1 Introduction

The *general* problem of induction is the problem of (i) identifying the range of the inductive inferences that we can justify and (ii) exploring the extent to which we can justify them. Under the general problem there are several subproblems. For example, there is the special subproblem of how it is possible to reliably infer causal structures solely from observational data, which has attracted many statisticians, computer scientists, and philosophers.[1] And there is the more general subproblem of whether it is possible to escape Hume's dilemma—a dilemma that aims to trash any justification of any kind of inductive inference.[2] In this paper I want to focus on a subproblem of induction that is more *normative and evaluative* in nature.

Here is the background. Normative/evaluative studies of inductive inferenc are pursued in many mathematical disciplines: formal epistemology, statistics, and theoretical computer science. But somewhat curiously, they all have said little about the most familiar kind of inductive inference, of which an instance is this:

> (Full Enumerative Induction) We have observed this many black ravens and they all are black; so all ravens are black.

To be sure, those disciplines have had much to say about enumerative induction, but typically only about a *restricted* version that weakens the conclusion or strengthens the premise. Here is an example:

> (Restricted Enumerative Induction) We have observed this many black ravens and they all are black; so all the ravens *observed in the future* will be black

This is the kind of enumerative induction studied by Bayesian confirmation theorists,[3] and also by formal/algorithmic learning theorists.[4] Classi-

[1] For a groundbreaking treatment of this problem, see Spirtes, Glymour, and Scheines (1993[2001]).

[2] See Hume (1777) for his formulation of the dilemma, and Reichenbach (1938, sec. 38) for a more powerful formulation of the dilemma. A quite comprehensive list of attempted responses is provided in Salmon (1966: chap. 2). My own response is provided in unpublished manuscript Lin (2017) "Hume's Dilemma and the Normative Turn", available upon request.

[3] See, for example, Carnap (1950), Hintikka (1966), and Hintikka and Niiniluoto (1980). They all talk about the probabilistic inference from evidence $F(a_1) \wedge \ldots \wedge F(a_n)$ to the countable conjunction $\bigwedge_{i=1}^{\infty} F(a_i)$, where a_i means the i-th individual (or raven) observed in one's inquiry. If the index i is understood as enumerating all ravens in the universe in an order unbeknownst to the inquirer, then the evidence cannot be formalized the way they do.

[4] See, for example, Kelly (1996) and Schulte (1996).

cal statisticians and statistical learning theorists would study an *even more* restricted version that adds a substantial assumption of IID (independent and identically distributed) random variables, where the randomness is due to objective chance.

So they all set aside a serious study of full enumerative induction. But why? The reason for statisticians is obvious: their primary job is to study inductive inferences under the IID assumption or the like. The reason for Bayesians and formal learning theorists is deeper. Let me explain.

It is logically possible that there are nonblack ravens but an inquirer never observes one throughout her entire life (even if she is immortal). Call such a possibility a *Cartesian scenario of induction*, for it can be (but need not be) realized by a Cartesian-like demon who always hides nonblack ravens from the inquirer's sight. Each Cartesian scenario of induction has a *normal counterpart*, in which the inquirer receives exactly the same data in time (without observing a nonblack raven) and, fortunately, all ravens are black. A Cartesian scenario of induction and its normal counterpart are empirically indistinguishable, but in the first it is false that all ravens are black, while in the second it is true. And that causes some trouble for both formal learning theorists and Bayesians.

For Bayesians, justification of full enumerative induction requires justifying a prior probability distribution that strongly disfavors a Cartesian scenario of induction and favors its normal counterpart. Carnap and other Bayesian confirmation theorists seem to never mention such an anti-Cartesian prior in their papers, possibly because they cannot justify it or simply because they do not care. A radically subjective Bayesian would say that such a prior is epistemically permissible, but only because she thinks that any probabilistic prior is epistemically permissible—*even* including those that strongly favor a Cartesian scenario of induction and will thereby lead to counterinduction:

> (Counterinduction) We have observed this many ravens and they all are black; so, *not* all ravens are black.

So a radically subjective Bayesian must concede that counterinduction is epistemically permissible as well. To make this concession explicit is to invite worries and complaints from those who would like to have a justification for full enumerative induction and *against* counterinduction—and that is probably why we seldom hear those Bayesians talk about full enumerative induction.

Formal learning theorists do not fare better. When they justify the use of an inductive principle for tackling a certain empirical problem, they justify

it as a necessary means of achieving a certain epistemic ideal for tackling that problem. The epistemic ideal they have in mind is called *identification in the limit* (Gold 1965 and Putnam 1965), which is a logical guarantee to find the true hypothesis in *every* possible way for the inquiry to unfold indefinitely, by virtue of following a learning method. When applied to the problem of whether all ravens are black, identification in the limit sets an extremely high standard: finding the truth both in a Cartesian scenario and in its normal counterpart. So, that standard is too high to be achieved by any learning method (see proposition 3.6 below).[5] Where it is impossible to achieve identification in the limit, nothing can be justified as a necessary means for achieving that.

So those are the reasons why normative/evaluative studies of inductive inference have said little about full enumerative induction. And we are left with the problem of giving a justification for full enumerative induction, against counterinduction, and against the "no induction" policy. I call this problem the *Cartesian problem of induction*, because of the role played by Cartesian scenarios of the sort mentioned above. The point of pursuing this problem is not to respond to every conceivable kind of inductive skeptic, but to push ourselves to the limit—to explore the extent to which we can justify full enumerative induction.

The present paper aims to take a first step toward the first positive solution to that problem. The key to my proposal is that Bayesians should have been *more* learning-theoretic, and learning theorists should have been *truly* learning-theoretic, true to what I identify as the spirit of learning theory. Here is the idea:

1. (Mathematics) There are various modes of convergence to the truth. For example, formal learning theory studies identification in the limit, which may be called *everywhere* convergence to the truth, for it quantifies over all possible ways for the inquiry to unfold indefinitely. Statistics studies modes of stochastic convergence, such as *almost sure* convergence to the truth.

2. (Epistemology) Certain modes of convergence to the truth are epistemic ideals for an inquirer to achieve where possible. Some modes or their combinations correspond to higher epistemic ideals than some

[5]The same impossibility result remains even if we resort to *partial* identification in the limit, a weakening of identification in the limit due to Osherson et al. (1986), which requires that, in each possible way for the inquiry to unfold indefinitely, the true hypothesis is output infinitely often and each false hypothesis is output at most finitely often.

others do. We, as inquirers, ought to *achieve the best we can* when tackling an empirical problem.[6]

3. (Crux) But for tackling the raven problem—i.e. the problem of whether all ravens are black—it is provably impossible to achieve everywhere convergence to the truth. So, to achieve the best we can, we should *look for what can be achieved*. How? Well, given that it is impossible to converge everywhere, let's try to see whether it is at least possible to converge "almost everywhere"—to converge in "almost all" possible ways for the inquiry to unfold indefinitely. And it should not be difficult to try, because topologists have worked out a geometric conception of "almost everywhere" and "almost all".

4. (Proposal) Let's try to define various modes of convergence to the truth, including "almost everywhere" convergence, and study their combinations, such as "almost everywhere" convergence plus "monotonic" convergence. Find the combinations that are achievable for tackling the raven problem. Then, from among the achievable combinations, identify the one that corresponds to the highest epistemic ideal. And check whether that ideal is achieved *only* by following full enumerative induction rather than counterinduction. If the answer is positive, then that methodological principle is justified as a necessary means of achieving the highest achievable epistemic ideal for tackling the raven problem. But is the answer positive?

5. (Result) Yes, according to corollary 6.6. And the story just told can be retold for Bayesians.

All this is done by being true to the spirit of learning theory:

Look for what can be achieved.
Achieve the best you can.

In fact, that is the spirit leading Gold (1965) and Putnam (1965) to create one of the earliest branches of learning theory: formal/algorithmic learning theory.[7] And that is how I address the Cartesian problem of induction.

[6] For precursors to this epistemological idea, see Reichenbach (1938), Kelly (1996), and Schulte (1996).

[7] They address certain mathematical or empirical problems that are impossible to solve with an effective procedure—or, in the terminology of this paper, impossible to solve by achieving *everywhere* convergence to the truth with *perfect monotonicity*. What they do is in effect to drop perfectly monotonic convergence and see whether it is at least possible to achieve everywhere convergence.

The philosophical view presented above is what I call *learning-theoretic epistemology*, which has to be defended at greater length in a future work, building upon what other learning-theoretic epistemologists have done.[8] This paper is devoted to developing its logical and mathematical foundation, without which my presentation of that philosophical view would be mere hand-waving. That said, I will make several philosophical points to motivate the mathematical steps to be taken in this paper. I will also explain in detail how the mathematical results are meant to solve the Cartesian problem of induction, *assuming* learning-theoretic epistemology.

Here is the roadmap for this paper: Section 2 implements the above proposal and provides a pictorial but very informative sketch of the main results. So those who do not care about the mathematical details may stop reading at the end of that section. Sections 3 and 4 provide the mathematical preliminaries, from learning theory and from topology, respectively. Then section 5 defines the concept key to the present work: "almost everywhere convergence to the truth". Section 6 finishes the main result that justifies full enumerative induction. Section 7 provides an important application of the present work: justification of a certain kind of Ockham's razor. I conclude in section 8 by going back to the big picture, especially how the idea of modes of convergence to the truth *unifies* the many branches of learning theory and learning-theoretic epistemology. The presentation of the main results is kept to a minimum. So if you are interested in some of the details not required for understanding the statements of the main results (such as proofs, examples, and discussions of open problems), you will be directed to the relevant parts of the very long appendix.

To declare the style in use: Emphasis is indicated by *italics*, while the terms to be defined are presented in **boldface**.

2 Pictorial Sketch of Main Results

This section sketches the key definitions and the main results, and presents their philosophical applications.

The problem of whether all ravens are black, which I call the **hard raven problem**, can be partially represented by the tree in figure 1. There are two competing hypotheses: `Yes` and `No`. The observation of a black raven is represented by a datum `+`; a nonblack raven, `-`. Observations of nonravens

[8] See Kelly (2001, 2004) and Kelly and Glymour (2004) for their responses to the usual worries, such as the Keynesian worry "in the long run we all will die" and the Carnapian worry "no convergence criterion can constrain short-run inferential practices".

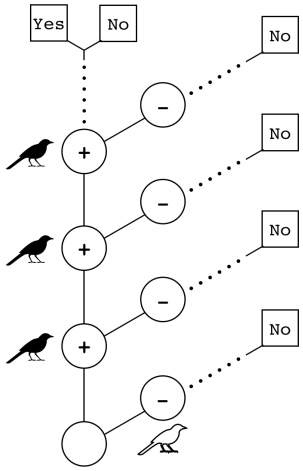


Figure 1: A partial representation of the hard raven problem

will be considered in subsequent sections but are omitted from the picture for simplicity. A data sequence (such as $(+, +, -)$) is a finite initial segment of a branch in the tree. A state of the world—i.e. a possible way for the inquiry to unfold indefinitely—is represented by an entire branch, which produces an infinite data stream (such as $(+, +, +, \ldots)$) and makes one of the completing hypotheses true (i.e. either `Yes` or `No`). The vertical branch that makes `No` true is one of the Cartesian scenarios of induction. There are actually an infinity of Cartesian scenarios if observations of nonravens are considered.

In general, an empirical **problem** specifies a set of competing hypotheses and a set of possible finite data sequences (possibly with some presuppositions that come with the problem). A **learning method** for that problem is a mapping that sends each finite data sequence to one of the competing hypotheses or to the question mark that represents suspension of judgment. A learning method is evaluated in terms of its truth-finding performance in each state contained in a **state space** $\mathcal{S}$, which consists of all possible ways for the inquiry to unfold indefinitely (without violating the presuppositions of the problem under discussion). Each such state makes exactly one of the competing hypotheses true and produces an infinite data stream $(e_1, e_2, e_3, \ldots)$, to be processed incrementally by a learning method M to output a sequence of conjectures $M(e_1), M(e_1, e_2), M(e_1, e_2, e_3), \ldots$ in time.

A learning method M is said to **converge to the truth** in a state $s \in \mathcal{S}$

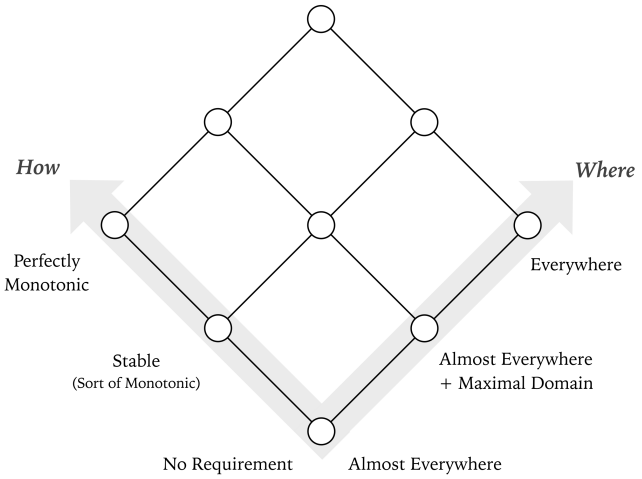


Figure 2: Modes of convergence to the truth, arranged by two dimensions

if, in state s, method M will eventually output the true hypothesis and then always continue to do so. To achieve **everywhere** convergence to the truth is to achieve convergence to the truth in *all* states in state space $\mathcal{S}$. This convergence criterion is what formal learning theorists call *identification in the limit*. This is only one of the many convergence criteria that concern the question of *where to converge*.

Let's consider the question of *where to converge* together with that of *how to converge*. Have a look at figure 2, in which various modes of convergence to the truth are arranged by two dimensions. The dimension that stretches to the upper right concerns where to converge. The other dimension, which stretches to the upper left, concerns how to converge. I introduce three modes for each of the two dimensions, so in combination there are nine modes to be considered in this paper. (Modes of stochastic convergence will not be considered because they are irrelevant to full enumerative induction.) Now I turn to explaining those modes of convergence.

A learning method M for a problem is said to achieve **almost everywhere** convergence to the truth if, for every competing hypothesis h considered in the problem, M converges to the truth in "almost all" states that make h true—or speaking geometrically, M converges to the truth "almost everywhere" on the topological space of the states that make h true. "**Almost everywhere**" is defined in a quite standard way in geometry and topology; it means "everywhere except possibly on a region that is negligible, i.e. nowhere dense, i.e. like a slice of hyper Swiss cheese that is

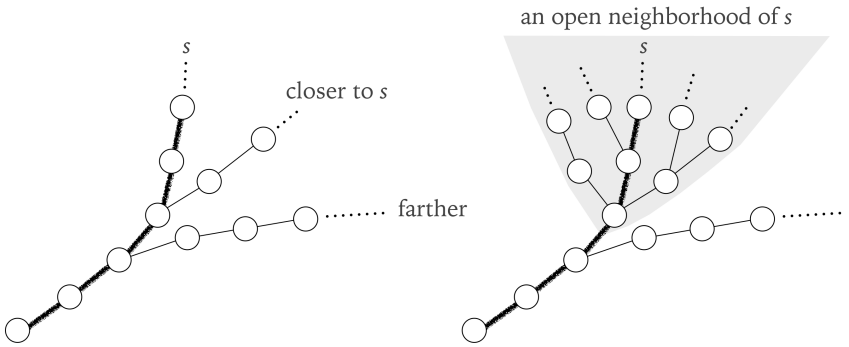


Figure 3: The empirical topology on a space of data streams or states

incredibly fully of holes". But which topological structure to use? I adopt what may be called **empirical topology**,[9] according to which:

(i) an open neighborhood of a state s is characterized as the set of states that are "close" to s to a certain degree (as depicted on the righthand side of figure 3);

(ii) a state "closer" to s is one that requires more data to empirically distinguish from s (as depicted on the lefthand side of figure 3).

It is a lot easier to explain the other modes of convergence. A learning method is said to converge to the truth on a **maximal domain** if there exists no learning method that converges to the truth in the same states and in strictly more states.

A learning method is said to achieve **perfectly monotonic** convergence to the truth if, whenever it outputs one of the competing hypotheses, it gets the truth and will continue to have the same output regardless of any further data received—it basically halts, just like an effective problem-solving method as studied in computability theory.

To be stable is to be "sort of" monotonic but not necessarily perfectly so. Specifically, a learning method is said to achieve **stable** convergence to the truth if, whenever it outputs a *true* competing hypothesis, it will continue

[9]Empirical topology is proposed by Vickers (1989) in computer science and by Kelly (1996) in epistemology.

to have the same output.[10] So this mode of convergence corresponds to the condition that, whenever the inquirer forms a belief in the true hypothesis considered in the problem, this belief is not merely a true opinion but has been "stabilized" or "tethered" to the truth, attaining the epistemic status that Plato values in *Meno* (see section 6). Finally, by "no requirement" I mean no requirement on how to converge.

The above finishes the sketch of the three modes of convergence on each of the two axes in figure 2. So there are nine "combined" modes of convergence arranged into a two-dimensional lattice structure, in which some modes are ordered higher than some others. A mode, if ordered higher, is mathematically stronger; it implies all the modes ordered lower in the lattice. That is mathematics. The following is epistemology. I make the evaluative assumption that:

> A mode of convergence to the truth, if ordered higher in the lattice in figure 2, corresponds to a higher epistemic ideal.

In fact, I even think that this assumption is obvious—or will become so after the definitions involved are rigorously stated and fully understood.[11]

The first main result is theorem 6.5 together with corollary 6.6, which says that, for tackling the hard raven problem, the achievable modes of convergence to the truth are the four in the shaded area of figure 4. So the highest achievable mode is the one marked by a star: "almost everywhere" + "maximal domain" + "stable". Furthermore, it can be achieved *only* by learning methods that implement full enumerative induction rather than counterinduction. This result relies on a crucial lemma—lemma 4.3—which says that, within the topological space of the states that make it false that

[10] Stable convergence to the truth is closely related to some properties that have been studied in learning theory, such as: Putnam's (1965) and Schulte's (1996) "mind-change"; Kelly and Glymour's (2004) "retraction"; Kelly, Genin, and Lin's (2016) "cycle". But these properties are defined only in terms of belief change without reference to truth. Stable convergence to the truth is a variant of the "no U-shaped learning" condition studied in Carlucci et al. (2005) and Carlucci et al. (2013), where U-shaped learning means the three-step process of believing in the truth, retracting it later, and then believing in the truth again.

[11] You might ask: When we compare two modes that are not ordered in the lattice—such that neither implies the other—how can we tell which corresponds to a higher epistemic ideal? This is a substantial issue, but in the present paper I do not need to take a position in order to justify full enumerative induction (although I do have an opinion, tending to think that the consideration about "where to converge" should be prioritized over the consideration about "how to converge"). See appendix A.5 for an interesting case study on this issue.

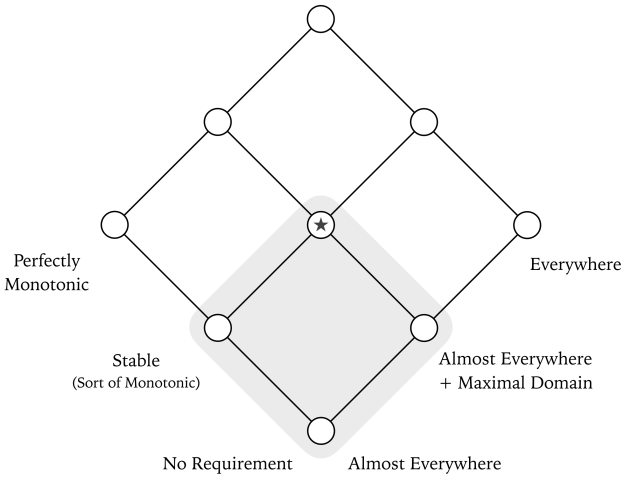


Figure 4: Modes of convergence achievable for the hard raven problem

all ravens are black, the set of the Cartesian scenarios is one of the (many) negligible regions.

Thanks to the above result, learning-theoretic epistemology can provide a justification for full enumerative induction and against counterinduction. The justification consists in the following argument:

Learning-Theoretic Argument

1. (Evaluative Premise) A mode of convergence to the truth, if ordered higher in the lattice in figure 2, corresponds to a higher epistemic ideal.

2. (Normative Premise) Given that an inquirer tackles a problem, she ought to follow a learning method that achieves, of the nine epistemic ideals in that lattice, the highest one achievable for tackling that problemprovided that such a uniquely highest one exists.

3. (Mathematical Premise) For tackling the hard raven problem, the achievable modes in that lattice are the four in the shaded area depicted in figure 4. (By corollary 6.6.)

4. So, given that an inquirer tackles the hard raven problem, she ought to follow a learning method that achieves the starred mode: "almost everywhere" + "maximal domain" + "stable". (By 1, 2, and 3.)

5. (Mathematical Premise) But that mode is achieved only by learning methods that implement full enumerative induction rather than counterinduction. (By corollary 6.6.)

6. *Therefore*, given that an inquirer tackles that problem, she ought to follow one of those inductive learning methods. (By 4 and 5.)

This is a *deductively valid* argument for a *normative conclusion* about induction, only with premises that are *mathematical, evaluative, or normative*.[12] This argument applies a general normative framework—premises 1 and 2—to the inquirers who tackle the hard raven problem. For those inquirers, premise 3 finds what can be achieved, step 4 identifies the best that can be achieved, and premise 5 points to a necessary means for achieving that: following one of the learning methods that implement full enumerative induction rather than counterinduction.

But exactly which of those inductive learning methods should those inquirers follow? Any of them, or only some, or only a unique one? On this issue the above argument is silent, although we might be able to refine it and argue for a stronger norm when we introduce *additional* modes of convergence to the truth. That said, the above argument seems to represent significant progress in justifying full enumerative induction.

The second main result is theorem 7.3, and one of its consequences is that, for tackling *any* problem, Ockham's razor of a certain kind is a necessary means for achieving any mode of convergence to the truth that implies "almost everywhere" + "stable", e.g. any of those in the shaded area in figure 5.[13] This kind of Ockham's razor says: "Do not accept a competing hypothesis more complicated than necessary for fitting the data you have", where a competing hypothesis is simpler (less complicated) if it is more parsimonious in terms of the capacity to fit data. In light of this result, a learning-theoretic epistemologist can argue that, given that an inquirer

[12]These features of the argument (as indicated by italics) are key to escaping from Hume's dilemma—or so I argue elsewhere *without* assuming learning-theoretic epistemology. The strategy is to defend and develop Reichenbach's idea that, *pace* Hume, a justification of induction need not be an argument for an empirical thesis about the uniformity of nature but can be an argument for a normative/evaluative thesis about how to pursue an inquiry (Reichenbach 1938: sec. 39). See my unpublished manuscript Lin (2017) "Hume's Dilemma and the Normative Turn", available upon request.

[13]This theorem, 7.3, extends and strengthens some aspects of my earlier work on Ockham's razor with co-authors (Kelly, Genin, and Lin 2016) in order to cover the hard raven problem and the like. But this theorem also simplifies and weakens some other aspects in order to highlight the core idea; in particular, the concept of Ockham's razor used here is simpler and weaker (but strong enough for present purposes).

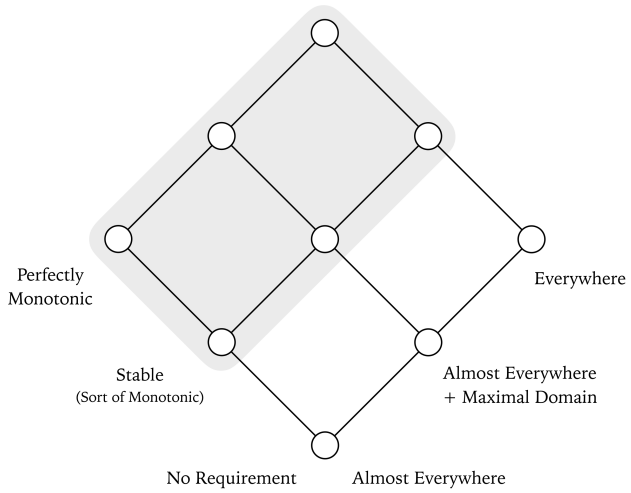


Figure 5: Those that can only be achieved by following Ockham's razor

tackles a problem for which the highest achievable mode of convergence to the truth implies "almost everywhere" + "stable", she ought to comply with the kind of Ockham's razor just mentioned.

The connection between the two main results is that, as we will see, any counterinductive inference violates that kind of Ockham's razor. In fact, a part of the first main result is proved as a corollary of the second main result.

The results sketched above all have their Bayesian versions. Those who believe in radically subjective Bayesianism would not care, but some other Bayesians would and should. According to what I call *learning-theoretic Bayesianism*, the Bayesian inquirers who tackle a problem can be evaluated as good or bad inquirers for finding the truth among the competing hypotheses, depending partly on what probabilistic priors they have. For example, a Bayesian inquirer can be evaluated as a good inquirer for tackling a certain problem only if her probabilistic belief converges via conditionalization to full certainty in the true hypothesis everywhere (or almost everywhere)—provided that, of course, that epistemic ideal is achievable for tackling that problem. Learning-theoretic Bayesianism also evaluates inquirers (or their priors) in terms of other modes of convergence; for example, modes of stochastic convergence have to be considered in order to evaluate inquirers who tackle statistical problems. It seems to me that many Bayesians are at least sometimes learning-theoretic rather than radically

subjective. Indeed, Bayesian statisticians caution against assigning a zero prior to a statistical hypothesis under discussion. And one of their typical reasons is that doing so will make it impossible to converge stochastically to full certainty in the truth if that statistical hypothesis is true. This reason is learning-theoretic in nature. It is just that, now, the modeling of belief is not qualitative but probabilistic.

The above summarizes the main results and the way they are used to solve the Cartesian problem of induction, *assuming* learning-theoretic epistemology. The rest of this paper is devoted to the mathematical details.

3 Preliminaries: Learning Theory

This section reviews some definitions familiar in formal learning theory.

Definition 3.1. A **problem** is a triple $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ consisting of:

- a hypothesis space $\mathcal{H}$, which is a set of competing **hypotheses**,
- an evidence space $\mathcal{E}$, which is a set of finite **data sequences** $(e_1, \ldots, e_n)$,
- a state space $\mathcal{S}$, which is a set of possible **states** of the world taking this form: $(h, \vec{e})$, where:
  - h, called the uniquely **true** hypothesis in this state, is an element of $\mathcal{H}$;
  - $\vec{e}$, called the **data stream** produced in this state, is an infinite sequence of data, written $\vec{e} = (e_1, e_2, e_3, \ldots)$, whose initial segments $(e_1, \ldots, e_n)$ are all in $\mathcal{E}$.

The state space $\mathcal{S}$ of a problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is meant to capture the *presupposition* of that problem in this way: $\mathcal{S}$ consists of all possible ways for the inquiry to unfold indefinitely without violating the presupposition.

Definition 3.2. A **learning method** for a problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is a function:

$$M : \mathcal{E} \to \mathcal{H} \cup \{?\},$$

where ? represents suspension of judgment. Given each data sequence $(e_1, \ldots, e_n) \in \mathcal{E}$, the output of M is written as $M(e_1, \ldots, e_n)$.

Example 3.3. The **hard raven problem** poses this question: *"Are all ravens black?"* This problem is partially represented by the tree structure in figure 1 and is formally defined as follows.

- The hypothesis space $\mathcal{H}$ is $\{\texttt{Yes}, \texttt{No}\}$, where:
  - `Yes` means that all ravens are black,
  - `No` means that not all ravens are black.
- The evidence space $\mathcal{E}$ consists of all finite sequences of `+`, `0`, and/or `-`, where:
  - datum `+` denotes the observation of a black raven;
  - datum `-`, a nonblack raven;
  - datum `0`, something else.
- The state space $\mathcal{S}$ consists of all states in one of the following three categories:[14]
  - (a) the states $(\texttt{Yes}, \vec{e})$ in which $\vec{e}$ is a `+/0` sequence,
  - (b) the states $(\texttt{No}, \vec{e})$ in which $\vec{e}$ is a `+/0` sequence.
  - (c) the states $(\texttt{No}, \vec{e})$ in which $\vec{e}$ is a `+/0/-` sequence that contains at least one occurrence of `-`.

The second category (b) contains the states in which there are nonblack black ravens but the inquirer will never observe one, so they are the **Cartesian scenarios of induction**.

Example 3.4. The **easy raven problem** is basically the same as the hard raven problem except that its state space consists only of the states in categories (a) and (c), ruling out the Cartesian scenarios of induction. So the easy raven problem *presupposes* that the inquirer is not living in a Cartesian scenario of induction. It poses this question: *"Suppose that you are not living in a Cartesian scenario of induction, then are all ravens black?"*

Definition 3.5. Let M be a method for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$. M is said to **converge to the truth** in a state $(h, \vec{e}) \in \mathcal{S}$ if

$$\lim_{n\to\infty} M(e_1, \ldots, e_n) \;\; = \;\; h\,,$$

namely, there exists a positive integer k such that, for each $n \geq k$, we have that $M(e_1, \ldots, e_n) = h$. M is said to converge to the truth **everywhere** for $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ if it converges to the truth in every state contained in $\mathcal{S}$.

[14] Well, there is a fourth category: the states $(\texttt{Yes}, \vec{e})$ in which $\vec{e}$ contains some occurrence of `-` (a nonblack raven). But such states are logically impossible, so they need not be considered.

Then we have the following negative result:

Proposition 3.6. *Although the easy raven problem has a learning method that converges to the truth everywhere, the hard raven problem does not.*

Proof. The first part is a classic, well-known result in learning theory. To prove the second part, let $\vec{e}$ be an infinite `+/0` sequence. Consider state $s = (\texttt{Yes}, \vec{e})$ and its Cartesian counterpart $s' = (\texttt{No}, \vec{e})$. It suffices to note that, for any learning method M for the hard raven problem, M converges to the truth in s iff M fails to do so in s' (because these two states are, so to speak, empirically equivalent). So there is no learning for the hard raven problem that converges to the truth everywhere. □

No one can escape such a negative result; Bayesians are no exceptions. The rest of this section is meant to make this clear.

Definition 3.7. Let a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ be given. Subsets of state space $\mathcal{S}$ are called **propositions**. Hypothesis $h \in \mathcal{H}$ and data sequence $(e_1, \ldots, e_n) \in \mathcal{E}$ are understood to express the following propositions:

$$\begin{aligned} |h| &= \{(h', \vec{e}\,') \in \mathcal{S} : h' = h\}; \\ |(e_1, \ldots, e_n)| &= \{(h', \vec{e}\,') \in \mathcal{S} : \vec{e}\,' \text{ extends } (e_1, \ldots, e_n)\}. \end{aligned}$$

That is, $|h|$ is the set of states in $\mathcal{S}$ that make hypothesis h true, and $|(e_1, \ldots, e_n)|$ is the set of states in $\mathcal{S}$ that produce data sequence $(e_1, \ldots, e_n)$.[15] Let $\mathcal{A}_\mathcal{P}$ denote the smallest σ-algebra that contains the above propositions for all $h \in \mathcal{H}$ and all $(e_1, \ldots, e_n) \in \mathcal{E}$. Given a probability function $\mathbb{P}$ defined on that algebra, I will write $\mathbb{P}(h)$ as a shorthand for $\mathbb{P}(|h|)$. Similarly, I will write $\mathbb{P}(e_1, \ldots, e_n)$ and $\mathbb{P}(h \,|\, e_1, \ldots, e_n)$, where the latter stands for conditional probability as defined the standard way.[16]

Definition 3.8. A **probabilistic prior** for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is a probability function $\mathbb{P}$ defined on σ-algebra $\mathcal{A}_\mathcal{P}$ with $\mathbb{P}(e_1, \ldots, e_n) > 0$ for each data stream $(e_1, \ldots, e_n) \in \mathcal{E}$. $\mathbb{P}$ is said to (have its posteriors) **converge to the truth** in a state $s = (h, \vec{e}) \in \mathcal{S}$ if

$$\lim_{n \to \infty} \mathbb{P}(h \,|\, e_1, \ldots, e_n) \;=\; 1\,,$$

[15] If you like, the concept of problems and other learning-theoretic concepts can be defined purely in terms of propositions, as done in Baltag et al. (2015) and Kelly et al. (2016).

[16] Namely, $\mathbb{P}(A \,|\, B) = \mathbb{P}(A \cap B)/\mathbb{P}(B)$.

that is, for any $\epsilon > 0$, there exists a positive integer k such that, for each $n \geq k$, $\mathbb{P}(h \,|\, e_1, \ldots, e_n) > 1 - \epsilon$. $\mathbb{P}$ is said to converge to the truth **everywhere** for problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ if it converges to the truth in each state in $\mathcal{S}$.

Proposition 3.9. *The hard raven problem has no probabilistic prior that converges to the truth everywhere.*

Proof. Copy the proof of proposition 3.6, paste it here, replace 'learning method' by 'probabilistic prior', and replace 'M' by '$\mathbb{P}$'. □

4 Preliminaries: Topology

When everywhere convergence to the truth cannot be achieved, let's see whether it is possible to achieve at least *almost* everywhere convergence to the truth. This section introduces a conception of "almost everywhere" that is geometrical, topological, and empirical.

Let a space X of data streams be given. Choose an arbitrary data stream $\vec{e}$ therein, and take it as the actual data stream to be received incrementally by the inquirer, as depicted on the lefthand side of figure 3. Consider an alternative data stream $\vec{e}\,'$ that is **identical** to $\vec{e}$ **up until** stage n; namely, $e'_i = e_i$ for each $i \leq n$ but $e'_{n+1} \neq e_{n+1}$. The larger n is, the later the point of departure is and the more data one needs to distinguish those two data streams. So, the larger n is, the harder it is to empirically distinguish those two data streams, and the "closer" $\vec{e}\,'$ is to the actual data stream $\vec{e}$—"closer" in an empirical sense. Consider the set of the data streams that are at least "that close" to $\vec{e}$:

$$N_n(\vec{e}) \;\; = \;\; \{\vec{e}\,' \in X : e'_i = e_i \text{ for each } i \leq n\}.$$

Take that as a *basic open neighborhood* of point $\vec{e}$ in space X, as depicted on the righthand side of figure 3. Such open neighborhoods provably form a *topological base* of X.[17]

Similarly, given the space $|h|$ of states that make a certain hypothesis h true, two states are close (hard to distinguish empirically) iff the data streams therein are close. So a basic open neighborhood of a state $s = (h, \vec{e})$

[17] Here is the standard definition of topological bases. Given a set X of points, a family $\mathcal{B}$ of subsets of X is called a **topological base** if (i) every point in X is contained in some set in $\mathcal{B}$ and (ii) for any $B_1, B_2 \in \mathcal{B}$ and any point $x \in B_1 \cap B_2$, there exists $B_3 \in \mathcal{B}$ such that $x \in B_3 \subseteq B_1 \cap B_2$.

in $|h|$ takes the following form:

$$\begin{aligned} N_n\Big((h,\vec{e})\Big) &= \{(h,\vec{e}') \in |h| : e'_i = e_i \text{ for each } i \leq n\} \\ &= |h| \cap |(e_1,\ldots,e_n)|\,. \end{aligned}$$

Such neighborhoods provably form a topological base of $|h|$, turning $|h|$ into a topological space. This is the topological base we will use. It is determined by the empirical distinguishability between states—distinguishability in terms of (finite) data sequences.

Definition 4.1. Given a problem $\mathcal{P} = \big(\mathcal{H},\mathcal{E},\mathcal{S}\big)$ and a hypothesis $h \in \mathcal{H}$, the **empirical topological base** of $|h|$ is the family of open neighborhoods constructed above; namely, it is defined as follows:

$$\begin{aligned} \mathcal{B}_{|h|} &= \Big\{N_n(s) : s \in |h| \text{ and } n \in \mathbb{N}^+\Big\} \\ &= \Big\{|h| \cap |(e_1,\ldots,e_n)| : (e_1,\ldots,e_n) \in \mathcal{E}\Big\} \smallsetminus \Big\{\varnothing\Big\}\,. \end{aligned}$$

I now turn to some concepts borrowed from general topology.

Definition 4.2. Let X be a topological space equipped with a topological base $\mathcal{B}_X$. A **negligible** (or **nowhere dense**) region within X is a subset R of X such that, for each nonempty open neighborhood $N \in \mathcal{B}_X$, there exists a nonempty open neighborhood $N' \in \mathcal{B}_X$ that is nested within N and disjoint from R.

A negligible region is like a slice of "hyper" Swiss cheese, incredibly full of holes: wherever you are in the ambient space X, say point x, and however small a basic open neighborhood of x is considered, say N, then within N you can always find an open "hole" N' of that slice of Swiss cheese. Here is an example:

Lemma 4.3. *In the hard raven problem, the Cartesian scenarios of induction (i.e. the states* $(\texttt{No},\vec{e})$ *with* $\vec{e}$ *being a* `+/0` *sequence) form a negligible region within the topological space* $|\texttt{No}|$.

Proof. Each nonempty basic open neighborhood in the topological space $|\texttt{No}|$, say $N = |\texttt{No}| \cap |(e_1,\ldots,e_n)|$, includes a nonempty basic open neighborhood, namely $N' = |\texttt{No}| \cap |(e_1,\ldots,e_n,\texttt{-})|$, which is disjoint from the set of the Cartesian scenarios of induction. □

With "negligible", we can define "almost everywhere" and "almost all" the standard way in topology:

Definition 4.4. Consider a property that may or may not apply to points in a topological space X. That property is said to apply **almost everywhere** on space X if it applies to all points in $X \setminus X'$, where X' is some negligible region within X. In that case, also say that it applies to **almost all** points in space X.

See appendix A.1 for a more detailed review of "almost everywhere" in general topology.[18]

5 Mode (I): "Almost Everywhere"

We are finally in a position to define the key concept that kickstarts the new learning theory:

Definition 5.1. A learning method M for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is said to converge to the truth **almost everywhere** if, for each hypothesis $h \in \mathcal{H}$, M converges to the truth in almost all states that make h true—or speaking geometrically, M converges to the truth almost everywhere on topological space $|h|$.

This definition makes possible a series of positive results. Here is the first one:

Proposition 5.2. *The hard raven problem has a learning method that converges to the truth almost everywhere.*

Proof. By the preceding result, lemma 4.3, the topological space $|\texttt{No}|$ has the following negligible region:

$$C \;=\; \{(\texttt{No}, \vec{e}) : \vec{e} \text{ is a } \texttt{+/0} \text{ sequence}\},$$

which consists of the *C*artesian scenarios of induction. So it suffices to an example of a learning method that converges to the truth in:

- every state in $|\texttt{Yes}|$,
- every state in $|\texttt{No}| \setminus C$.

The following learning method does the job:

M^*: "Output hypothesis `No` if you have observed a nonblack raven (`-`); otherwise output `Yes`."

[18] Also see Oxtoby (1996) for an elegant presentation.

This finishes the proof. □

Despite the above positive result, there are problems for which it is impossible to achieve almost everywhere convergence to the truth. Examples are provided in appendix A.2; their philosophical implications are briefly discussed in appendix A.3. The above positive result can be reproduced for Bayesians as follows.

Definition 5.3. A probabilistic prior $\mathbb{P}$ is said to converge to the truth **almost everywhere** for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ if, for each hypothesis $h \in \mathcal{H}$, $\mathbb{P}$ converges to the truth almost everywhere on topological space $|h|$.

Proposition 5.4. *The hard raven problem has a probabilistic prior that converges to the truth almost everywhere.*

Proof. Immediate from theorem 6.8, to be presented and proved below. □

Well, the above is only a first step toward justifying full enumerative induction. For there remains a subproblem, which may be called the *problem of counterinduction*: Almost everywhere convergence to the truth, alone, is so liberal that it is witnessed also by some crazy learning methods that apply counterinduction. Here is an example:

$M^\dagger$: "Output hypothesis `No` if you have observed a nonblack raven (`-`) or if everything you have observed is a black raven (`+`); output `Yes` if every raven you have observed is black and you have observed a nonraven (`0`)."

This method converges to the truth almost everywhere for the hard raven problem because it converges to the truth in:

- every state in $|\texttt{Yes}| \setminus \{s\}$,
  where $s = (\texttt{Yes}, \text{the constant sequence of } \texttt{+})$,

- every state in $|\texttt{No}| \setminus C'$,
  where $C' = C \setminus \{s'\}$ and $s' = (\texttt{No}, \text{the constant sequence of } \texttt{+})$.

The singleton $\{s\}$ is a negligible region within $|\texttt{Yes}|$; C', within $|\texttt{No}|$.

The learning method $M^\dagger$ defined above applies counterinduction. It will be ruled out by the modes of convergence to be introduced below.

6 Modes (II) and (III): "Stable" and "Maximal"

Before one's opinion converges to the truth, it might be false, or it might be true but to be retracted as data accumulate. But when one's opinion has converged to the truth, it is "tied" to the truth and will not "run away", which seems epistemically valuable. Plato expresses the same idea in *Meno*:

> True opinions are a fine thing and do all sorts of good so long as they stay in their place, but they will not stay long. They run away from a man's mind; so they are not worth much until you *tether* them by working out a reason. ... Once they are *tied down*, they become knowledge, and are *stable*. That is why knowledge is something more valuable than right opinion. What distinguishes the one from the other is the *tether*. (Emphasis mine.)

Hence the following definition:

Definition 6.1. A learning method M is said to **have converged** to the truth given the n-th stage of inquiry in a state $s = (h, \vec{e})$ if

$$M(e_1, \ldots, e_k) \quad = \quad h \quad \text{for each } k \geq n\,.$$

With the above concept we can define the following two epistemic ideals:

Definition 6.2. A learning method M for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is said to converge to the truth with **perfect monotonicity** if

> M has converged to the truth given a stage n in a state $s = (h, \vec{e})$ whenever $M(e_1, \ldots, e_n) \neq \texttt{?}$.

Say that M converges to the truth with **stability** if the following (weaker) condition holds:

> M has converged to the truth given a stage n in a state $s = (h, \vec{e})$ whenever $M(e_1, \ldots, e_n) = h$ (the hypothesis true in s).

Stable convergence is sort of monotonic but not necessarily perfectly so, while perfect monotonicity can be very demanding. Indeed, when a learning method for a problem achieves *everywhere* convergence to the truth with *perfect monotonicity*, it is basically what a computability theorist would call an *effective procedure* for solving that problem.[19] That combination of modes, "everywhere" + "perfectly monotonic", is a great thing to have

[19] Well, ignoring whether the learning method in question is a computable function.

where achievable. But it is too high to be achievable for any problem that is inductive in nature, such as the hard raven problem. In fact, we have a stronger negative result:

Proposition 6.3. *For the hard raven problem, it is impossible to simultaneously achieve the following two modes of convergence to the truth: "almost everywhere" and "perfectly monotonic".*

And the following is the last mode of convergence needed to state the first main result:

Definition 6.4. A learning method M for a problem is said to converge to the truth on a **maximal domain** if there is no learning method for the same problem that converges to the truth in all states where M does and in strictly more states.[20]

Then we have the first main result:

Theorem 6.5. *The hard raven problem has a learning method that converges to the truth (i) almost everywhere, (ii) on a maximal domain, and (iii) with stability. Every such learning method M has the following properties:*

1. *M is* **never counterinductive** *in that, for any data sequence $(e_1, \ldots, e_n)$ that has not witnessed a nonblack raven, $M(e_1, \ldots, e_n) \neq$ `No`;*
2. *M is* **enumeratively inductive** *in that, for any data stream $\vec{e}$ that never witnesses a nonblack raven, $M(e_1, \ldots, e_n)$ converges to* `Yes` *as $n \to \infty$.*

The idea that underlies the proof of the above theorem is explained in appendix B.1, which I have tried to make as instructive as possible. You do not want to miss it if you are interested in how exactly stable convergence helps to argue against counterinduction. The proof itself is in appendix B.3.

Corollary 6.6. *Consider the modes of convergence to the truth arranged in the lattice in figure 4. The four modes in the shaded area are exactly those achievable for the hard raven problem. To achieve the strongest of those four, namely "almost everywhere" + "maximal" + "stable", a necessary means is to follow one of the learning methods that are enumeratively inductive and never counterinductive.*

[20] I am indebted to Konstantin Genin for bringing this concept to my attention.

This corollary follows immediately from preceding results: propositions 3.6 and 6.3 and theorem 6.5.

The rest of this section retells the essential part of the above story in Bayesian terms.

Definition 6.7. Let $\mathbb{P}$ be a probabilistic prior for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$. $\mathbb{P}$ is said to **have started to stably converge** to the truth given stage n in state $s = (h, \vec{e}) \in \mathcal{S}$ if

1. $\mathbb{P}(h \,|\, e_1, \ldots, e_n, \ldots, e_{n+i})$ is monotonically increasing as a function of i defined on $\mathbb{N}$,

2. $\mathbb{P}(h \,|\, e_1, \ldots, e_n, \ldots, e_{n+i})$ converges to 1 as $i \to \infty$.

$\mathbb{P}$ is said to converge to the truth with **stability** if, for each hypothesis $h \in \mathcal{H}$, for each state $s = (h, \vec{e}) \in \mathcal{S}$ that makes h true, and for each stage n as a positive integer, if $\mathbb{P}(h \,|\, e_1, \ldots, e_n) > 1/2$, then $\mathbb{P}$ has started to converge to the truth given stage n in state s.

Theorem 6.8. *The hard raven problem has a probabilistic prior that converges to the truth (i) almost everywhere, (ii) on a maximal domain, and (iii) with stability. Every such probabilistic prior $\mathbb{P}$ has the following properties:*

1. *$\mathbb{P}$ is* **never counterinductive** *in that, for any data sequence $(e_1, \ldots, e_n)$ that has not witnessed a nonblack raven, $\mathbb{P}(\texttt{No} \,|\, e_1, \ldots, e_n) \leq 1/2$;*

2. *$\mathbb{P}$ is* **enumeratively inductive** *in that, for any data stream $\vec{e}$ that never witnesses a nonblack raven, $\mathbb{P}(\texttt{Yes} \,|\, e_1, \ldots, e_n)$ converges to 1 as $n \to \infty$.*

All the results of this section are proved in appendix B.3. The above finishes the mathematical result employed to justify full enumerative induction. I now turn to an immediate application: justification of Ockham's razor.

7 "Almost Everywhere" + "Stable" $\Longrightarrow$ "Ockham"

I have presented a result (corollary 6.6) that can be used to justify a norm that says when to follow this methodological principle:

Do not accept a counterinductive hypothesis.

(When? At least when tackling the hard raven problem.) This section presents a similar result, which can be used to justify a norm that says when to follow this methodological principle:

Do not accept a hypothesis if it is more complicated than necessary for fitting data.

This is Ockham's razor of a certain kind, where a hypothesis is simpler iff it is parsimonious in terms of the capacity to fit data.

To be more precise:

Definition 7.1. Let a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ be given. A data sequence $(e_1, \ldots, e_n)$ and a hypothesis h therein are said to be **compatible** if the propositions they express have a nonempty overlap, which also means that there exists a state in $\mathcal{S}$ that makes hypothesis h true and produces data sequence $(e_1, \ldots, e_n)$. For each hypothesis $h \in \mathcal{H}$, let $\mathcal{E}(h)$ denote the set of data sequences in $\mathcal{E}$ that are compatible with h (so $\mathcal{E}(h)$ captures the data-fitting capacity of h). The **empirical simplicity order**, written $\prec$, is defined on $\mathcal{H}$ as follows: for all hypotheses h and $h' \in \mathcal{H}$,

$$h \prec h' \quad \text{iff} \quad \mathcal{E}(h) \subset \mathcal{E}(h').$$

Or in words, h is **simpler** then h' iff h "fits" strictly less data sequences than h' does. Say that h is **no more complex** than h' if $h' \not\prec h$.

In the hard raven problem, for example, the inductive hypothesis `Yes` is simpler than the counterinductive hypothesis `No`.

Definition 7.2. A learning method M for a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is said to follow **Ockham's tenacious razor** just in case, for each hypothesis $h \in \mathcal{H}$ and each data sequence $(e_1, \ldots, e_n) \in \mathcal{E}$, h is the output of M given $(e_1, \ldots, e_n)$ only if

- (*Razor Condition*) h is no more complex than any hypothesis in $\mathcal{H}$ that is compatible with $(e_1, \ldots, e_n)$;
- (*Tenacity Condition*) h continues to be the output of M given any data sequence in $\mathcal{E}$ that extends $(e_1, \ldots, e_n)$ and is compatible with h.

In other words, a learning method M follows Ockham's tenacious razor just in case, whenever M outputs a hypothesis h, h is no more complex than necessary for fitting the available data and h will continue to be the output until it is refuted by the accumulated data. In the hard raven problem, to

comply with the razor condition is exactly to be never counterinductive—to never infer `No` whenever one has not observed a nonblack raven.

Then we have the second main result:

Theorem 7.3 (Ockham Stability Theorem)**.** *Let M a learning method for a problem that converges to the truth almost everywhere. Then following two conditions are equivalent:*

1. *M converges to the truth with stability.*
2. *M follows Ockham's tenacious razor.*

I call it the *Ockham stability theorem*. I understand its epistemological significance as follows. Almost everywhere convergence to the truth is a fundamental epistemic ideal to strive for whenever it is achievable. Convergence with stability is also good epistemically, but without almost everywhere convergence, it is not clear what value there is in achieving only stability. So, almost everywhere convergence first, stable convergence second. Given almost everywhere convergence, we might want to strive (further) for stable convergence (if that is possible), and to achieve that is *exactly* to follow Ockham's tenacious razor, as stated in the above theorem.

An immediate application of the *1* $\Rightarrow$ *2* side of the Ockham stability theorem is to prove, as a corollary, the "never be counterinductive" part of theorem 6.5. For, when tackling the hard raven problem, to be never counterinductive is exactly to comply with the razor condition. So the Ockham stability theorem helps to justify a local norm of Ockham's razor: Given that an inquirer tackles the hard raven problem, she ought to always follow Ockham's tenacious razor and, hence, never apply counterinduction.

We can use the preceding theorem to justify the use of Ockham's razor for tackling other problems, such as curve-fitting problems. See appendix A.4 for an example.

The rest of this section retells the essential part of the above story in Bayesian terms.

Definition 7.4. A probabilistic prior $\mathbb{P}$ for a problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ is said to follow **Ockham's tenacious razor** just in case, for each hypothesis $h \in \mathcal{H}$ and each data sequence $(e_1, \ldots, e_n) \in \mathcal{E}$, $\mathbb{P}(h \,|\, e_1, \ldots, e_n) > 1/2$ only if

- (*Razor Condition*) h is no more complex than any hypothesis in $\mathcal{H}$ that is compatible with the given data sequence $(e_1, \ldots, e_n)$;
- (*Tenacity Condition*) for any data sequence $(e_1, \ldots, e_n, \ldots, e_{n+n'})$ in $\mathcal{E}$ that extends $(e_1, \ldots, e_n)$ and is compatible with h, $\mathbb{P}(h \,|\, e_1, \ldots, e_{n+i})$ is monotonically increasing as a function of $i \in \{0, \ldots, n'\}$.

Theorem 7.5 (Ockham Stability Theorem, Bayesian Version)**.** *Let* $\mathbb{P}$ *be a probabilistic prior for a problem that converges to the truth almost everywhere. Then, condition 1 below implies condition 2 below (but the converse does not hold):*

1. $\mathbb{P}$ *converges to the truth with stability.*

2. $\mathbb{P}$ *follows Ockham's tenacious razor.*

Although the converse does not hold,[21] we might be able to formulate a stronger version of tenacity or a weaker version of stability in Bayesian terms in order to restore the equivalence between conditions *1* and *2*. But that will not be attempted here. For there is no loss in application to epistemology: to justify Ockham's razor, what is really needed is just the implication relation from the epistemic ideal *1* to the methodological principle *2*, which shows that the latter is a necessary means for achieving the former. The converse does no justificatory work. Showing that Ockham's razor achieves an epistemic ideal does not suffice to argue that one *has to* follow Ockham's razor, for there might be other means that also achieves the epistemic ideal.

All the results of this section are proved in appendix B.4.

8 Conclusion

Let me conclude by presenting a new perspective on learning theory—or perhaps an old perspective that deserves an explicit formulation and more attention. In a nutshell, it is the idea of modes of convergence to the truth that *unifies* the many branches of learning theory, together with learning-theoretic epistemology. To be more specific:

1. There are different modes of convergence to the truth, which correspond to different epistemic, truth-directed ideals for an inquirer to achieve where possible.

2. Learning theory, as a mathematical theory, is the general theory of various modes of convergence to the truth (or the correct learning target, which can be a concept, a function, or a predictive model). So learning theory has a number of branches, each studying a certain group of modes of convergence. I can count at least five branches:

[21] Here is the reason why the converse does not hold: the tenacity condition—as defined in the Bayesian framework—only requires posterior probability to remain the same or go up as data accumulate, but does not require it to go up high enough to ensure convergence to 1, let along convergence with stability.

(a) Formal/algorithmic learning theory in the tradition of Gold (1965) and Putnam (1965) studies a very strong mode of convergence, which can be called *everywhere* convergence to the truth.

(b) Computability theory studies an even stronger mode of convergence to the truth, which combines *everywhere* convergence and *perfectly monotonic* convergence.

(c) Statistical learning theory studies modes of stochastic convergence to the truth. One example is *uniform* convergence *in probability*, which is basically Valiant's (1984) PAC (probably approximately correct) learning criterion. An example familiar in asymptotic statistics is *almost sure* convergence. So statistical learning theory is construed broadly to include both the PAC learning theory and much of asymptotic statistics.

(d) Topological learning theory, the branch I develop here, studies *almost everywhere* convergence to the truth and possibly other modes of topological convergence.[22]

(e) Learning theory for Bayesian inquirers studies the Bayesian versions of all the above modes of convergence.[23] Instead of convergence of qualitative beliefs to the truth, it studies convergence of probabilistic beliefs to full certainty in the truth.[24]

3. Learning-theoretic epistemology is meant to provide justifications for local norms of inductive inference—"local" in the sense of being sensitive to the problem tackled by the inquirer.[25] But the justifications are systematically based on (i) a system of epistemic values and (ii) a general epistemic norm. So this epistemology has two parts, one evaluative and the other normative.

   (i) The evaluative part determines which mode of convergence to the truth (or which combination of modes) corresponds to a higher

[22]For example, it is possible to talk about the almost sure convergence of a learning method with respect to almost all chance distributions, taken as points in a topological space, whose topology is generated by the total variation metric between chance distributions. That can be called *almost-everywhere almost-sure* convergence to the truth, which I conjecture to be achievable for many causal discovery problems that do not make the causal faithfulness assumption. This is work in progress.

[23]For an earlier advocate of learning theory for Bayesian inquirers, see Kelly (1996: chap. 13).

[24]Where convergence to full certainty in the truth is too demanding, we should see whether it is possible to achieve at least convergence to high credence in the truth.

[25]For discussion of this kind of sensitivity, see appendix A.3, which presupposes appendix A.2.

epistemic ideal for an inquirer to achieve where possible. The present paper provides an example, as depicted in the lattice in figure 2.

(ii) The normative part operates with this general guideline: "Look for what can be achieved; achieve the best you can." The present paper provides a case study on the hard raven problem, full enumerative induction, and Ockham's razor.

It is from this unificatory perspective that I develop my solution to the Cartesian problem of induction. For tackling the problem of whether all ravens are black, the highest achievable epistemic ideal (among the ideals considered in this paper) is the combination of three modes of convergence to the truth: "almost everywhere" + "maximal domain" + "stable", as depicted in the lattice in figure 4. And a necessary means for achieving that is to follow full enumerative induction rather than counterinduction.

I believe that learning-theoretic epistemology is a promising approach to normative studies of inductive inference, but I will have to defend this philosophical claim at greater length in a future work. The goal of this paper is modest: to develop a solution to the Cartesian problem of induction, to lay its logical and mathematical foundation, to help spark the development of competing solutions for future comparison, and—last but not least—to do all this from a unificatory perspective on learning theories and learning-theoretic epistemology, unified by the idea of modes of convergence to the truth.

9 Acknowledgements

This paper would have been impossible without the years of discussions I have had with my mentor and friend Kevin T. Kelly. Many ideas in this paper were inspired by our innumerable discussions. I am indebted to Konstantin Genin, I-Sen Chen, Conor Mayo-Wilson, and Liam Bright for their very helpful comments on a precursor to the present paper. I thank Alan Hájek for discussion and for his encouragement and support during my pursuit of the present project. I am also indebted to the following people for discussions: Jason Konek, Ray Briggs, Ben Levenstein, Tina Rulli, David Copp, Alexandru Baltag, and Harvey Lederman. The present project was initiated when I taught learning theory in the Foundations Seminar on Formal Methods at the Australian National University, June 2015; I thank the participants for sparkling the present project.

10 References

Baltag, A., Gierasimczuk, N., and Smets, S. (2015) "On the Solvability of Inductive Problems: A Study in Epistemic Topology", in Ramanujam, R. (Ed.) *Proceedings of the 15th Conference on Theoretical Aspects of Rationality and Knowledge (TARK-2015)*, ACM 2015 (ILLC Prepublication Series PP-2015-13).

Carlucci, L. and Case, J. (2013) "On the Necessity of U-Shaped Learning", *Topics in Cognitive Science*, 5(1): 56-88.

Carlucci, L., Case, J., Jain, S., and Stephan, F. (2005) "Non U-Shaped Vacillatory and Team Learning", *Algorithmic Learning Theory*, 241-255, Springer Berlin Heidelberg.

Gold, E. M. (1965) "Limiting Recursion", *Journal of Symbolic Logic*, 30(1): 27-48.

Gold, E. M. (1967) "Language identification in the limit", *Information and Control*, 10(5): 447-474.

Hintikka, J. (1966), "A Two-Dimensional Continuum of Inductive Methods", in J. Hintikka and P. Suppes (eds.) *Aspects of Inductive Logic*, Amsterdam: North-Holland.

Hintikka, J., and I. Niiniluoto (1980), "An Axiomatic Foundation for the Logic of Inductive Generalization", in Jeffrey, R. (ed.) *Studies in Inductive Logic and Probability*, vol. 2. Berkeley and Los Angeles: University of California Press.

Hume, D. (1777) *Enquiries Concerning Human Understanding and Concerning the Principles of Morals*, reprinted and edited with introduction, comparative table of contents, and analytical index by Bigge, L.A. Selby (1975), MA. Third edition with text revised and notes by P. H. Nidditch. Oxford, Clarendon Press.

Kelly, K. T. (1996) *The Logic of Reliable Inquiry*, Oxford: Oxford University Press.

Kelly, K. T. (2001) "The Logic of Success", *the British Journal for the Philosophy of Science*, Special Millennium Issue 51: 639-666.

Kelly, K. T. (2004) "Learning Theory and Epistemology", in *Handbook of Epistemology*, I. Niiniluoto, M. Sintonen, and J. Smolenski, (eds.) Dordrecht: Kluwer.

Kelly, K. T. and C. Glymour (2004) "Why Probability Does Not Capture the Logic of Scientific Justification", in C. Hitchcock (ed.) *Contemporary Debates in the Philosophy of Science*, London: Blackwell.

Kelly, T. K, K. Genin, and H. Lin (2016) "Realism, Rhetoric, and Reliability", *Synthese* 193(4): 1191-1223.

Lin, H. (2017) "Hume's Dilemma and the Normative Turn", unpublished manuscript.

Osherson, D., S. Micheal, and S. Weinstein (1986) *Systems that Learn: An Introduction to Learning Theory for Cognitive and Computer Scientists*, MIT Press.

Oxtoby, J. C. (1996) *Measure and Category: A Survey of the Analogies between Topological and Measure Spaces*, 2nd Edition, Springer.

Putnam, H. (1963) "Degree of Confirmation and Inductive Logic", in Schilpp, P. A. (ed.) *The Philosophy of Rudolf Carnap*, La Salle, Ill: Open Court.

Putnam, H. (1965) "Trial and Error Predicates and a Solution to a Problem of Mostowski", *Journal of Symbolic Logic*, 30(1): 49-57.

Reichenbach, H. (1938) *Experience and Prediction: An Analysis of the Foundation and the Structure of Knowledge*, Chicago, University of Chicago Press.

Schulte, O. (1996) "Means-Ends Epistemology", *The British Journal for the Philosophy of Science* 79(1), 141147.

Spirtes, P., C. Glymour, and R. Scheines (2001) Causation, Prediction, and Search.

Vickers, S. (1989) *Topology via Logic*, Cambridge University Press, Cambridge.

A Examples, Discussions, and Open Questions

This section contains materials that might be of interest to some but not all readers.

A.1 Review of "Almost Everywhere" in Topology

Let X be a topological space (equipped with a distinguished topological base). Let π be a property that may or may not apply to points in X. The following conditions are equivalent:

1. π applies to almost all points in X—or speaking geometrically, π applies almost everywhere on X.

2. Every nonempty (basic) open set of X has a nonempty (basic) open subset on which π applies everywhere.

3. The set of points to which π applies is comprehensive enough to include a dense open subset of X.

The equivalence between conditions 1 and 2 is used in some of the proofs in this paper. Condition 3 emphasizes the fact that "being a dense subset of X" alone does not suffice for "containing almost all points in X". For example, the set of rationals is dense in the set of reals, but the former is too small to include an open subset of the latter. So the property of being a rational does not apply almost everywhere on the real line.

Sometimes topologists adopt a more lenient criterion of "almost all", according to which a property π is said to apply to almost all points in X just in case π applies to all points in $X \setminus X'$, where X' is a countable union of negligible (i.e. nowhere dense) subsets of X. This more lenient criterion is used for proving the well-known theorem that almost all continuous functions defined on the unit interval are nowhere differentiable.[26]

The present paper adopts the more stringent criterion of "almost everywhere", requiring that X' be a negligible subset of X. This choice is made for a reason that is both epistemological and exploratory. The more stringent convergence criterion corresponds to a higher epistemic ideal. I propose to see whether the higher ideal is achievable for the hard raven problem, and the answer is positive. If that were too high to be achievable, I would try to see whether the lower ideal is achievable.

A.2 Too Hard to Achieve "Almost Everywhere"

A problem can be too hard to allow for the achievement of almost everywhere convergence to the truth. There are multiple ways of generating such problems. A Cartesian skeptic has one way to offer, making use of two empirically equivalent hypotheses:

[26] See Oxtoby (1996).

Example A.1. The **very hard raven problem** poses the following joint question:

> *Are all ravens black? If not, will all the ravens observed in the future be black?*

There are three potential answers: `Yes`, `NoYes`, and `NoNo`. Note that `NoYes` is a Cartesian skeptical hypothesis, a hypothesis that is akin to (but not as terrible as) the proposition that one is a brain in a vat. Hypotheses `Yes` and `NoYes` are empirically equivalent—they are compatible with exactly the same data sequences. The present problem can be formally defined as follows:

- the hypothesis space $\mathcal{H}$ is $\{\texttt{Yes}, \texttt{NoYes}, \texttt{NoNo}\}$,
- the evidence space $\mathcal{E}$ consists of all finite sequences of `+`, `0`, and/or `-`,
- the state space $\mathcal{S}$ consists of all states in the following three categories:
  - (a) the states $(\texttt{Yes}, \vec{e})$ in which $\vec{e}$ is a `+/0` sequence,
  - (b) the states $(\texttt{NoYes}, \vec{e})$ in which $\vec{e}$ is a `+/0` sequence.
  - (c) the states $(\texttt{NoNo}, \vec{e})$ in which $\vec{e}$ is a `+/0/-` sequence that contains at least one occurrence of `-`.

Then we have this negative result:

Proposition A.2. *For the very hard raven problem, it is impossible to achieve almost everywhere convergence to the truth.*

Sketch of Proof. Suppose for *reductio* that some learning method M converges to the truth almost everywhere for the very hard raven problem. By almost everywhere convergence on the space $|\texttt{Yes}|$, there exists a `+/0` sequence $(e_1, \ldots, e_n)$ such that M converges to the truth everywhere on $|\texttt{Yes}| \cap |(e_1, \ldots, e_n)|$. By almost everywhere convergence on the space $|\texttt{NoYes}|$, $(e_1, \ldots, e_n)$ can be extended to some `+/0` sequence $(e_1, \ldots, e_n, \ldots, e'_n)$ such that M converges to the truth everywhere on $|\texttt{NoYes}| \cap |(e_1, \ldots, e_n, \ldots, e'_n)|$. Choose an (infinite) data stream $\vec{e} \in |(e_1, \ldots, e_n, \ldots, e'_n)|$. So:

$$\begin{aligned} (\texttt{Yes}, \vec{e}) &\in |\texttt{Yes}| \cap |(e_1, \ldots, e_n)|\,, \\ (\texttt{NoYes}, \vec{e}) &\in |\texttt{NoYes}| \cap |(e_1, \ldots, e_n, \ldots, e'_n)|\,. \end{aligned}$$

It follows that M converges to the truth both in state $(\texttt{Yes}, \vec{e})$ and in state $(\texttt{NoYes}, \vec{e})$. But that is impossible because those two states are empirically indistinguishable and make distinct hypotheses true. □

Due to that negative result, learning-theoretic epistemologists make no normative recommendation as to how to tackle the very hard raven problem. This raises some philosophical worries and questions, especially about the nature and purpose of learning-theoretic epistemology—see the next subsection, A.3, for a short philosophical discussion.

Here I would like to give more examples to show that, to construct a problem for which almost everywhere convergence is unachievable, it is not necessary to invoke two empirically indistinguishable *states*, and it is not sufficient to invoke two empirically equivalent *hypotheses*.

Example A.3. The **cardinality problem** poses the following question:

> *Given that the incoming data stream will be a* 0/1 *sequence, how many occurrences of* 1 *will there be? Zero, one, two, . . . , or infinite?*

This problem can be formally defined as follows:

- the hypothesis space $\mathcal{H}$ is $\{\mathtt{0}, \mathtt{1}, \mathtt{2}, \ldots, \infty\}$,
- the evidence space $\mathcal{E}$ consists of all finite sequences of 0 and/or 1,
- the state space $\mathcal{S}$ consists of all states of the following form:
  - the states $(\mathtt{n}, \vec{e})$ in which $\mathtt{n}$ is a natural number and $\vec{e}$ is a 0/1 sequence that contains exactly $\mathtt{n}$ occurrences of 1;
  - the states $(\infty, \vec{e})$ in which $\vec{e}$ is a 0/1 sequence that contains infinitely many occurrences of 1.

In the above problem, any two states are empirically distinguishable, but we still have the following negative result:

Proposition A.4. *For the cardinality problem, it is impossible to achieve almost everywhere convergence.*

Sketch of Proof. Suppose for *reductio* that there exists a learning method M that converges to the truth almost everywhere for the cardinality problem. So, in particular, M converges to the truth ∞ almost everywhere in topological space $|\infty|$. It follows that, for some finite data sequence σ_*, M converges to the truth ∞ everywhere on basic open set $|\infty| \cap |\sigma_*|$. Let $\mathtt{k}$ be the least hypothesis compatible with σ_*. By the forcing lemma (in appendix B.2), there exists a data sequence σ_k that extends σ_* and is compatible with hypothesis $\mathtt{k}$ such that $M(\sigma_k) = \mathtt{k}$. Continue applying the forcing lemma to obtain this result: for each $n \geq k$, data sequence σ_n is extended into data

sequence σ_{n+1} compatible with hypothesis `n+1` such that $M(\sigma_{n+1}) =$ `n+1`. Let σ be the infinite data sequence that extends σ_n for all natural numbers $n \geq k$. Then it is not hard to argue that M fails to converge to the truth in state $s = (\infty, \sigma)$. But this state s is in basic open set $|\infty| \cap |\sigma_*|$. Contradiction. □

The presence of two empirically equivalent hypotheses, alone, does not imply the impossibility of achieving almost everywhere convergence. Here is a counterexample:

Example A.5. The **even-vs-odd problem** poses the following question:

> *Given that the incoming data stream will be a* $0/1$ *sequence with finitely many occurrences of* 1*, will there be evenly many or oddly many?*

This problem can be formally defined as follows:

- the hypothesis space $\mathcal{H}$ is $\{\texttt{Even}, \texttt{Odd}\}$,
- the evidence space $\mathcal{E}$ consists of all finite $0/1$ sequences,
- the state space $\mathcal{S}$ consists of all states of the following form:
  - the states $(\texttt{Even}, \vec{e})$ in which $\vec{e}$ is a $0/1$ sequence that contains evenly many occurrences of 1;
  - the states $(\texttt{Odd}, \vec{e})$ in which $\vec{e}$ is a $0/1$ sequence that contains oddly many occurrence of 1.

The two competing hypotheses, `Even` and `Odd`, are empirically equivalent because no data sequence refutes one and saves the other. But we still have the following positive result:

Proposition A.6. *For the even-vs-odd problem, it is possible to achieve convergence to the truth everywhere—and, a fortiori, almost everywhere.*

Sketch of Proof. Everywhere convergence is achievable for this problem, as witnessed by this method: "Output `Even` if you have observed evenly many occurrences of 1; otherwise output `Odd`." □

A.3 Sensitivity to the Chosen Set of Hypotheses

It was remarked earlier that, for the very hard raven problem, it is even impossible to achieve almost everywhere convergence. As a consequence, learning-theoretic epistemologists have been unable to make a normative

recommendation for an inquirer tackling that problem. Let me say why they have nothing to apologize.

The very hard raven problem embodies not just the philosophical problem of responding to the inductive skeptic, but also the problem of responding to the Cartesian skeptic, as highlighted by the two empirically equivalent hypotheses put on the table:

- `Yes`: "Yes, all ravens are black."
- `NoYes`: "No, not all ravens are black; and yes, all ravens to be observed are black."

Learning-theoretic epistemology is not designed to respond to the Cartesian skeptic, and we may conjoin it with a good, independent reply to the Cartesian skeptic. To be sure, learning-theoretic epistemologists can, and should, insist that when an inquirer tackles the hard raven problem *rather than* the very hard one, she ought to be inductive and never be counterinductive, thanks to the argument and the results provided above.

So learning-theoretic epistemology typically makes a normative recommendation of this form: "*If* one tackles such and such a problem, one ought to follow a learning method having such and such properties." Such a normative recommendation is *sensitive* to, or *conditional* upon, the problem pursued by the inquirer.

In fact, learning theorists have long recognized that epistemology needs such sensitivity, for a reason that is not tied to Cartesian skepticism but can be traced back to the genesis of learning theory. The development of learning theory was historically motivated by the observation that it is mathematically impossible for us learn everything by meeting the epistemic ideal of everywhere convergence to the truth. That is, it is provably impossible to design a learning machine that is so powerful as to be capable of convergently solving the "ultimate" problem, the problem that entertains all hypotheses that human beings can understand (Putnam 1963). The cardinality problem corresponds to one such example, which makes the point even without invoking two empirically indistinguishable states or a Cartesian-like demon who is always hiding some observable items from the inquirer.

Given that it is impossible to design a learning machine for learning everything that one can understand, one has to prioritize certain things to learn. That is, in a context of inquiry, an inquirer has to identify the hypotheses whose truth values she really wants to learn, and to pursue the problem consisting of those hypotheses. When she switches to a different context of inquiry, she might need to reconsider the priority and decide to

pursue a different problem. For example, an inquirer in a philosophy seminar on Cartesian skepticism might take the very hard raven problem to be of the utmost importance and decide to pursue it. But when she returns to the laboratory, the only important problem to pursue might just be the hard raven problem, rather than the very hard one. Here I only claim that she might switch that way. As to whether she ought to switch that way or is at least epistemically permitted to switch that way, the positive answer has to be defended elsewhere.

To sum up, learning-theoretic epistemologists recognize two groups of important issues to address:

> (*Mathematical Issues*) What can be learned? Which set of hypotheses can be learned in the limit? Which problem can be solved with which combinations of modes of convergence to the truth?
>
> (*Normative Issues*) One has no alternative but to prioritize certain things to learn. But which to prioritize? Which hypotheses and which problem are the things that one really cares about, or ought to care about, in which context of inquiry, such as a laboratory or a philosophy seminar?

While the mathematical issues have driven the development of learning theory, learning-theoretic epistemologists still have a lot to do to address the normative issues.

A.4 How to Justify Ockham's Razor: One More Example

Here is one more example that illustrates the application of theorem 7.3 to justification of Ockham's razor:

Example A.7. Let x and y be real-valued variables, and suppose that y depends functionally on x. The **hard polynomial degree problem** poses the following question:

> *Given that y is a polynomial function of x, what is the degree of that polynomial function?*

This problem considers "rectangular" data on the x-y plane. A rectangular datum e_i is an open rectangle on the x-y plane that is axis-aligned and has only rational endpoints. Understand e_i to say: "The true polynomial function passes through rectangle e_i." A (finite or infinite) sequence of such rectangles is said to be compatible with a polynomial function if that polynomial function passes through all rectangles therein. The present problem can be formally defined as follows:

- the hypothesis space $\mathcal{H}$ is the set of possible polynomial degrees, $\{\mathtt{0}, \mathtt{1}, \mathtt{2}, \ldots\}$;
- the evidence space $\mathcal{E}$ is the set of finite sequences of rectangular data that are compatible with at least some polynomial function;
- the state space $\mathcal{S}$ is the set of states taking the following form:

$$(\mathtt{d}, \vec{e}\,),$$

  where $\mathtt{d}$ is a polynomial degree in $\mathcal{H}$ and $\vec{e}$ is an infinite sequence of rectangular data that is compatible with at least one polynomial function of degree $\mathtt{d}$.

A hypothesis of a lower polynomial degree is simpler:

$$\mathtt{0} \prec \mathtt{1} \prec \mathtt{2} \prec \ldots \mathtt{n} \prec \mathtt{n{+}1} \ldots$$

Here is an example of a learning method that follows Ockham's tenacious razor:

$M^*_{\mathtt{ock}}$ "Output degree $\mathtt{d}$ whenever $\mathtt{d}$ is the lowest polynomial degree that can fit the data you have (namely, whenever the data sequence you have is compatible with some polynomial function of degree $\mathtt{d}$ but with no polynomial function of any lower degree)."

This method never suspends judgment. There are other methods that also follow Ockham's tenacious razor, and they differ from the previous one by being less opinionated, willing to suspend judgment occasionally before jumping to a conclusion.

Before we apply theorem 7.3 to the hard polynomial degree problem, we have to figure out what can be achieved for that problem:

Proposition A.8. *For the hard polynomial degree problem, it is possible to achieve convergence to the truth almost everywhere with stability on a maximal domain.*

Sketch of Proof. The existential claim is witnessed by the method $M^*_{\mathtt{ock}}$ defined above, and can be proved in a way that mimics the proof of the existential claim of theorem 6.5. □

For the hard polynomial problem, is it possible achieve a higher mode, such as one that implies everywhere convergence or perfectly monotonic convergence? The answer is negative. To secure at least almost everywhere convergence, perfectly monotonic convergence is impossible because the problem in question is essentially an inductive problem, with a hypothesis that goes beyond the logical consequences of data. Everywhere convergence is unachievable because the state space is liberal enough to allow for two empirically indistinguishable states that make distinct hypotheses true. For example, consider a data stream $\vec{e}$ being so unspecific that it is compatible with some polynomial function of degree `d` and also with some other polynomial function of degree `d+1`. So there are (at least) two empirically indistinguishable states that make distinct hypotheses true, namely $(\texttt{d}, \vec{e})$ and $(\texttt{d+1}, \vec{e})$. Therefore, this problem does not have a learning method that converges to the truth everywhere. This is why it is called a "hard" problem, which suggests that we can obtain an "easy" version if we are willing to make a sufficiently strong presupposition to constrain the state space.[27]

So here is what we have: When tackling the hard polynomial problem, an inquirer ought to achieve the highest achievable epistemic ideal among those in the lattice in figure 4, and that is the joint mode of convergence to the truth "almost everywhere" + "stable" + "maximal". A necessary means for achieving that is to follow Ockham's tenacious razor, thanks to theorem 7.3. This is why an inquirer tackling the hard polynomial degree problem ought to follow Ockham's tenacious razor—or so I submit.

You might wonder whether Ockham's razor can be justified in a simpler way than I just did. Suppose that we have proved that a problem $\mathcal{P}$ is easy enough to make it possible to achieve at least "almost everywhere" + "stable", setting aside all other modes of convergence. Then it is tempting to quickly conclude that $\mathcal{P}$ ought to be tackled with a method that achieves "almost everywhere" + "stable", and then immediately apply theorem 7.3 to conclude that $\mathcal{P}$ ought to be tackled with a method that follows Ockham's tenacious razor. It is tempting to do all this and rush to justify Ockham's razor, without considering higher epistemic ideals, such as one that adds convergence on a maximal domain. Is it OK to rush to justify Ockham's razor that way?

[27] To be more specific, the **easy polynomial degree problem** is the same as the hard one except that it has a more constrained state space, in which each state $(\texttt{d}, \vec{e})$ is required to be such that its data stream $\vec{e}$ is compatible with exactly one polynomial function and that unique polynomial function has degree `d`. For this problem, everywhere convergence to the truth is achievable.

The answer is negative, and it is important to know why:[28] Some problems involve a trade-off between "stable" and "maximal" that forces the inquirer to sacrifice one in order to secure the other—see appendix A.5 for an example. In that case, it may not be immediately clear as to whether the inquirer should opt for the package "almost everywhere" + "stable" or side with "almost everywhere" + "maximal". Only the former requires following Ockham's tenacious razor; the latter does not.

A.5 Trade-off Between "Stable" and "Maximal"

There are situations in which the inquirer is, in a sense, forced to make a trade-off between two desirable modes of convergence, such as stability and maximality. Here is an example:

Example A.9. The **bounded even-vs-odd problem** poses the following question:[29]

> *Given that the incoming data stream will be a* $0/1$ *sequence with at most two occurrences of* 1*, will there be evenly or oddly many occurrences of* 1*?*

This problem can be formally defined as follows:

- the hypothesis space $\mathcal{H}$ is $\{\texttt{Even}, \texttt{Odd}\}$,
- the evidence space $\mathcal{E}$ consists of all finite $0/1$ sequences that have at most two occurrences of 1,
- the state space $\mathcal{S}$ consists of all states of the following form:
  - the states $(\texttt{Even}, \vec{e})$ in which $\vec{e}$ is a $0/1$ sequence that contains exactly zero or two occurrences of 1;
  - the states $(\texttt{Odd}, \vec{e})$ in which $\vec{e}$ is a $0/1$ sequence that contains exactly one occurrence of 1.

In this problem, `Odd` is simpler than `Even`.

Then we have the following trade-off result:

[28] The point made in this paragraph is a supplement to the way that I and co-authors justify Ockham's razor in Kelly, Genin, and Lin (2016). In that earlier work, the achievability of "everywhere convergence" plus "cycle-free" (a variant of stability) is taken to be sufficient for justifying the use of Ockham's razor.

[29] I thank Konstantin Genin for bringing this problem to my attention.

Proposition A.10. *Consider the following two modes of convergence:*

(1) convergence to the truth with stability,

(2) convergence to the truth on a maximal domain.

For the bounded even-vs-odd problem, each of those modes is achievable, but they are not jointly achievable.

Sketch of Proof. Everywhere convergence is achievable for this problem, as witness by this method: "Output `Even` if you have observed evenly many occurrences of 1; otherwise output `Odd`." As a consequence, convergence on a maximal domain is equivalent to everywhere convergence. Suppose that M converges to the truth on a maximal domain. So M converges to the truth everywhere and, hence, in the states $(\texttt{Even}, \vec{e})$ with $\vec{e}$ containing no occurrence of 1. But convergence in those states can be argued to violate the razor condition in Ockham's tenacious razor. Then, by the Ockham stability theorem 7.3, M fails to achieve stable convergence. □

Fortunately, for the bounded even-vs-odd problem, the inquirer is forced to choose between stability and maximality only in a weak sense: she could have easily fine-grain the hypothesis space and ask instead: "How many occurrences of 1 will there be? Zero, one, or two?" Call this fine-grained problem the **zero-vs-one-vs-two problem**. For this problem, stability and maximality are jointly achievable together with everywhere convergence. This observation leads to the following open questions:

> (*Open Questions*) Are there problems for which it is possible to achieve stability, possible to achieve maximality, impossible to achieve both simultaneously, and even impossible to achieve both no matter how the hypothesis space is fine-grained? If there are such problems, which mode of convergence should one sacrifice in exchange for the other?

I tend to think that, in such an unfortunate problem, stability should be sacrificed in exchange for maximality—in general, the consideration about "where to converge" should be prioritized over the consideration about "how to converge". But this normative claim will have to be defended in another paper.

B Proofs

B.1 The Idea of Proof of Theorem 6.5

Theorem 6.5 has three parts. The existential claim and the universal claim about "be enumeratively inductive" are two easier parts. The crucial part is the universal claim about "never be counterinductive". The reason why it is crucial is two-fold: first, it is a very instructive special case of the $1 \Rightarrow 2$ side of the Ockham stability theorem 7.3; second, it is a lemma for proving the part about "be enumeratively inductive". So let me separate the crucial part for a closer examination:

Proposition B.1 (Never Be Counterinductive)**.** *Let M be a learning method for the hard raven problem that converges to the truth almost everywhere with stability. Then M is never counterinductive.*

Note that we do not need convergence on a maximal domain here. It will be convenient to have the following concept:

Definition B.2. Given a problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$, a data sequence $(e_1, \ldots, e_n) \in \mathcal{E}$ is said to be **compatible** with a hypothesis $h \in \mathcal{H}$ if the propositions they express have a nonempty overlap, namely:

$$|h| \cap |(e_1, \ldots, e_n)| \quad \neq \quad \varnothing \,,$$

which also means that $(e_1, \ldots, e_n)$ can be extended into a data stream $\vec{e}$ such that $(h, \vec{e})$ is a state in $\mathcal{S}$.

The proof of the above proposition proceeds as follows. Let M be a learning method for the hard raven problem that converges to the truth almost everywhere. Suppose that M is sometimes counterinductive, namely, for some `+/0` sequence $(e_1, \ldots, e_n)$, we have that:

$$M(e_1, \ldots, e_n) \quad = \quad \texttt{No}\,. \tag{1}$$

It suffices to show that M fails to converge to the truth with stability. Since $(e_1, \ldots, e_n)$ is a `+/0` sequence, it is compatible with `Yes`. To summarize, we have had:

- M converges to the truth almost everywhere.
- $(e_1, \ldots, e_n)$ is compatible with `Yes`.

Given these two conditions, we can apply the so-called *forcing lemma* (to be stated soon) in order to "force" M to output `Yes` by extending $(e_1, \ldots, e_n)$ into a certain `+/0` sequence $(e_1, \ldots, e_n, \ldots, e_{n'})$ such that:

$$M(e_1, \ldots, e_n, \ldots, e_{n'}) \quad = \quad \texttt{Yes}\,. \tag{2}$$

Now, choose a state s such that:

$$s \quad \in \quad |\texttt{No}| \cap |(e_1, \ldots, e_n, \ldots, e_{n'})|\,. \tag{3}$$

We can always make this choice because every data sequence is compatible with `No`. By (1)-(3), we have: given the earlier stage n in state s, M outputs the truth `No` but fails to have converged to the truth. So M does not converge to the truth with stability. This finishes the proof of the part "never be counterinductive" in theorem 6.5—as soon as the forcing lemma is stated and established.

Let me state the forcing lemma here, remark on its importance, and leave its proof to appendix B.2:

Lemma (**Forcing Lemma**). *Let $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ be an arbitrary problem. Suppose that M is a learning method for it that converges to the truth almost everywhere, and that $(e_1, \ldots, e_n) \in \mathcal{E}$ is compatible with $h \in \mathcal{H}$. Then the above data sequence can be extended into a data sequence $(e_1, \ldots, e_n, \ldots, e_{n'}) \in \mathcal{E}$ such that:*

1. *$(e_1, \ldots, e_n, \ldots, e_{n'})$ is still compatible with h,*
2. *$M(e_1, \ldots, e_n, \ldots, e_{n'}) = h$.*

This lemma has a weaker and classic version, which deletes 'almost' and applies only to learning methods that converge to the truth everywhere. The weaker version has played an important role in proving many results in formal learning theory. Now, with the forcing lemma strengthened to cover almost everywhere convergence, many old proof techniques can be carried over to the learning theory developed here.[30] In fact, most of the results of this paper—positive or negative—are proved with the help of the forcing lemma.

Now let me turn to sketching the proof of the part "be enumeratively inductive". Suppose that learning method M converges to the truth almost

[30] In case you are interested: the forcing lemma can even be strengthened further to apply to *convergence to the truth on a dense set*. I wonder whether such a weak convergence criterion is interesting epistemologically, but I will not address this question here.

everywhere with stability (and we are going to suppose that M converges on a maximal domain only when we really need to). Then, by the preceding result, M is never counterinductive, and hence it fails to converge to the truth in every Cartesian scenario of induction, say $(\texttt{No}, \vec{e})$, where $\vec{e}$ contains no occurrence of a nonblack raven. This failure of convergence in the Cartesian state $(\texttt{No}, \vec{e})$ opens the possibility for M to converge to the truth in its normal counterpart $(\texttt{Yes}, \vec{e})$. To turn this possibility into a reality, it suffices to invoke the last supposition of the theorem, that M converges to the truth on a maximal domain. It can be shown that, in order for M to converge to the truth on a maximal domain, the domain of convergence of M has to be so comprehensive that it contains all states that make hypothesis `Yes` true, which implies that M is enumeratively inductive.

As to the proof of the existential claim, it is almost routine to verify that it is witnessed by the method M^* constructed above, which says: "Output hypothesis `No` if you have observed a nonblack raven (`-`); otherwise output `Yes`."

This finishes the proof sketch of theorem 6.5.

B.2 Proof of the Forcing Lemma

The forcing lemma has two versions, one for (qualitative) learning methods and the other for probabilistic priors.

Lemma B.3 (**Forcing Lemma, Qualitative Version**). *Let $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ be an arbitrary problem. Suppose that M is a learning method for it that converges to the truth almost everywhere, and that $(e_1, \ldots, e_n) \in \mathcal{E}$ is compatible with $h \in \mathcal{H}$. Then the above data sequence can be extended into a data sequence $(e_1, \ldots, e_n, \ldots, e_{n'}) \in \mathcal{E}$ such that:*

1. *$(e_1, \ldots, e_n, \ldots, e_{n'})$ is still compatible with h,*
2. *$M(e_1, \ldots, e_n, \ldots, e_{n'}) = h$.*

Proof. Suppose that $(e_1, \ldots, e_n)$ is compatible with h. Namely,

$$|h| \cap |(e_1, \ldots, e_n)|$$

is a nonempty basic open set of topological space $|h|$. We are going to make use of the following characterization of "almost everywhere" in general topology:

> A property applies almost everywhere on a topological space (with a distinguished topological base) if, and only if, each nonempty (basic) open set U has a nonempty (basic) open subset U' such that the property applies everywhere on U'.

So, by the "only if" side and the hypothesis that M converges to the truth almost everywhere, it follows that $|h| \cap |(e_1, \ldots, e_n)|$ has a nonempty basic open subset:

$$|h| \cap |(e_1, \ldots, e_n, \ldots, e_k)|$$

on which M converges to the truth everywhere. Now, within this nonempty set, choose an arbitrary state $(h, \vec{e})$. So, in that state, M converges to the truth. Then there exists a positive integer $n' \geq k$ such that M outputs the truth h given the n'-th stage along data stream $\vec{e}$. That is:

$$M(e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'}) = h\,.$$

It is not hard to see that the input is still compatible with h. □

Lemma B.4 (**Forcing Lemma, Bayesian Version**). *Let $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ be an arbitrary problem. Suppose that $\mathbb{P}$ is a probabilistic prior for it that converges to the truth almost everywhere, and that $(e_1, \ldots, e_n) \in \mathcal{E}$ is compatible with $h \in \mathcal{H}$. Then the above data sequence can be extended into a data sequence $(e_1, \ldots, e_n, \ldots, e_{n'}) \in \mathcal{E}$ such that:*

1. *$(e_1, \ldots, e_n, \ldots, e_{n'})$ is still compatible with h,*
2. *$\mathbb{P}(h \,|\, e_1, \ldots, e_n, \ldots, e_{n'}) > 1/2$.*

Proof. Copy the proof of the qualitative version of the forcing lemma, and paste it here. Now, replace the only occurrence of $M(e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'}) = h$ by $\mathbb{P}(h \,|\, e_1, \ldots, e_n, \ldots, e_k, \ldots, e_{n'}) > 1/2$. As the last step, replace each occurrence of M by $\mathbb{P}$. □

B.3 Proofs for Section 6: Enumerative Induction

The proofs presented in this section rely on the forcing lemma proved in section B.2.

Proof of Proposition 6.3. Suppose that a learning method M for the hard raven problem achieves perfectly monotonic convergence. Then M is a "non-inductive" method in that it never outputs `Yes`, so it fails to converge to the truth in every state in $|\texttt{Yes}|$. So M fails to converge to the truth almost everywhere in the topological space $|\texttt{Yes}|$. So M fails to converge to the truth almost everywhere. □

Proof of Theorem 6.5. To establish the existential claim, it suffices to show that it is witnessed by the learning method M^* we have discussed: "Output hypothesis `No` if you have observed a nonblack raven (`-`); otherwise output `Yes`." Proposition 5.2 has established that M^* converges to the truth almost everywhere. It is routine to verify that M^* converges to the truth with stability. To show that M^* has a maximal domain of convergence, note that it converges to the truth in all states in $|\texttt{Yes}|$ and in all states in $|\texttt{No}|$ except the Cartesian scenarios of induction. No learning method converges to the truth in strictly more states. For to do so is is to converge to the truth both in a normal state $(\texttt{Yes}, \vec{e})$ and its Cartesian counterpart $(\texttt{No}, \vec{e})$, which is impossible. This establishes maximal convergence for M^*, and finishes the proof of the existential claim.

To establish the first part of the universal claim "never be counterinductive", it suffices to invoke the proof that has already been detailed in appendix B.1, or simply to note that it is a corollary of theorem 7.3. Note that the proof relies only on the two modes of convergence to the truth, "almost everywhere" and "stable". To establish the second part "be enumeratively inductive", suppose that M is a learning method for the hard raven problem that converges to the truth on a maximal domain, and that M is never counterinductive (making use of the first part). It suffices to show that M is enumeratively inductive, as follows. Since M is never counterinductive, M fails to converge to the truth in each Cartesian scenario of induction. So the domain of convergence of M is included in that of M^*, which has been proved to be a maximal domain of convergence. But M converges on a maximal domain, so M must have the same domain of convergence as M^*. Then M converges to the truth in every state $(\texttt{Yes}, \vec{e})$ contained in $|\texttt{Yes}|$. It follows that M is enumeratively inductive. □

Proof of Theorem 6.8. The proof of the existential claim is the crux, so let me first present the proof of the easy part, the universal claim. Just copy the proof of the universal claim in theorem 6.5 (i.e. the preceding paragraph), paste it here, and apply the following replacements: First, replace the reference to theorem 7.3 by the reference to its Bayesian counterpart, theorem 7.5. Second, replace 'learning method' by 'probabilistic prior'. Third, replace M by $\mathbb{P}$. As the last step, replace M^* by $\mathbb{P}^*$, which is the probabilistic prior to be constructed below for proving the existential claim.

To prove the existential claim, construct a witness $\mathbb{P}^*$ as a linear combination of two other probabilistic priors:

$$\mathbb{P}^* \;=\; \frac{1}{2}\,\mathbb{P}_0 \,+\, \frac{1}{2}\,\mathbb{P}_1\,,$$

where $\mathbb{P}_0$ and $\mathbb{P}_1$ are defined as follows. Let $\mathbb{P}_0$ be the probability function generated by, so to speak, assuming that `Yes` is true and observations of `+`, `0`, `-` are i.i.d. (independent and identically distributed) random variables, with equal probability 1/2 for `+` and for `0`, and with probability 0 for `-`. So:

$$\begin{aligned} \mathbb{P}_0(\texttt{Yes}) &= 1\,. \\ \mathbb{P}_0(e_1,\ldots,e_n) &= \begin{cases} \left(\frac{1}{2}\right)^n & \text{if } e_i \neq \texttt{-} \text{ for each } i \leq n, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Similarly, let $\mathbb{P}_1$ be the probability function generated by, so to speak, assuming that `No` is true and observations of `+`, `0`, `-` are i.i.d. random variables with equal probability 1/3 for `+`, for `0`, and for `-`. So:

$$\begin{aligned} \mathbb{P}_1(\texttt{No}) &= 1\,. \\ \mathbb{P}_1(e_1,\ldots,e_n) &= \left(\frac{1}{3}\right)^n\,. \end{aligned}$$

It suffices to show that $\mathbb{P}^*$, defined as the half-and-half mixture of $\mathbb{P}_0$ and $\mathbb{P}_1$, converges to the truth with all the three modes mentioned in the existential claim. By the construction of $\mathbb{P}^*$, we have:

$$\begin{aligned} \mathbb{P}^*(\texttt{Yes}) &= 1/2\,. \\ \mathbb{P}^*(\texttt{No}) &= 1/2\,. \\ \mathbb{P}^*(e_1,\ldots,e_n \,|\, \texttt{Yes}) &= \mathbb{P}_0(e_1,\ldots,e_n) = \begin{cases} \left(\frac{1}{2}\right)^n & \text{if } e_i \neq \texttt{-} \text{ for each } i \leq n, \\ 0 & \text{otherwise.} \end{cases} \\ \mathbb{P}^*(e_1,\ldots,e_n \,|\, \texttt{No}) &= \mathbb{P}_1(e_1,\ldots,e_n) = \left(\frac{1}{3}\right)^n\,. \end{aligned}$$

Now, calculate conditional probability $\mathbb{P}^*(\texttt{Yes} \,|\, e_1,\ldots,e_n)$ by plugging the above probability values into the following instance of Bayes' theorem:

$$\begin{aligned} &\mathbb{P}^*(\texttt{Yes} \,|\, e_1,\ldots,e_n) \\ = \;& \frac{\mathbb{P}^*(e_1,\ldots,e_n \,|\, \texttt{Yes})\,\mathbb{P}^*(\texttt{Yes})}{\mathbb{P}^*(e_1,\ldots,e_n \,|\, \texttt{Yes})\,\mathbb{P}^*(\texttt{Yes}) + \mathbb{P}^*(e_1,\ldots,e_n \,|\, \texttt{No})\,\mathbb{P}^*(\texttt{No})}\,. \end{aligned}$$

Then we have:

$$\mathbb{P}^*(\texttt{Yes} \,|\, e_1,\ldots,e_n) = \begin{cases} \frac{1}{1+(2/3)^n} & \text{if } e_i \neq \texttt{-} \text{ for each } i \leq n, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

$$\mathbb{P}^*(\texttt{No} \,|\, e_1,\ldots,e_n) = 1 - \mathbb{P}^*(\texttt{Yes} \,|\, e_1,\ldots,e_n)\,. \tag{5}$$

$$\lim_{n\to\infty} \frac{1}{1+(2/3)^n} = 1\,. \tag{6}$$

By the above three equations, (4)-(6), it follows that $\mathbb{P}^*$ converges to the truth in all states in $|\texttt{Yes}|$, and in all states in $|\texttt{No}|$ except the Cartesian scenarios of induction. But recall that, by lemma 4.3, the set of the Cartesian scenarios of induction is negligible within the topological space $|\texttt{No}|$. So $\mathbb{P}^*$ converges to the truth almost everywhere. Argue for stable convergence by considering the following two cases.

Case (i): suppose that $\mathbb{P}^*(\texttt{Yes} \,|\, e_1, \ldots, e_n) > 1/2$ and state $s \in |\texttt{Yes}| \cap |(e_1, \ldots, e_n)|$. So $s = (\texttt{Yes}, \vec{e})$, where $\vec{e}$ is an infinite +/0 sequence. By equation (4) and the fact that $1/\big(1 + (2/3)^n\big)$ is a monotonically increasing function of n that converges to 1 as $n \to \infty$, we have: $\mathbb{P}^*$ converges to the truth in s and $\mathbb{P}^*(\texttt{Yes} \,|\, e_1, \ldots, e_n) \geq \mathbb{P}^*(\texttt{Yes} \,|\, e_1, \ldots, e_n, \ldots, e_{n'})$ for any $n' \geq n$.

Case (ii): suppose that $\mathbb{P}^*(\texttt{No} \,|\, e_1, \ldots, e_n) > 1/2$ and state $s \in |\texttt{No}| \cap |(e_1, \ldots, e_n)|$. So, by equations (4) and (5), $(e_1, \ldots, e_n)$ contains an occurrence of -. Then $s = (\texttt{No}, \vec{e})$, where $e_i = \texttt{-}$ for some $i \leq n$. So $\mathbb{P}^*(\texttt{No} \,|\, e_1, \ldots, e_n, \ldots, e_{n'}) = 1$ for all $n' \geq n$. So we have: $\mathbb{P}^*$ converges to the truth in s and $\mathbb{P}^*(\texttt{No} \,|\, e_1, \ldots, e_n) \geq \mathbb{P}^*(\texttt{No} \,|\, e_1, \ldots, e_n, \ldots, e_{n'})$ for any $n' \geq n$.

By the results of cases (i) and (ii), $\mathbb{P}^*$ converges to the truth with stability. To establish maximal domain of convergence, suppose for *reductio* that there is a probability function $\mathbb{P}$ that has a strictly more inclusive domain of convergence than $\mathbb{P}^*$ does. But $\mathbb{P}^*$ converges to the truth in all states except the Cartesian scenarios of induction. So $\mathbb{P}$ must converge to the truth in a certain normal state and in its Cartesian counterpart, which is impossible. So $\mathbb{P}^*$ converges to the truth on a maximal domain. □

B.4 Proofs for Section 7: Ockham's Razor

The proofs presented in this section rely on the forcing lemma proved in section B.2.

Proof of Theorem 7.3. Suppose that learning method M converges to the truth almost everywhere for problem $\big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$. To prove the side $1 \Rightarrow 2$ by contraposition, suppose that M does not follow Ockham's tenacious razor. It suffices to show that M does not converge to the truth with stability. Discuss the following two exhaustive cases.

Case (i): Suppose that M violates the tenacity condition. That is, $M(e_1, \ldots, e_n) = h$ and $M(e_1, \ldots, e_n, \ldots, e_{n'}) \neq h$, where $(e_1, \ldots, e_n, \ldots, e_{n'})$ is compatible with h. By that compatibility, choose a state s in the nonempty set $|h| \cap (e_1, \ldots, e_n, \ldots, e_{n'})$. It follows that, given stage n in state s, M out-

puts the truth h but it has not converged to the truth. So M fails to converge to the truth with stability.

Case (ii): Suppose that M violates Ockham's razor. Then $M(e_1, \ldots, e_n) = h$, for some $(e_1, \ldots, e_n) \in \mathcal{E}$ and some $h \in \mathcal{H}$, but there exists another hypothesis $h' \in \mathcal{H}$ that is compatible with $(e_1, \ldots, e_n)$ and simpler than h. Since $(e_1, \ldots, e_n)$ is compatible with h', by the forcing lemma B.3 and the almost everywhere convergence of M, we have: $(e_1, \ldots, e_n)$ can be extended into a data sequence $(e_1, \ldots, e_n, \ldots, e_{n'}) \in \mathcal{E}$ such that, first, $M(e_1, \ldots, e_n, \ldots, e_{n'}) = h'$ and, second, $(e_1, \ldots, e_n, \ldots, e_{n'})$ is compatible with h'. Since $(e_1, \ldots, e_n, \ldots, e_{n'})$ is compatible with h' and since h' is simpler than h, it follows that $(e_1, \ldots, e_n, \ldots, e_{n'})$ is also compatible with h. By that compatibility, choose a state $s \in |h| \cap |(e_1, \ldots, e_n, \ldots, e_{n'})|$. So, given the earlier stage n in state s, M outputs the truth, h, but has not converged to the truth, for $M(e_1, \ldots, e_n, \ldots, e_{n'}) \neq h$. It follows that M fails to converge to the truth with stability.

To prove the side $2 \Rightarrow 1$, it suffices to show that the tenacity condition (alone) implies convergence to the truth with stability. Suppose that M has the tenacity property, and that M outputs the truth h given stage n in state $s = (h, \vec{e})$. It suffices to show that M has converged to the truth given the same stage n in the same state s. Note that, for any natural number i, the data sequence $(e_1, \ldots, e_{n+i})$ extends $(e_1, \ldots, e_n)$ and is compatible with h. So, by the tenacity condition, $M(e_1, \ldots, e_{n+i}) = h$, for all $i \geq 0$. It follows that M has converged to the truth, h, given stage n in state s. □

Proof of Theorem 7.5. Copy the proof of the $1 \Rightarrow 2$ side of theorem 7.3, and paste it here. Then apply the following replacements. For case (i):

- First, replace $M(e_1, \ldots, e_n) = h$ and $M(e_1, \ldots, e_n, \ldots, e_{n'}) \neq h$ by $1/2 < \mathbb{P}(h \,|\, e_1, \ldots, e_n) > \mathbb{P}(h \,|\, e_1, \ldots, e_n, \ldots, e_{n'})$.
- And then replace each occurrence of M by $\mathbb{P}$.

For case (ii):

- First, replace $M(e_1, \ldots, e_n) = h$ by $\mathbb{P}(h \,|\, e_1, \ldots, e_n) > 1/2$.
- Then replace $M(e_1, \ldots, e_n, \ldots, e_{n'}) = h'$ by $\mathbb{P}(h' \,|\, e_1, \ldots, e_n, \ldots, e_{n'}) > 1/2$.
- Then replace $M(e_1, \ldots, e_n, \ldots, e_{n'}) \neq h$ by $\mathbb{P}(h \,|\, e_1, \ldots, e_n, \ldots, e_{n'}) \not> 1/2$

- And, as the last step, replace each occurrence of M by $\mathbb{P}$.

This finishes the proof of the $1 \Rightarrow 2$ side.

To prove that the converse $2 \Rightarrow 1$ does *not* hold, construct a problem $\mathcal{P} = \big(\mathcal{H}, \mathcal{E}, \mathcal{S}\big)$ as follows. Consider only the following data streams, where m and n are arbitrary natural numbers:

$$\begin{aligned} s_\omega &= 0^\omega\,. \\ s_m &= 0^m 1^\omega\,. \\ s_{mn} &= 0^m 1^n 2^\omega\,. \end{aligned}$$

Their initial segments form the evidence space $\mathcal{E}$. The hypothesis space $\mathcal{H}$ consists of

- h = "The actual sequence will not end with occurrences of 2."
- h' = "It will."

The state space $\mathcal{S}$ consists of (h, s_ω), (h, s_m), and (h', s_{mn}), for all natural numbers m and n. Construct a countably additive probability function $\mathbb{P}$ that assigns the following probabilities to singletons of states:

$$\begin{aligned} \mathbb{P}\{s_\omega\} &= 0\,. \\ \mathbb{P}\{s_m\} &= \left(\frac{1}{2}\right)^{m+1} \times 60\%\,. \\ \mathbb{P}\{s_{mn}\} &= \left(\frac{1}{2}\right)^{m+1} \times 40\% \times \left(\frac{1}{2}\right)^{n+1}\,. \end{aligned}$$

Those assignments of probabilities are designed to ensure the following:

$$\begin{aligned} \mathbb{P}\{s_m\} &= \left(\frac{1}{2}\right)^{m+1} \times 60\%\,. \\ \mathbb{P}\{s_{m0}, s_{m1}, s_{m2}, \ldots\} &= \left(\frac{1}{2}\right)^{m+1} \times 40\%\,. \\ \mathbb{P}\{s_m, s_{m0}, s_{m1}, s_{m2}, \ldots\} &= \left(\frac{1}{2}\right)^{m+1}\,. \\ \sum_{m=0}^{\infty} \mathbb{P}\{s_m, s_{m0}, s_{m1}, s_{m2}, \ldots\} &= 1\,. \end{aligned}$$

It follows that, for each natural number m, we have:

$$\mathbb{P}(h \,|\, 0^m) = 60\%\,.$$

So $\mathbb{P}$ fails to converge to the truth H in state 0^ω. It is routine to verify that $\mathbb{P}$ converges to the truth in all the other states and, hence, does so almost everywhere. It is also routine to verify that $\mathbb{P}$ follows Ockham's tenacious razor. In state (h, s_ω) and given information 0^m, $\mathbb{P}$ assigns a probability greater than $1/2$ (namely 60%) to the truth (namely h) but fails to have started to stably converge to the truth, because it even fails to converge to the truth in that state. So $\mathbb{P}$ fails to converge to the truth with stability. This finishes the proof. □

How and When Chances Guide Credences via the Principal Principle

Ilho Park (Chonbuk National University)

Abstract

This paper is intended to examine the relationship between the Principal Principle and Conditionalization. For this purpose, I will first formulate several versions of the Principal Principle and Conditionalization in Section 2. In regard to the relationship between the two norms in question, I will give both good and bad news in this paper. The good news, which will be given in Section 3, is that the Principal Principle and Conditionalization are complementary in a particular sense. This news can be regarded as a result that criticizes and supplements some existing works about the relationship between the norms. The bad news, which will be delivered in Section 4, is that our credences are sensitive to when objective chances guide our rational credences. This news reveals another kind of non-commutativity of Bayesian belief updating.

1 Introduction

Many philosophers of probability think that there are some epistemic norms governing how objective chances ought to guide our credences. The most famous chance-credence norm is what is called 'Principal Principle', which is formulated and named by David Lewis (1980).[1] On the other hand, the philosophers also think that there are other epistemic norms, which govern how our credence should evolve after experience. One of the most well-known norms is Conditionalization. Roughly speaking, this norm requires us to update our credences by conditionalizing on experience that directly changes some parts of our credal state. It is very desirable that the former should not rule out the latter in the sense that rational agents could obey such norms at the same time. Moreover, the credences obeying

[1] There are other kinds of chance-credence norm, except Lewis's Principal Principle, that relate objective chances to our rational credences. Lewis, Hall, and Thau's *New Principle*, and Ismael's *General Recipe* are representative of such norms. See Lewis (1994), Ismael (2008), Hall (1994), and Thau (1994).

these norms should be insensitive to any irrelevant factor to our epistemic rationality.

This paper is intended to examine whether such desiderata are satisfied by the Principal Principle and Conditionalization. In this regard, I will deliver both good and bad news in this paper. The good news, which will be given in Section 3, is that the Principal Principle and Conditionalization are *complementary* in a particular sense. The bad news, which will be given in Section 4, is that our credences are *sensitive* to when objective chances guide our rational credences. For this purpose, I will first formulate several versions of the Principal Principle and Conditionalization in Section 2. With such versions at hand, I will examine the aforementioned relationship between the two kinds of epistemic norm.

Before I proceed further, some notes about assumptions, terminologies, and notations are in order. First, I will assume that credences and chances are all probabilities. That is, credence and chance functions are assumed to satisfy the standard probability axioms. In this regard, I will also assume that those credence and chance functions are defined on the same outcome space Ω, which is a set of all possible worlds. For the sake of the mathematical simplicity, it will be assumed that Ω has only a finitely many possible worlds. 'w', 'w_1', 'w^*', and so forth will be used to refer to possible worlds. Propositions in this paper are defined as a subset of Ω. The unit set $\{w\}$ is a proposition, but I will use interchangeably '$\{w\}$' and w if there is no danger of confusion. When A is a proposition, '$\neg A$' refers to the complement of A with respect to Ω—namely, the negation of A. Similarly, when A and B are propositions, the conjunction 'AB' and the disjunction '$A \vee B$', respectively, refer to the intersection and union of A and B. Let $\mathcal{F}$ refer to the algebra on Ω—that is, the set of subsets of Ω. $\mathcal{F}$ has Ω as a member, and is closed under complementation and union. Lastly, '$\mathbb{E}$' refers to a partition $\{E_1, \cdots, E_n\}$ whose members are mutually exclusive and collectively exhaustive. Similarly, '$\mathbb{F}$' also refers to a partition $\{F_1, \cdots, F_m\}$. Note that $\mathbb{E} \subseteq \mathcal{F}$ and $\mathbb{F} \subseteq \mathcal{F}$.

Throughout, I will assume that each possible world has only one ur-chance function. Here, 'ur-chance function at a particular world' refers to a chance function that does not reflect any history of that world. I will not rule out a metaphysical possibility that two different worlds have the same ur-chance function. Thus, there are only finitely many ur-chance functions. Following Hall (2004) and Pettigrew (2014), I will formulate chance-credence norms by means of ur-chance functions. 'ch', 'ch_1', 'ch^*', and so on will be used to refer to chance functions. 'U_{ch}' stands for the proposition that ch is the ur-chance function. U_{ch} is true at all and only possible worlds whose ur-chance function is ch. Thus, U_{ch} is equivalent to

the disjunction of all possible worlds whose ur-chance function is ch. It is noteworthy that the set of U_{ch}s is a partition, which is a subset of $\mathcal{F}$. This partition will be denoted by '$\mathbb{U}$'.

It is also assumed that the chance function at a time t and world w is obtained by *conditionalizing* the ur-chance function at w on the history of w up to t. That is, it is assumed that $ch^w_t(\cdot) = ch^w(\cdot|H_{tw})$, where ch^w_t is the chance function at a time t and world w, ch^w is the ur-chance function at w, and H_{tw} is a proposition describing the complete and accurate history of w up to t. Admittedly, there may be a chance function at w that reflects only a part of the history of w up to t. For example, it entirely makes sense that there is a chance function at w that reflects E that is a subset of H_{tw}. In this regard, I assume that such a chance function is obtained by conditionalizing the ur-chance function at w on E—that is, $ch^w_E(\cdot) = ch^w(\cdot|E)$, where ch^w is the ur-chance function at w, ch^w_E is the chance function at w that reflects only E.

In this paper, I will consider only one chance-credence norm—that is, Lewis's Principal Principle. As is well known, its original version suffers from the so-called 'Big Bad Bug', and the some philosophers—e.g., Lewis (1994), Thau (1994), and Hall (1994)—provided a modified version of the original principle, which is often called 'New Principle'. Roughly speaking, the bug in question is that Lewis's original Principal Principle is incompatible with his Humean Supervenience. This issue is very interesting, but, in this paper, I will not consider the bug and the New Principle. This is because my discussions have little bearing on Humean Supervenience, and my results remain the same even if we regard the New Principle as a more plausible chance-credence norm.

Let me finish this section with introducing another notation. I will use 'C' to denote a coherent *initial* credence function. Here, the modifier 'initial' is intended to express that the credence function in question undergoes no course of experience, and so has no evidence. I assume that the initial credence functions are *regular* in the sense that the functions assign a non-zero value to any non-empty proposition. Additionally, 'C_E' refers to a credence function that has total evidence E. I will assume, following many philosophers of probability, that if a credence function has total evidence E, then the function is certain that E is true—that is, $C_E(E) = 1$. Admittedly, it is not the case that experience always leads us to a certainty. After undergoing a course of experience, we might change our credences to values less than 1. I will not rule out this kind of epistemic possibility. In this regard, '$C_{\mathbb{E}}$' refers to the credence function that is updated from C after a course of experience directly changes the credences in some member of a partition $\mathbb{E}$, and nothing else.

Here it is assumed that $C_{\mathbb{E}}(E_i) < 1$ for any $E_i \in \mathbb{E}$. More generally, I will use 'C_{old}' to denote a credence function that an agent has before she undergoes a course of experience that directly changes some credences. In addition, 'C_{new}' will be used to stand for a credence function that is updated from C_{old} after she undergoes the experience in question.

2 Chances and Updating

In this section, I will formulate several versions of the Principal Principle and Conditionalization. First, I will provide some versions of the Principal Principle, which are applied to various evidential situations. Second, two versions of Conditionalization will be given. Each version of Conditionalization states how our credences should evolve when we are placed in a particular evidential situation.

2.1 The Principal Principle

Let me first consider Lewis's original Principal Principle.

PP$_{\text{initial}}$: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$C(A|U_{ch}) = ch(A), \text{ if defined.}$$

As explained, 'C' refers to an initial credence function that is coherent and regular.

It is natural, on the other hand, that credence-chance norms including the Principal Principle should not rule out any kind of credences. In particular, such norms should be able to constrain any credences no matter what evidential situation the credences are placed in. However, Lewis's original version concerns only a special kind of evidential situations, in which the credences undergo no course of experience and so has no evidence. Thus, the principle should be regarded as staying incomplete unless it constrains the relationship between chances and posterior credences that agents obtain after experiencing something. That is, the Principal Principle should be able to formulate the relationships between C_E and ch, and between $C_{\mathbb{E}}$ and ch.

Let's consider the first relationship. How should C_E respect or be guided by an ur-chance function ch, via the Principal Principle? In other words, to what chance function should $C_E(\cdot|U_{ch})$ be equal? Obviously, it should not be the case that: For

any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$C_E(A|U_{ch}) = ch(A), \text{ if defined.} \tag{1}$$

This is because (1) yields a contradiction when $ch(E) < 1$. In particular, it follows from (1) that $C_E(E|U_{ch}) = 1 \neq ch(E) < 1$ when $C_E(U_{ch}) > 0$. Here it is noteworthy that if chance functions can be regarded as the epistemic experts for C_E, then the chance functions ought to have at least the total evidence E. In other words, C_E should be guided by the chance functions that reflect at least E. Then, how about the chance functions that reflect more than C_E's total evidence? Let ch_{EF} be the chance function that reflects EF. Assume that E does not entail F. Then, ch_{EF} is the chance function that reflects more than C_E's total evidence. Now, consider that: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$C_E(A|U_{ch}) = ch_{EF}(A), \text{ if defined.} \tag{2}$$

Could it be correct? No. Note that E is the total evidence of C_E, and E does not entail F. Thus, it should hold that $C_E(F) < 1$. According to (2) and the probability calculus, however, this cannot be the case.[2]

As a result, if a credence function is to respect or be guided by a chance function in accordance with the Principal Principle, then the chance function should reflect the same proposition as the total evidence that the credence function has. In other words, C_E should be guided by ch_E, rather than ch or ch_{EF}. Note that it was assumed that $ch_E(\cdot) = ch(\cdot|E)$. Then, we have that:[3]

PP$_{\text{certainty}}$: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$C_E(A|U_{ch}) = ch(A|E), \text{ if defined.}$$

Now, let's consider the relationship between $C_{\mathbb{E}}$ and ch. Here, someone may pay attention to the similarity between C and $C_{\mathbb{E}}$. As assumed, the initial credence

[2] According to the aforementioned assumptions, $ch_{EF}(F) = ch(F|EF) = 1$, where ch is the relevant ur-chance function. Then, we have that:

$$\begin{aligned} C_E(F) &= \sum_{U_{ch} \in \mathbb{U}} C_E(U_{ch}) C_E(F|U_{ch}) \\ &= \sum_{U_{ch} \in \mathbb{U}} C_E(U_{ch}) ch_{EF}(F) = \sum_{U_{ch} \in \mathbb{U}} C_E(U_{ch}) = 1, \end{aligned}$$

and this result contradicts the assumption that E is our total evidence, and so $C_E(F) < 1$.

[3] This kind of formulation is not new one. Some similar formulations can be found in Meacham (2010) and Pettigrew (2014).

function C has no evidence. How about $C_{\mathbb{E}}$? Note that $C_{\mathbb{E}}$ is the function updated from C after a course of experience directly changes the credences in some member of a partition $\mathbb{E}$, and nothing else. It was assumed in the previous section that $C_{\mathbb{E}}(E_i) < 1$ for any E_i. Thus, there is no proposition of which $C_{\mathbb{E}}$ gets to be newly certain after the experience, and so $C_{\mathbb{E}}$ also has no evidence. For this reason, someone may think that the relationship between $C_{\mathbb{E}}$ and ch should be formulated in a similar way to $\text{PP}_{\text{initial}}$. Here is such a formulation:[4]

PP′: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$C_{\mathbb{E}}(A|U_{ch}) = ch(A), \text{ if defined.}$$

However, a little consideration reveals that PP′ is problematic. Note that it follows from PP′ and the probability calculus that:[5]

$$inf(\{ch(E_i) : U_{ch} \in \mathbb{U}\}) \leq C_{\mathbb{E}}(E_i) \leq sup(\{ch(E_i) : U_{ch} \in \mathbb{U}\}). \quad (3)$$

Note that $\{ch(E_i) : U_{ch} \in \mathbb{U}\}$ is a set of the values that each ur-chance function assigns to E_i. Thus, if there are some E_is that violate (3), then PP′ will yield a contradiction. So, if we are to take PP′ as a plausible rule governing the relationship between ch and $C_{\mathbb{E}}$, we should be committed to a norm stating that the new credences in E_is should have upper and lower bounds, which heavily depend on the values of objective chances. However, such a norm seems to be epistemologically implausible. Consider a coin toss, for example. Let H_t be the proposition

[4] A similar formulation can be found in Nissan-Rozen (2013). Admittedly, there is a slight difference between PP′ and the formulation in Nissan-Rozen (2013). Unlike PP′, Nissan-Rozen formulates the principle, as follows: For any A,

$$C_{\mathbb{E}}(A|XE) = x,$$

where X is the proposition that the chance of A, at a time, is x, and E is an admissible proposition. Note that Lewis's original principle is also formulated as follows: For any A,

$$C(A|XE) = x,$$

where X is the proposition that the chance of A, at a time, is x, and E is an admissible proposition. That is, Nissan-Rozen formulates the principle related to $C_{\mathbb{E}}$ in a very similar way to the principle related to C. For this reason, his formulation has the same spirit as PP′.

[5] Since $\mathbb{U}$ is a partition, it holds that $\sum_{U_{ch} \in \mathbb{U}} C_{\mathbb{E}}(U_{ch}) = 1$. Thus, it follows from PP′ and the probability calculus that: For any $E_i \in \mathbb{E}$,

$$\begin{aligned} C_{\mathbb{E}}(E_i) &= \sum_{U_{ch} \in \mathbb{U}} C_{\mathbb{E}}(U_{ch}) C_{\mathbb{E}}(E_i|U_{ch}) \\ &= \sum_{U_{ch} \in \mathbb{U}} C_{\mathbb{E}}(U_{ch}) ch(E_i). \end{aligned} \quad (*)$$

That is, $C_{\mathbb{E}}(E_i)$ must be a mixture of $ch(E_i)$s, with the weights being the new credence in U_{ch}. Then, (*) entails (3), as required.

the the coin lands heads at time t. Suppose that you know that the ur-chances of H_t are in a particular interval—[0.3, 0.8], say. After the coin is tossed at time t, you observe the result, and then updates you credence in H_t. Should the new credence, which is directly affected by the observation, be in [0.3, 0.8]? It seems not. Your credence in H_t could have any value in [0,1], depending on the observational conditions and the result of the coin toss. It is no wonder that it may appear to you that the coin lands heads, and so you get to believe H_t to the degree of a value close to 1. However, PP' prohibits you from updating the credence in H_t in this way.

Then, how should we formulate the relationship between $C_{\mathbb{E}}$ and ch? The relationship between $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{certainty}}$ may provide a clue to such a formulation. Note that C_E is the credence function updated from C after experience directly changes the credence in E to the maximal value, i.e. 1, and nothing else. Moreover, it follows from $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{certainty}}$ that: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$C_E(A|EU_{ch}) = C(A|EU_{ch}) = ch(A|E), \text{ if defined.} \tag{4}$$

This equation says that the conditional credences given EU_{ch} should remain the same through the belief update from C to C_E. Here E is the proposition whose credence is directly changed by experience.

On the other hand, $C_{\mathbb{E}}$ is the credence function updated from C after experience directly changes the credences in the members of $\mathbb{E}$ to a value less than 1, and nothing else. In this case, E_is are the propositions whose credences are directly changed by experience. Similar to (4), then, it may be required that: For any $A \in \mathcal{F}$, $E_i \in \mathbb{E}$, and $U_{ch} \in \mathbb{U}$,

$$C_{\mathbb{E}}(A|E_iU_{ch}) = C(A|E_iU_{ch}) = ch(A|E_i), \text{ if defined.} \tag{5}$$

As explained above, there seems to be some rationale for $\text{PP}_{\text{certainty}}$, which, together with $\text{PP}_{\text{initial}}$, entails (4). Moreover, it seems hard to find plausible ways of discriminating (5) from (4). As a result, if we regard $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{certainty}}$ as norms governing the relationship between chances and credences, then we should also take (5) as such a norm. Then, $\text{PP}_{\text{initial}}$ and (5) jointly entail that:

$\text{PP}_{\text{uncertainty}}$: For any $A \in \mathcal{F}$, $E_i \in \mathbb{E}$ and $U_{ch} \in \mathbb{U}$,

$$C_{\mathbb{E}}(A|E_iU_{ch}) = ch(A|E_i), \text{ if defined.}$$

Let me briefly note some desirable features of $\text{PP}_{\text{uncertainty}}$. First, $\text{PP}_{\text{uncertainty}}$ is free from the problem that besets PP′. That is, the new credence in E, which is not in between $inf(\{ch(E_i) : U_{ch} \in \mathbb{U}\})$ and $sup(\{ch(E_i) : U_{ch} \in \mathbb{U}\})$, does not lead to any contradiction. Unlike PP′, thus, $\text{PP}_{\text{uncertainty}}$ does not render us to be committed to the claim that the new credences should have upper and lower bounds. Second, $\text{PP}_{\text{uncertainty}}$ is a generalized version of $\text{PP}_{\text{certainty}}$—but, PP′ not. When one member of $\mathbb{E}$ is total evidence, $\text{PP}_{\text{uncertainty}}$ is equivalent to $\text{PP}_{\text{certainty}}$. However, PP′ may lead us to a contradiction when one member of $\mathbb{E}$ is total evidence. Suppose that $E_k (\in \mathbb{E})$ is total evidence, and so $C_{\mathbb{E}}(E_k) = 1$. Then, a contradiction follows when there is an ur-chance function ch such that $ch(E_k) < 1$. Suppose that there is such a chance function. Then, we have that $1 = C_{\mathbb{E}}(E_k|U_{ch}) = ch(E_k) < 1$. A more interesting difference between $\text{PP}_{\text{uncertainty}}$ and PP′ is intimately related to the Bayesian belief updating. Some authors formulate the relationship between C and $C_{\mathbb{E}}$ using PP′, and show that there is a kind of incompatibility between the Principal Principle and the Bayesian updating rules (For example, see Levi (1967) and Nissan-Rozen (2014)). As will be shown in what follows, $\text{PP}_{\text{uncertainty}}$ does not suffer from such an incompatibility. I will consider briefly this incompatibility in the next section. Before that, I should formulate the Bayesian updating rules.

2.2 The Bayesian Updating Rules

Roughly speaking, Conditionalization is a rule stating how the *direct* impact of experience on an agent's particular credences should be propagated into her overall credal system. Thus, Conditionalization can be formulated in several ways, according to what credences are directly changed by experience. The so-called Simple Conditionalization (SC) is a rule governing a way of updating an agent's overall credal states when she gets to have some evidence with certainty. On the other hand, what is called Jeffrey Conditionalization (JC) is a rule that can be applied to cases in which a course of experience directly changes an agent's credences without any certainty. Here are such versions of Conditionalization:[6]

SC: When a course of experience directly changes an agent's credence in E

[6] More exactly, the equation in JC should be

$$C_{new}(A) = \sum_{E_i \in \{E_i : C_{old}(E_i) > 0\}} C_{new}(E_i) C_{old}(A|E_i).$$

In what follows, for notational simplicity, I will use 'E_i' rather than '$E_i \in \{E_i : C_{old}(E_i) > 0\}$' when there is no danger of confusion.

to 1 and nothing else, for any $A \in \mathcal{F}$,

$$C_{new}(A) = C_{old}(A|E), \text{ if defined.}$$

JC: When a course of experience directly changes an agent's credences in some members in $\mathbb{E}$ to values less than 1 and nothing else, for any $A \in \mathcal{F}$.

$$C_{new}(A) = \sum_{E_i} C_{new}(E_i) C_{old}(A|E_i).$$

With the help of SC and JC, we can formulate the relationships between C and C_{E}, and between C and $C_{\mathbb{E}}$. That is, we have that:

SC$_0$: For any $A \in \mathcal{F}$,

$$C_{E}(A) = C(A|E), \text{ if defined.}$$

JC$_0$: For any $A \in \mathcal{F}$,

$$C_{\mathbb{E}}(A) = \sum_{E_i} C_{\mathbb{E}}(E_i) C(A|E_i).$$

Moreover, we are also able to formulate sequential belief updating. Let $C_{E,F}$ be the credence function that is sequentially updated from C by SC on E, and then by SC on F. Then, it is easily ascertained that: For any $A \in \mathcal{F}$,

$$C_{E,F}(A) = C(A|EF), \text{ if defined.}$$

We also obtain that $C_{E,F} = C_{F,E}$. That is, the sequential belief updating by SC is *commutative* in the sense that the final credence function is insensitive to the order in which the pieces of evidence are incorporated into the credal system.

On the other hand, the so-called 'Bayes factors' may be helpful in formulating sequential belief updating by JC. Some philosophers think that the Bayes factors properly represent the impact of experience *itself* with the old credences factored out.[7] Suppose that an agent's credence function is updated from C_{old} to C_{new} after a course of experience directly changes her credences in some members of

[7] Some discussions about the Bayes factors have been suggested by several authors like Field (1978), Jeffrey (2004, pp.55-59) and Wagner (2003, pp.360-362).

$\mathbb{E}$. Then, the Bayes factor of E_i against E_1 is defined as follows:

$$\beta^{C_{new}}_{C_{old}}(E_i : E_1) = \frac{C_{new}(E_i)/C_{old}(E_i)}{C_{new}(E_1)/C_{old}(E_1)}.$$

Here E_1 is an arbitrary anchored proposition that is a member of $\mathbb{E}$. Using this factor, we can re-parameterize JC, as follows:

JC*: When a course of experience directly changes an agent's credences in some members in $\mathbb{E}$ to values less than 1, and nothing else, for any $A \in \mathcal{F}$,

$$C_{new}(A) = \frac{\sum_{E_i} \beta^{C_{new}}_{C_{old}}(E_i : E_1) C_{old}(AE_i)}{\sum_{E_i} \beta^{C_{new}}_{C_{old}}(E_i : E_1) C_{old}(E_i)}.$$

Now, let '$C_{\mathbb{E},\mathbb{F}}$' refer to a credence function that is sequentially updated from C by JC on a partition $\mathbb{E}$, and then by JC on another partition $\mathbb{F}$. Then, it follows from JC* that: For any $A \in \mathcal{F}$,

$$C_{\mathbb{E},\mathbb{F}}(A) = \frac{\sum_{E_i,F_j} \beta^{C_{\mathbb{E}}}_{C}(E_i : E_1) \beta^{C_{\mathbb{E},\mathbb{F}}}_{C_{\mathbb{E}}}(F_j : F_1) C(AE_iF_j)}{\sum_{E_i,F_j} \beta^{C_{\mathbb{E}}}_{C}(E_i : E_1) \beta^{C_{\mathbb{E},\mathbb{F}}}_{C_{\mathbb{E}}}(F_j : F_1) C(E_iF_j)}.$$

From this equation, it follows that $C_{\mathbb{E},\mathbb{F}} = C_{\mathbb{F},\mathbb{E}}$ if $\beta^{C_{\mathbb{E}}}_{C}(E_i : E_1) = \beta^{C_{\mathbb{F},\mathbb{E}}}_{C_{\mathbb{F}}}(E_i : E_1)$ and $\beta^{C_{\mathbb{E},\mathbb{F}}}_{C_{\mathbb{E}}}(F_j : F_1) = \beta^{C_{\mathbb{F}}}_{C}(F_j : F_1)$ for any $E_i \in \mathbb{E}$ and $F_j \in \mathbb{F}$. That is, the sequential belief updating by JC is *commutative* in the sense that the final credence function is insensitive to the order in which the Bayes factors are incorporated into the credal system.

Heretofore, I provide some explanations regarding two groups of epistemic norms. The first is a norm stating how credence functions should respect or be guided by objective chances; the second is a norm how credence functions should evolve after experience something. In the next section, I will examine and explicate the relationship between these two groups. In doing so, I will deliver good news regarding the relationship—that is, the Principal Principle and Conditionalization are complementary to each other in a particular sense.

3 The Complementarity between the Principal Principle and Conditionalization

In this section, I will examine how the Principal Principle is related to Conditionalization. In particular, I will show that:

Comp1. If an agent satisfies $\text{PP}_{\text{initial}}$ and updates her credence function by means of Conditionalization, then she must satisfy the Principal Principle no matter what evidential situations she is placed in.

Comp2. If an agent updates her credences in several particular propositions by means of Conditionalization and she satisfies the Principal Principle no matter what evidential situations she is placed in, then her overall credal state must evolve in accordance with Conditionalization.

These results, I think, could be regarded as stating a kind of complementarity between the epistemic norms in question. According to Comp1, we don't have to formulate *every* version of the Principal Principle if the credences evolves by means of Conditionalization. Similarly, Comp2 says that we don't have to require for rational agents to update *overall* credal state by means of Conditionalization if the agent satisfies the Principal Principle.

First, let me consider some propositions related to Comp1. The relevant proofs are given in Appendix. Here are such propositions:

Proposition 3.1: If C satisfies $\text{PP}_{\text{initial}}$ and C_E is the credence function updated from C by SC on E, then C_E satisfies $\text{PP}_{\text{certainty}}$.

Proposition 3.2: If C satisfies $\text{PP}_{\text{initial}}$ and $C_{\mathbb{E}}$ is the credence function updated from C by JC on $\mathbb{E}$, then $C_{\mathbb{E}}$ satisfies $\text{PP}_{\text{uncertainty}}$.

Proposition 3.1 and its proof can be found in several works like Pettigrew (2014), but Proposition 3.2 is a new result. According to these propositions, $\text{PP}_{\text{certainty}}$ and $\text{PP}_{\text{uncertainty}}$ are theoretically redundant since Conditionalization, i.e., SC and JC, guarantees that the credence function, which is updated from the initial function satisfying $\text{PP}_{\text{initial}}$, also satisfies the Principal Principle, i.e., $\text{PP}_{\text{certainty}}$ and $\text{PP}_{\text{uncertainty}}$.

In this regard, I would like to pay attention to the following proposition:

Proposition 3.3: It is *not* the case that $C_{\mathbb{E}}$ satisfies PP′ if C satisfies $\text{PP}_{\text{initial}}$ and $C_{\mathbb{E}}$ is the credence function updated from C by JC on $\mathbb{E}$.

Nissan-Rozen (2013) shows a similar thing to this proposition, and then he concludes that the Principal Principle is not preserved under JC. Indeed, this is old news for Bayesians. For example, Levi (1967) has already provided a similar claim in order to argue that JC leads us to a contradiction under the assumption that some relevant likelihoods should remain the same.[8] However, Proposition 3.3, I think, shows neither that the Principal Principle is not preserved under JC, nor that JC suffers from a contradiction. This is because PP′ can not be regarded as a plausible chance-credence norm governing the relationship between C and $C_{\mathbb{E}}$. As noted earlier, PP′ imposes an epistemologically implausible restriction on the degree to which experience directly changes our credences. Rather, we should think that $\text{PP}_{\text{uncertainty}}$, which is free from such a problem, is more plausible, and so that, as Proposition 3.2 states, the Principal Principle is preserved under JC.

Now, let me consider some propositions related to Comp2. To do so, we need to make a distinction between *local* and *global* updating. Here are the associated definitions:

Definition 3.1: Local Updating

(3.1a) C_E is *locally* updated from C by SC on E relative to a set $S \subset \mathcal{F}$ if and only if $C_E(A) = C(A|E)$ for any $A \in S \subset \mathcal{F}$.

(3.1b) $C_{\mathbb{E}}$ is *locally* updated from C by JC on $\mathbb{E}$ relative to a set $S \subset \mathcal{F}$ if and only if $C_{\mathbb{E}}(A) = \sum_{E_i} C_{\mathbb{E}}(E_i)C(A|E_i)$ for any $A \in S \subset \mathcal{F}$.

When a credence function is locally updated by means of Conditionalization relative to a proper subset S of $\mathcal{F}$, Conditionalization stays silent on how the credences in the propositions that are not members of S should evolve. On the other hand, *Global Updating* is defined as follows:

Definition 3.2: Global Updating

(3.2a) C_E is *globally* updated from C by SC on E if and only if $C_E(A) = C(A|E)$ for any $A \in \mathcal{F}$.

(3.2b) $C_{\mathbb{E}}$ is *globally* updated from C by JC on $\mathbb{E}$ if and only if $C_{\mathbb{E}}(A) = \sum_{E_i} C_{\mathbb{E}}(E)C(A|E)$ for any $A \in \mathcal{F}$.

Unlike Local Updating, when a credence function are globally updated, the credences in all propositions should be updated in accordance with Conditionalization.

[8] Several discussions about Levi's argument can be found in Jeffrey (1970) and Harper and Kyburg (1968).

Interestingly, the Principal Principle globalizes the local updating relative to a particular set of propositions. Here is the relevant proposition:

Proposition 3.4: Suppose that C and C_E, respectively, satisfy $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{certainty}}$. Suppose also that C_E is locally updated from C by SC on E relative to $\mathbb{U}$. Then, C_E is the credence function globally updated from C by SC on E.

Note that $\mathbb{U}$, which is a set of U_{ch}s, is a subset of $\mathcal{F}$. This proposition, I think, reveals how objective chances guide our credences. Suppose that an agent satisfies the Principal Principle, i.e., $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{certainty}}$, and that she updates her credences by SC on total evidence. Then, it can be said, with the help of Proposition 3.4, that, when an agent satisfies the Principal Principle, the impact of the total evidence on her overall credal system comes entirely by way of *her new credences about chance functions*.

A similar observation regarding $C_{\mathbb{E}}$ can be made as follows:

Proposition 3.5: Suppose that C and $C_{\mathbb{E}}$ respectively satisfy $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{uncertainty}}$. Suppose also that $C_{\mathbb{E}}$ is locally updated from C by JC on $\mathbb{E}$ relative to $\{U_{ch}E_i\}$. Then, $C_{\mathbb{E}}$ is the credence function globally updated from C by JC on $\mathbb{E}$.

Here, the set $\{U_{ch}E_i\}$ is a partition that has $U_{ch}E_i$s as members, and that is a subset of $\mathcal{F}$. Similar to Proposition 3.4, this proposition states that the impact of experience on the new credences in $U_{ch}E_i$s, which are locally updates by JC on E_i, can be extended into the overall credal system via the Principal Principle. In regard to $C_{\mathbb{E}}$, however, we cannot say that the impact of the experience on the overall credal system comes entirely by way of the new credences in U_{ch}s (when the Principal Principle is satisfied). Rather, it should be said that the impact in question comes entirely by way of the new credences in $U_{ch}E_i$s (when the Principal Principle is satisfied). Anyway, it is still true that the Principal Principle globalizes our local updating so that the impact of experience is propagated into the overall credal system.

In this section, I have provided some results showing that the Principal Principle and Conditionalization are complementary to each other in a particular sense. This result, I think, can be regarded as good news about the relationship between the two norms in question. However, this kind of complementarity is just one side of the relationship between them. There is a problem that seems to beset the epistemic relationship between objective chances and our rational credences. I will deliver this bad news in what follows.

4 When Chances Guide Credences

The bad news that will be given in this section is that the credences, which obey both the Principal Principle and Conditionalization, is sensitive to a seemingly irrelevant factor to our epistemic rationality—in particular, it will be shown that the credences, which we should have when the two norms in question are satisfied, depends on which of chances and experience comes first. To see this, let me reconsider initial credence functions.

4.1 Chance-fed and Chance-free Credence Functions

As stated, initial credence functions satisfying $\text{PP}_{\text{initial}}$ have no information that is obtained by experience—that is, experience has not caused any change in credences, and so the initial credence function has no evidence. Then, can we say the initial credence functions satisfying $\text{PP}_{\text{initial}}$ do not reflect any kind of information? Note that the principle can be regarded as an answer to the question, 'What should the initial credences be like when they are guided by chances?'. Thus, the initial credence functions satisfying $\text{PP}_{\text{initial}}$ can be regarded as one that reflects or is guided by some information about chances. In this sense, it is not the case that the initial credence functions satisfying $\text{PP}_{\text{initial}}$ do not have any kind of information.

To express this point more clearly, let me introduce another notation '$\mathfrak{C}$' that denotes an initial credence function that has not reflected the information about chances yet. I will call this kind of credence function 'a chance-free initial credence function'. Chance-free initial credence functions are assumed to be probabilistically coherent and regular. This kind of initial function may be updated in a certain way so as to satisfy $\text{PP}_{\text{initial}}$. I will call the function so obtained 'a chance-fed initial credence function', which is denoted by '$\mathfrak{C}_{\mathbb{CH}}$'. Here, the subscript '$\mathbb{CH}$' is intended to express that the information about chances is fed to the initial credence function $\mathfrak{C}$. With these notations at hand, $\text{PP}_{\text{initial}}$ can be paraphrased as follows:

PP*$_{\text{initial}}$: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{\mathbb{CH}}(A|U_{ch}) = ch(A), \text{ if defined.}$$

Note that the Principal Principle does not require that any chance-free credence function $\mathfrak{C}$ should satisfy that $\mathfrak{C}(A|U_{ch}) = ch(A)$ for any $A \in \mathcal{F}$. The principle should be thought as of a norm governing what chance-fed initial credences, not chance-free ones, should be like.

Then, how should chance-fed initial functions be updated from chance-free initial ones? To put it another way, is there any rule governing the relationship between $\mathfrak{C}$ and $\mathfrak{C}_{\mathbb{CH}}$? Here we need to consider what credences are directly changed when information about chances are incorporated, via the Principal Principle, into our credal system. The Principal Principle, in particular PP*$_{\text{initial}}$, could be thought of as a norm stating what values should be newly assigned to some conditional credences—that is, it requires that, for any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$, $ch(A)$ should be newly assigned to the conditional credence in A given U_{ch}. Equivalently, it require that, for any $w \in \Omega$ and $U_{ch} \in \mathbb{U}$, $ch(w)$ should be newly assigned to the conditional credence in w given U_{ch}.[9]

Note that this requirement does not directly impose any constraint on other credences than the conditional credences in question. Then, is there any Bayesian updating rule stating how our credence function should evolve when the new conditional credences in question are obtained, and nothing else? More generally, is there any Bayesian updating rule stating how our credence function should evolve when a conditional credence in E given F is directly changed, and nothing else? Fortunately, yes.

4.2 Feeding Chances to Chance-free Credence Functions

What is often called Adams Conditionalization (AC) plays such a role.[10] Suppose that the conditional credences in E_is ($\in \mathbb{E}$) given F are directly changed from $C_{old}(E_i|F)$ to $C_{new}(E_i|F)$, and nothing else. Then, Adams Conditionalization is formulated as follows:

AC: When a course of experience directly changes an agent's conditional credences in some members in $\mathbb{E}$ given F, and nothing else, for any $A \in \mathcal{F}$,

$$C_{new}(A) = \sum_{E_i} C_{old}(F) C_{new}(E_i|F) C_{old}(A|E_iF) + C_{old}(A\neg F).$$

AC *itself* cannot handle a case in which the conditional credence in E given $\neg F$, as well as the conditional credence in E given F , is directly changed. That is, AC *itself*

[9] Note that it is assumed that ch is a probability function. Thus, the following two propositions are equivalent to each other: (i) For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}\mathfrak{C}_{\mathbb{CH}}(A|U_{ch}) = ch(A)$, if defined ; (ii) For any $w \in \Omega$ and $U_{ch} \in \mathbb{U}$,$\mathfrak{C}_{\mathbb{CH}}(w|U_{ch}) = ch(w)$, if defined.

[10] Adams Conditionalization is named by Bradley (2005). A similar discussion is also found in Wagner (2003). There are some attempts that solve several difficulties of Bayesian epistemology by means of this kind of Conditionalization. For example, Douven and Romeijn (2011) suggest a solution for the so-called Judy Benjamin Problem using Adams conditionalization.

is not a rule that governs the way of updating our credence function when conditional credences with various conditioning propositions are directly changed. However, we can easily generalize AC so that it can deal with such a case. Suppose that the conditional credences in E_is given F_1, the conditional credences in E_is given F_2, $\cdots$ are directly changed and nothing else. Here, it is assumed that the set of F_is is a partition, which will be denoted by '$\mathbb{F}$'.

For the present purpose, let me consider the following equation:

$$C_{old}(A) = \sum_{E_i} C_{old}(F)C_{old}(E_i|F)C_{old}(A|E_iF) + C_{old}(A\neg F), \qquad (6)$$

which follows from the probability calculus. It is noteworthy that, according to AC, the new credence function C_{new} is obtained by replacing $C_{old}(E_i|F)$s in (6) with $C_{new}(E_i|F)$. Now, consider the following equation, which is entailed by the probability calculus:

$$\begin{aligned} C_{old}(A) = &\sum_{E_i} C_{old}(F_1)C_{old}(E_i|F_1)C_{old}(A|E_iF_1) \\ &+ \sum_{E_i} C_{old}(F_2)C_{old}(E_i|F_2)C_{old}(A|E_iF_2) + \cdots \end{aligned} \qquad (7)$$

Then, we can obtain a generalized version of AC, which will be abbreviated to 'GAC', by replacing $C_{old}(E_i|F_j)$ with $C_{new}(E_i|F_j)$. Here is such a generalization:

GAC: When a course of experience directly changes an agent's conditional credences in some members in $\mathbb{E}$ given F_1, some members in $\mathbb{E}$ given F_2, $\cdots$, and nothing else, for any $A \in \mathcal{F}$,

$$C_{new}(A) = \sum_{E_i, F_j} C_{old}(F_j)C_{new}(E_i|F_j)C_{old}(A|E_iF_j).$$

It is noteworthy that GAC is equivalent to AC when $\mathbb{F} = \{F, \neg F\}$ and $C_{old}(E_i|\neg F) = C_{new}(E_i|\neg F)$ for any $E_i \in \mathbb{E}$.

Now, we can provide a Bayesian updating rule governing the relationship between $\mathfrak{C}$ and $\mathfrak{C}_{\mathbb{CH}}$. In other words, we can formulate a norm governing the way of feeding chances to initial chance-free credence functions. As explained, $\text{PP}_{\text{initial}}$ requires that, for any $w \in \Omega$ and $U_{ch} \in \mathbb{U}$, $ch(w)$ should be newly assigned to the conditional credence in w given U_{ch}. Then, we can think that, in the belief updating from $\mathfrak{C}$ to $\mathfrak{C}_{\mathbb{CH}}$, the conditional credences in w given U_{ch} are directly changed

from $\mathfrak{C}(w|U_{ch})$ to $ch(w)$, and nothing else. Then, we can derive a way of feeding chances to a chance-free initial credence function via the Principal Principle, as follows:[11]

Feeding Chances to Initial Credences (FCI): For any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A) = \sum_{ch} \mathfrak{C}(U_{ch})ch(A).$$

The derivation of FCI from PP*$_{\text{initial}}$ (or PP$_{\text{initial}}$) and GAC is given in Appendix.

FCI is a rule that constrains a chance-free initial credence function that does not undergo any course of experience, and so has no evidence. However, we can suggest a similar rule that constrains chance-free credence functions that undergo some course of experience before chances are fed to them. Let $\mathfrak{C}_E$ be a chance-free credence function with total evidence E. Moreover, let $\mathfrak{C}_{\mathbb{E}}$ be a chance-free function that is updated from $\mathfrak{C}$ after a course of experience directly changes the credences in some member of a partition $\mathbb{E}$, and nothing else. Then, chances may be fed to these chance-free functions with the help of the Principal Principle. Let $\mathfrak{C}_{E,\mathbb{CH}}$ be the chance-fed credence function updated from $\mathfrak{C}_E$ via PP$_{\text{certainty}}$. Similarly, let $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ be the chance-fed credence function updated from $\mathfrak{C}_{\mathbb{E}}$ via PP$_{\text{uncertainty}}$. Now, with these notations at hand, PP$_{\text{certainty}}$ and PP$_{\text{uncertainty}}$ can be paraphrased as follows:

PP*$_{\text{certainty}}$: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{E,\mathbb{CH}}(A|U_{ch}) = ch(A|E), \text{ if defined.}$$

PP*$_{\text{uncertainty}}$: For any $A \in \mathcal{F}$, $E_i \in \mathbb{E}$, and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A|U_{ch}E_i) = ch(A|E_i), \text{ if defined.}$$

Now, we can formulate, with the help of GAC and the Principal Principle, ways of updating from $\mathfrak{C}_E$ to $\mathfrak{C}_{E,\mathbb{CH}}$ and updating from $\mathfrak{C}_{\mathbb{E}}$ to $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$.

Similar to PP$_{\text{initial}}$, PP*$_{\text{certainty}}$ requires that, for any $w \in \Omega$ and $U_{ch} \in \mathbb{U}$, $ch(w|E)$ should be newly assigned to the conditional credence in w given U_{ch}, when $\mathfrak{C}_E$ is

[11] More exactly, the equation in FCI should be

$$\mathfrak{C}_{E,\mathbb{CH}}(A) = \sum_{ch \in \{ch : U_{ch} \in \mathbb{U}\}} \mathfrak{C}_E(U_{ch})ch(A|E).$$

For notational simplicity, I will use 'ch' rather than '$ch \in \{ch : U_{ch} \in \mathbb{U}\}$' when there is no danger of confusion in what follows. Similar kinds of simplification go with FCC and CCU.

updated to $\mathfrak{C}_{E,\mathbb{CH}}$. Then, the rule, which governs the way of feeding chances to $\mathfrak{C}_E$ via the Principal Principle, could be formulated as follows:

Feeding Chances to Credences under Certainty (FCC): For any $A \in \mathcal{F}$,

$$\mathfrak{C}_{E,\mathbb{CH}}(A) = \sum_{ch} \mathfrak{C}_E(U_{ch}) ch(A|E).$$

This rule follows from PP*$_{\text{certainty}}$and GAC, as shown in Appendix. A similar consideration goes with the relationship between $\mathfrak{C}_{\mathbb{E}}$ and $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$. When $\mathfrak{C}_{\mathbb{E}}$ is updated to $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$, $ch(w|E_i)$ should be newly assigned to the conditional credence in w given $U_{ch}E_i$ for any $w \in \Omega$, $E_i \in \mathbb{E}$, and $U_{ch} \in \mathbb{U}$. Then, the way of feeding chances to $\mathfrak{C}_{\mathbb{E}}$ can be formulated as follows:

Feeding Chances to Credences under Uncertainty (FCU): For any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) = \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i) ch(A|E_i).$$

As above, we can obtain this rule with the help of PP*$_{\text{uncertainty}}$ and GAC.

Heretofore, I formulate ways of feeding chances to chance-free credence functions. In doing so, I regarded such ways as a kind of Bayesian belief updating—in particular, belief updating by Adams Conditionalization related to some conditional credences, whose new values are given by the Principal Principle. Here it is noteworthy that $\mathfrak{C}_{E,\mathbb{CH}}$ can be regarded as one obtained by a sequential belief updating. That is, we can think that $\mathfrak{C}_{E,\mathbb{CH}}$ is updated from $\mathfrak{C}_E$, which has been updated from $\mathfrak{C}$ before. A similar consideration goes with $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$. Interestingly, this consideration yields some undesirable results.

4.3 When Chances Guide Credences

Note that, in the sequential belief updating from $\mathfrak{C}$ to $\mathfrak{C}_{E,\mathbb{CH}}$, chances guides credences *after* the total evidence E is obtained. The order could be reversed. That is, it is entirely possible that chances guides credences *before* the total evidence E is obtained. Let $\mathfrak{C}_{\mathbb{CH},E}$ be the credence function so obtained. That is, $\mathfrak{C}_{\mathbb{CH}}$ is updated from $\mathfrak{C}$ by means of FCI, and then $\mathfrak{C}_{\mathbb{CH},E}$ is updated from $\mathfrak{C}_{\mathbb{CH}}$ by means of SC. Similarly, let $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$ be the credence function that, by means of JC, is updated from $\mathfrak{C}_{\mathbb{CH}}$, which has been updated from $\mathfrak{C}$ by means of FCI.

Before I proceed further, note the following propositions.

Proposition 4.3.1 Suppose that our credences are updated by means of SC, FCI, and FCC. Then, for any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{E,\mathbb{CH}}(A|U_{ch}) = \mathfrak{C}_{\mathbb{CH},E}(A|U_{ch}) = ch(A|E), \text{ if defined.}$$

Proposition 4.3.2 Suppose that our credences are updated by means of JC, FCI, and FCU. Then, for any $A \in \mathcal{F}$, $E_i \in \mathbb{E}$, and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A|U_{ch}E_i) = \mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A|U_{ch}E_i) = ch(A|E_i), \text{ if defined.}$$

According to these propositions, when credence functions are updated by means of the relevant rules, $\text{PP}_{\text{certainty}}$ and $\text{PP}_{\text{uncertainty}}$ are satisfied no matter when chances is fed to our credal system. So far, so good.

However, we can derive a seemingly undesirable results—what final credence functions are obtained depends on *when chances guide credences*. More formally, we have that:

Proposition 4.3.3 Suppose that our credences are updated by means of SC, FCI, and FCC. Then, *it is not the case that* $\mathfrak{C}_{E,\mathbb{CH}}(A) = \mathfrak{C}_{\mathbb{CH},E}(A)$ for any $A \in \mathcal{F}$.

Note that $\mathfrak{C}_{E,\mathbb{CH}}$ and $\mathfrak{C}_{\mathbb{CH},E}$ have the same total evidence, and that they are updated from the same initial credence function. The only difference between them is when the impacts of chances is incorporated into the credal system. A very similar observation can be made in regard to the relationship between $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ and $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A)$, as follows:

Proposition 4.3.4 Suppose that our credences are updated by means of JC, FCI, and FCU. Then, it is not the case that $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) = \mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A)$ for any $A \in \mathcal{F}$ if $\beta^{\mathfrak{C}_{\mathbb{E}}}_{\mathfrak{C}}(E_i, E_1) = \beta^{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}}_{\mathfrak{C}_{\mathbb{CH}}}(E_i, E_1)$ for any $E_i \in \mathbb{E}$.

As explained in Section 2.2, some authors think that the Bayes factors well represent the impact of experience *itself* with old credences factored out. I assume, following such authors, that the Bayes factors play such a role.[12] Then, the clause

[12] Indeed, we can obtain the similar conclusion even if we does not make such an assumption. It is also shown, for example, that:

Proposition 4.3.4* Suppose that our credences are updated by means of JC, FCI, and FCU. Then, it is not the case that $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) = \mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A)$ for any $A \in \mathcal{F}$ if $\mathfrak{C}_{\mathbb{E}}(E_i) = \mathfrak{C}_{\mathbb{CH},\mathbb{E}}(E_i)$ for any $E_i \in \mathbb{E}$.

This proposition states that even if two agents, who share an initial chance-free credence function in common, obtain the same *new credences in* E_i*s*, their final credence functions could be different from each other according to when changes guide their credences.

'$\beta^{\mathfrak{C}_{\mathbb{E}}}_{\mathfrak{C}}(E_i, E_1) = \beta^{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}}_{\mathfrak{C}_{\mathbb{CH}}}(E_i, E_1)$ for any $E_i \in \mathbb{E}$' in Proposition 4.3.4 means that the experience that causes the belief updating from $\mathfrak{C}$ to $\mathfrak{C}_{\mathbb{E}}$ is the same as the experience that causes the belief updating from $\mathfrak{C}_{\mathbb{CH}}$ to $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$. That is, the proposition says that even if two agents, who share an initial chance-free credence function in common, undergo the same course of experience, their final credence functions could be different from each other, depending on when chances guide their credences.

In order to prove the above propositions, it is sufficient to provide belief updates in which $\mathfrak{C}_{E,\mathbb{CH}}(A) \neq \mathfrak{C}_{\mathbb{CH},E}(A)$ for some $A \in \mathcal{F}$, and in which $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) \neq \mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A)$ for some $A \in \mathcal{F}$ although $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ has the same Bayes factors as $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$. Here are the examples of such belief updates (Some relevant calculations are given in Appendix).

Example

There are six possible worlds, $w_1, \cdots,$ and w_6. The ur-chance function of w_1, w_2, and w_3 is ch'; the ur-chance function of the other worlds is ch^*. Thus, $U_{ch'} \equiv w_1 \vee w_2 \vee w_3$ and $U_{ch^*} \equiv w_4 \vee w_5 \vee w_6$. Let's assume that $E \equiv w_1 \vee w_2 \vee w_4 \vee w_5$ and $\neg E \equiv w_3 \vee w_6$. I will describe the relevant probability assignments by means of the following 2×3 tables.

Possible Worlds

w_1	w_4
w_2	w_5
w_3	w_6

Each cell of this table refers to the associated possible world. The tables whose cells are filled with a numerical value express the relevant probability assignments. Here are such tables.

$\mathfrak{C}$

1/10	2/10
2/10	2/10
2/10	1/10

ch'

1/3	0
1/3	0
1/3	0

ch^*

0	2/4
0	1/4
0	1/4

These tables express, for example, that $\mathfrak{C}(w_1) = 1/10$, $ch'(w_2) = 1/3$, and $ch^*(w_6) = 1/4$.

Now, we can calculate the credence assignments of $\mathfrak{C}_{\mathbb{CH},E}$ and $\mathfrak{C}_{E,\mathbb{CH}}$ by means of SC, FCI and FCC. Figure 1 displays two different sequential belief updates from

Figure 1: $\mathfrak{C}_{\mathbb{CH},E}$ vs. $\mathfrak{C}_{E,\mathbb{CH}}$

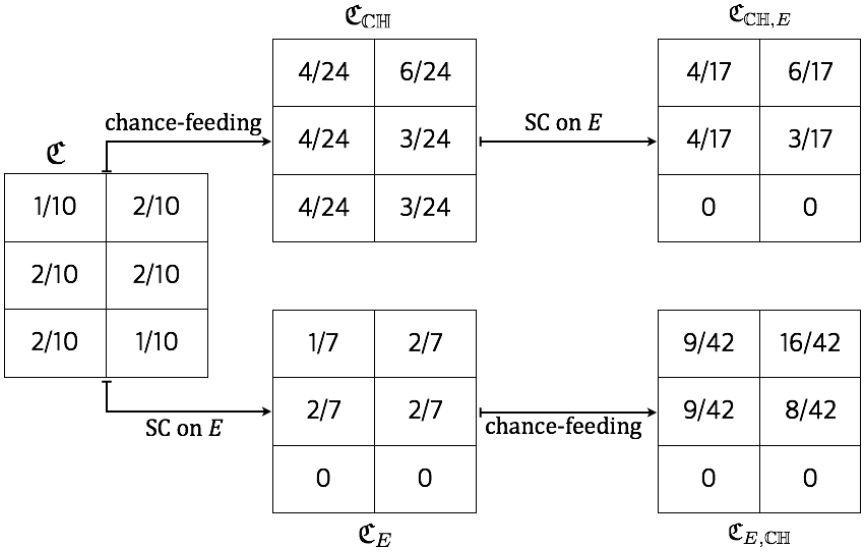
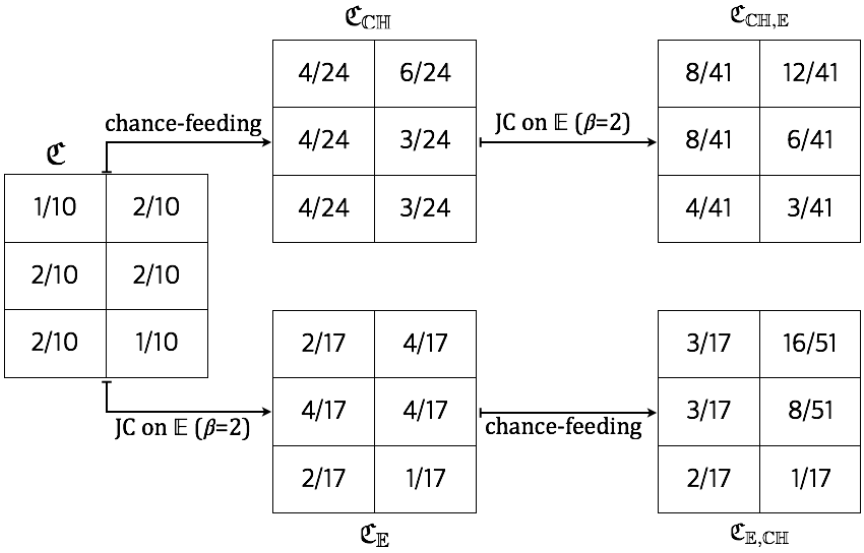


Figure 2: $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$ vs. $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$, where $\beta^{\mathfrak{C}_{\mathbb{E}}}_{\mathfrak{C}}(E,\neg E) = \beta^{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}}_{\mathfrak{C}_{\mathbb{CH}}}(E,\neg E) = 2$.



the chance-free initial credence function $\mathfrak{C}$. Note that $\mathfrak{C}_{\mathbb{CH},E}(w_i|U_{ch'}) = \mathfrak{C}_{E,\mathbb{CH}}(w_i|U_{ch'}) = ch'(w_i)$ and $\mathfrak{C}_{\mathbb{CH},E}(w_i|U_{ch^*}) = \mathfrak{C}_{E,\mathbb{CH}}(w_i|U_{ch^*}) = ch^*(w_i)$ for any w_i. That is, both of $\mathfrak{C}_{\mathbb{CH},E}$ and $\mathfrak{C}_{E,\mathbb{CH}}$ satisfy $\text{PP}_{\text{certainty}}$. However, $\mathfrak{C}_{\mathbb{CH},E} \neq \mathfrak{C}_{E,\mathbb{CH}}$. Note, for example, that $\mathfrak{C}_{\mathbb{CH},E}(w_1) = 4/17 \neq 9/42 = \mathfrak{C}_{E,\mathbb{CH}}(w_1)$. This result means that the final results vary depending on *when* chances guide credences. $\mathfrak{C}_{E,\mathbb{CH}}$ is obtained when chances guide credences *after* the credences obtain the total evidence E, whereas $\mathfrak{C}_{\mathbb{CH},E}$ is obtained when chances guide credence *before* the credences obtain E.

A similar observation can be made in regard to the relationship between $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$ and $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$. Suppose that $\mathbb{E} = \{E, \neg E\}$, and that a course of experience leads to directly change the credences in E and in $\neg E$, and nothing else. Suppose also that the belief updating from $\mathfrak{C}$ to $\mathfrak{C}_{\mathbb{E}}$ has the same Bayes factor of E against $\neg E$ as the belief updating from $\mathfrak{C}_{\mathbb{CH}}$ to $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$—in particular, suppose that

$$\beta^{\mathfrak{C}_{\mathbb{E}}}_{\mathfrak{C}}(E, \neg E) = \beta^{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}}_{\mathfrak{C}_{\mathbb{CH}}}(E, \neg E) = 2.$$

Then, we can calculate the credence assignments of $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$ and $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ by means of JC, FCI, and FCU. Figure 2 displays the results. It can be easily ascertained that both $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ and $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$ satisfy $\text{PP}_{\text{uncertainty}}$. However, $\mathfrak{C}_{\mathbb{E},\mathbb{CH}} \neq \mathfrak{C}_{\mathbb{CH},\mathbb{E}}$—for example, $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(w_1) = 8/41 \neq 3/17 = \mathfrak{C}_{\mathbb{E},\mathbb{CH}}(w_1)$. Then, it is also said that the final results vary according to whether chances guide credences before or after experience.

Propositions 4.3.3 and 4.3.4 reveal another kind of non-commutativity of Bayesian belief updating. As well-known, Bayesian belief updating by means of JC is non-commutative in the sense that the final credence function is sensitive to the order in which *the new credences* are incorporated into the credal system. As explained in Section 2.2, however, Bayesian belief updating by SC is commutative. Moreover, the belief updating by JC could also be regarded as commutative in the sense that the final credence function is *in*sensitive to the order in which *the Bayes factors* are incorporated into the credal system.

Interestingly, non-commutativity in Proposition 4.3.3 is not due to the belief updating by JC, and non-commutativity in Proposition 4.3.4 follows even if two different sequential belief updates have the associated Bayes factors in common. So, we should say that Propositions 4.3.3 and 4.3.4 reveal entirely new non-commutativity, which it seems hard to handle in existing ways. In response to such non-commutativity, it is not a good strategy, for example, to deny that JC is a rational belief updating rule, and to argue that the Bayes factors, rather than the new credences, properly represent the impact of experience itself.

Then, is there any way of responding these kinds of non-commutativity? Note again that the non-commutativity in question follows from two groups of epistemic norm. The first includes several versions of the Principal Principle—that is, $\text{PP*}_{\text{initial}}$, $\text{PP*}_{\text{certainty}}$, and $\text{PP*}_{\text{uncertainty}}$. The second consists of several versions of Conditionalization—that is, SC, JC, and (G)AC. Thus, we cannot circumvent such non-commutativity as far as the Principal Principle and Conditionalization should be regarded as plausible epistemic norms. Someone may attempt to argue that these kinds of non-commutativity are epistemologically intuitive, and so threaten neither the Principal Principle nor Conditionalization. For example, it may be argued that the impact of chances through the belief update from $\mathfrak{C}$ to $\mathfrak{C}_{\mathbb{CH}}$ on a credal system is different from the impact of chances through the belief update from $\mathfrak{C}_E$ to $\mathfrak{C}_{E,\mathbb{CH}}$on the system, and so that it is epistemically intuitive that $\mathfrak{C}_{\mathbb{CH},E}$ is different from $\mathfrak{C}_{E,\mathbb{CH}}$.

However, this attempt, even if it is successful, is at best a partial response to the non-commutativity in question. The real problem is *not* that the final credences could be different depending on when objective chances guide our rational credences, *but* that there seems to be no rationale that helps us decide when objective chances should guide our rational credences. We might understand intuitively and epistemologically why $\mathfrak{C}_{E,\mathbb{CH}}$ cannot help being different from $\mathfrak{C}_{\mathbb{CH},E}$. However, what is more important is to provide an unequivocal way of constraining our rational credences and their updating, and so stating *when chances should guide our rational credences*. As of now, it seems to be hard to find such a way. For this reason, Propositions 4.3.3 and 4.3.4 must be bad news for Bayesians.

Appendix

The following proofs and derivations have various conditional credences and chances. For presentational simplicity, I will assume in what follows that such conditional credences and chances are all well defined. I think this assumption yields no significant confusion.

A proof of Proposition 3.1:

Suppose that C satisfies $\text{PP}_{\text{initial}}$ and C_E is the credence function updated from C by SC on E. Then, we have that: For any $A \in \mathcal{F}$ and $U_{ch} \in \mathbb{U}$,

$$\begin{aligned} C_E(A|U_{ch}) = C(A|EU_{ch}) &= \frac{C(AE|U_{ch})}{C(E|U_{ch})} \\ &= \frac{ch(AE)}{ch(E)} = ch(A|E), \end{aligned}$$

as required. This proof is basically the same as one given by Pettigrew (2014).

A proof of Proposition 3.2:

Suppose that C satisfies $\text{PP}_{\text{initial}}$ and $C_{\mathbb{E}}$ is the credence function updated from C by JC on E. As is well known, it is equivalent to JC that: For any $A \in \mathcal{F}$ and $E_i \in \mathbb{E}$,

$$C(A|E_i) = C_{\mathbb{E}}(A|E_i).$$

Then, we have that: For any $E_i \in \mathbb{E}$ and $U_{ch} \in \mathbb{U}$,

$$C_{\mathbb{E}}(A|U_{ch}E_i) = \frac{C_{\mathbb{E}}(AU_{ch}|E_i)}{C_{\mathbb{E}}(U_{ch}|E_i)} = \frac{C(AU_{ch}|E_i)}{C(U_{ch}|E_i)} = C(A|E_iU_{ch}) = ch(A|E_i),$$

as required.

A proof of Proposition 3.3:

Suppose that $\mathbb{E} = \{E, \neg E\}$ and $C(E) \neq C_{\mathbb{E}}(E)$. Suppose also that C satisfies $\text{PP}_{\text{initial}}$ and $C_{\mathbb{E}}$ is the credence function updated from C by JC on $\mathbb{E}$. Suppose even, *for reductio*, that $C_{\mathbb{E}}$ satisfies PP'. Then, we have that:

$$\begin{aligned} ch(E) = C_{\mathbb{E}}(E|U_{ch}) &= \frac{C_{\mathbb{E}}(EU_{ch})}{C_{\mathbb{E}}(U_{ch})} \\ &= \frac{\beta C(E|U_{ch})}{\beta C(E|U_{ch}) + C(\neg E|U_{ch})} \\ &= \frac{\beta ch(E)}{\beta ch(E) + ch(\neg E)}, \end{aligned}$$

where $\beta = \frac{C_{\mathbb{E}}(E)/C_{\mathbb{E}}(\neg E)}{C(E)/C(\neg E)}$. Now, it follows from the above equation that $\beta = 1$, which contradicts the above assumption that $C_{\mathbb{E}}(E) \neq C(E)$. Thus, there is a case

in which $C_{\mathbb{E}}$ does not satisfy PP' even if C satisfies $\text{PP}_{\text{initial}}$ and $C_{\mathbb{E}}$ is the credence function updated from C by JC on $\mathbb{E}$. Done.

A proof of Proposition 3.4:

Suppose that C and C_{E}, respectively, satisfy $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{certainty}}$. Suppose also that C_{E} is *locally* updated from C by SC on E relative to $\{U_{ch}\}$. That is, it holds that $C_{E}(U_{ch}) = C(U_{ch}|E)$ for any $U_{ch} \in \mathbb{U}$. Note that $\text{PP}_{\text{initial}}$ entails that: For any $A \in \mathcal{F}$, $U_{ch} \in \mathbb{U}$, and $E \in \mathcal{F}$,

$$C(A|U_{ch}E) = ch(A|E).$$

Then, we have that: For any $A \in \mathcal{F}$,

$$\begin{aligned} C_{E}(A) &= \sum_{ch} C_{E}(U_{ch})C_{E}(A|U_{ch}) \\ &= \sum_{ch} C(U_{ch}|E)ch(A|E) = \sum_{ch} C(U_{ch}|E)C(A|U_{ch}E) \\ &= \sum_{ch} C(AU_{ch}|E) = C(A|E), \end{aligned}$$

as required.

A proof of Proposition 3.5:

Suppose that C and $C_{\mathbb{E}}$, respectively, satisfy $\text{PP}_{\text{initial}}$ and $\text{PP}_{\text{uncertainty}}$. Suppose also that $C_{\mathbb{E}}$ is *locally* updated from C by JC on $\mathbb{E}$ relative to $\{U_{ch}E_i\}$. Then, we have that: For any $E_k \in \mathbb{E}$ and $U_{ch} \in \mathbb{U}$,

$$\begin{aligned} C_{\mathbb{E}}(E_kU_{ch}) &= \sum_{E_i} C_{\mathbb{E}}(E_i)C(E_kU_{ch}|E_i) \\ &= C_{\mathbb{E}}(E_k)C(E_kU_{ch}|E_k) + \sum_{E_i \in \{E_i : E_i \neq E_k\}} C_{\mathbb{E}}(E_i)C(E_kU_{ch}|E_i) \\ &= C_{\mathbb{E}}(E_k)C(E_kU_{ch}|E_k) = C_{\mathbb{E}}(E_k)C(U_{ch}|E_k). \end{aligned}$$

Now, it follows from the above assumptions and result that: For any $A \in \mathcal{F}$,

$$\begin{aligned}
C_{\mathbb{E}}(A) &= \sum_{E_i, ch} C_{\mathbb{E}}(E_i U_{ch}) C_{\mathbb{E}}(A|E_i U_{ch}) \\
&= \sum_{E_i, ch} C_{\mathbb{E}}(E_i) C(U_{ch}|E_i) C_{\mathbb{E}}(A|E_i U_{ch}) \\
&= \sum_{E_i} \sum_{ch} C_{\mathbb{E}}(E_i) C(U_{ch}|E_i) C(A|E_i U_{ch}) \\
&= \sum_{E_i} C_{\mathbb{E}}(E_i) \sum_{ch} C(AU_{ch}|E_i) = \sum_{E_i} C_{\mathbb{E}}(E_i) C(A|E_i),
\end{aligned}$$

as required.

A derivation of FCI from PP*$_{\text{initial}}$ and GAC:

Suppose that $\mathfrak{C}_{\mathbb{CH}}$ satisfies PP*$_{\text{initial}}$, and that $\mathfrak{C}_{\mathbb{CH}}$ is updated from $\mathfrak{C}$ by means of GAC. Then we have that: For any $A \in \mathcal{F}$,

$$\begin{aligned}
\mathfrak{C}_{\mathbb{CH}}(A) &= \sum_{w, ch} \mathfrak{C}(U_{ch}) \mathfrak{C}(A|wU_{ch}) \mathfrak{C}_{\mathbb{CH}}(w|U_{ch}) \\
&= \sum_{w, ch} \mathfrak{C}(U_{ch}) \mathfrak{C}(A|wU_{ch}) ch(w) \\
&= \sum_{ch} \mathfrak{C}(U_{ch}) \sum_{w} \mathfrak{C}(A|wU_{ch}) ch(w) \\
&= \sum_{ch} \mathfrak{C}(U_{ch}) \sum_{w \in A} \mathfrak{C}(A|wU_{ch}) ch(w) \\
&\quad + \sum_{ch} \mathfrak{C}(U_{ch}) \sum_{w \in \neg A} \mathfrak{C}(A|wU_{ch}) ch(w) \\
&= \sum_{ch} \mathfrak{C}(U_{ch}) \sum_{w \in A} ch(w) = \sum_{ch} \mathfrak{C}(U_{ch}) ch(A).
\end{aligned}$$

as required.

A derivation of FCC from PP*$_{\text{certainty}}$ and GAC:

Suppose that $\mathfrak{C}_{E,\mathbb{CH}}$ satisfies PP*$_{\text{certainty}}$, and that $\mathfrak{C}_{E,\mathbb{CH}}$ is updated from $\mathfrak{C}_E$ by means of GAC. Then we have that: For any $A \in \mathcal{F}$,

$$\begin{aligned}
C_{E,\mathbb{CH}}(A) &= \sum_{w,ch} C_E(U_{ch})C_E(A|wU_{ch})C_{E,\mathbb{CH}}(w|U_{ch}) \\
&= \sum_{w,ch} C_E(U_{ch})C_E(A|wU_{ch})ch(w|E) \\
&= \sum_{ch} C_E(U_{ch}) \sum_{w} C_E(A|wU_{ch})ch(w|E) \\
&= \sum_{ch} C_E(U_{ch}) \sum_{w\in A} C_E(A|wU_{ch})ch(w|E) \\
&\quad + \sum_{ch} C_E(U_{ch}) \sum_{w\in \neg A} C_E(A|wU_{ch})ch(w|E) \\
&= \sum_{ch} C_E(U_{ch}) \sum_{w\in A} ch(w|E) = \sum_{ch} C_E(U_{ch})ch(A|E),
\end{aligned}$$

as required.

A derivation of FCU from PP*$_{\text{uncertainty}}$ and GAC:

Suppose that $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ satisfies PP*$_{\text{uncertainty}}$, and that $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$ is updated from $\mathfrak{C}_{\mathbb{E}}$ by means of GAC. Then we have that: For any $A \in \mathcal{F}$,

$$\begin{aligned}
\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) &= \sum_{w,ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)\mathfrak{C}_{\mathbb{E}}(A|wU_{ch}E_i)\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(w|U_{ch}E_i) \\
&= \sum_{w,ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)\mathfrak{C}_{\mathbb{E}}(A|wU_{ch}E_i)ch(w|E_i) \\
&= \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i) \sum_{w} \mathfrak{C}_{\mathbb{E}}(A|wU_{ch}E_i)ch(w|E_i) \\
&= \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i) \sum_{w\in A} \mathfrak{C}_{\mathbb{E}}(A|wU_{ch}E_i)ch(w|E_i) \\
&\quad + \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i) \sum_{w\in \neg A} \mathfrak{C}_{\mathbb{E}}(A|wU_{ch}E_i)ch(w|E_i) \\
&= \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i) \sum_{w\in A} ch(w|E_i) = \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)ch(A|E_i),
\end{aligned}$$

as required.

A proof of Proposition 4.3.1:

Suppose that our credences are updated by means of SC, FCI, and FCC. Then, it follows from FCC and the probability calculus that, for any $A \in \mathcal{F}$,

$$\begin{aligned}
\mathfrak{C}_{E,\mathbb{CH}}(A) &= \sum_{ch} \mathfrak{C}_E(U_{ch}) ch(A|E) \\
&= \sum_{ch} \mathfrak{C}_E(U_{ch}) \sum_{w \in A} ch(w|E) \\
&= \sum_{ch} \mathfrak{C}_E(U_{ch}) \sum_{w \in A} \mathfrak{C}_E(A|wU_{ch}) ch(w|E) \\
&\quad + \sum_{ch} \mathfrak{C}_E(U_{ch}) \sum_{w \in \neg A} \mathfrak{C}_E(A|wU_{ch}) ch(w|E) \\
&= \sum_{ch} \mathfrak{C}_E(U_{ch}) \sum_{w} \mathfrak{C}_E(A|wU_{ch}) ch(w|E) \\
&= \sum_{ch,w} \mathfrak{C}_E(U_{ch}) \mathfrak{C}_E(A|wU_{ch}) ch(w|E). \qquad \text{(4.3.1a)}
\end{aligned}$$

Similarly, it follows from from FCI and the probability calculus that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A) = \sum_{ch,w} \mathfrak{C}(U_{ch}) \mathfrak{C}(A|wU_{ch}) ch(w). \qquad \text{(4.3.1b)}$$

Then, (4.3.1a) entails that when $w^* \in \Omega$ and $U_{ch'} \in \mathbb{U}$,

$$\begin{aligned}
\mathfrak{C}_{E,\mathbb{CH}}(w^* U_{ch'}) &= \sum_{ch,w} \mathfrak{C}_E(U_{ch}) \mathfrak{C}_E(w^* U_{ch'}|wU_{ch}) ch(w|E) \\
&= \sum_{ch \in \{ch: ch \neq ch'\}, w \in \{w: w \neq w^*\}} \mathfrak{C}_E(U_{ch}) \mathfrak{C}_E(w^* U_{ch'}|wU_{ch}) ch(w|E) \\
&\quad + \mathfrak{C}_E(U_{ch'}) \mathfrak{C}_E(w^* U_{ch'}|w^* U_{ch'}) ch^*(w^*|E) \\
&= \mathfrak{C}_E(U_{ch'}) ch^*(w^*|E),
\end{aligned}$$

and, that when $U_{ch'} \in \mathbb{U}$,

$$\begin{aligned}\mathfrak{C}_{E,\mathbb{CH}}(U_{ch'}) &= \sum_{ch,w} \mathfrak{C}_E(U_{ch})\mathfrak{C}_E(U_{ch'}|wU_{ch})ch(w|E) \\ &= \sum_{ch\in\{ch:ch\neq ch'\}} \sum_{w} \mathfrak{C}_E(U_{ch})\mathfrak{C}_E(U_{ch'}|wU_{ch})ch(w|E) \\ &\quad + \sum_{w} \mathfrak{C}_E(U_{ch'})\mathfrak{C}_E(U_{ch'}|wU_{ch'})ch^{'}(w|E) \\ &= \sum_{w} \mathfrak{C}_E(U_{ch'})ch'(w|E) = \mathfrak{C}_E(U_{ch'}).\end{aligned}$$

Thus, we have that, for any $w \in \Omega$ and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{E,\mathbb{CH}}(w|U_{ch}) = ch(w|E),$$

and so that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{E,\mathbb{CH}}(A|U_{ch}) = \sum_{w\in A} \mathfrak{C}_{E,\mathbb{CH}}(w|U_{ch}) = \sum_{w\in A} ch(w|E) = ch(A|E). \qquad \text{(4.3.1c)}$$

In a similar way, it follows from (4.3.1b) that when $w^* \in \Omega$ and $U_{ch'} \in \mathbb{U}$,

$$\mathfrak{C}_{\mathbb{CH}}(w^*U_{ch'}) = \mathfrak{C}(U_{ch'})ch'(w*),$$

and that when $U_{ch'} \in \mathbb{U}$,

$$\mathfrak{C}_{\mathbb{CH}}(U_{ch'}) = \mathfrak{C}(U_{ch'}).$$

Thus, we have that, for any $w \in \Omega$ and $U_{ch} \in \mathbb{U}$,

$$\mathfrak{C}_{\mathbb{CH}}(w|U_{ch}) = ch(w),$$

and so that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A|U_{ch}) = \sum_{w\in A} \mathfrak{C}_{\mathbb{CH}}(w|U_{ch}) = \sum_{w\in A} ch(w) = ch(A). \qquad \text{(4.3.1d)}$$

Then, SC and (4.3.1d) entail that, for any $A \in \mathcal{F}$, $U_{ch} \in \mathbb{U}$, and $E \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH},E}(A|U_{ch}) = \mathfrak{C}_{\mathbb{CH}}(A|U_{ch}E) = \frac{\mathfrak{C}_{\mathbb{CH}}(AE|U_{ch})}{\mathfrak{C}_{\mathbb{CH}}(A|U_{ch})} = ch(A|E). \qquad \text{(4.3.1e)}$$

Finally, with the help of (4.3.1c) and (4.3.1e), we obtain Proposition 4.3.1. Done.

A proof of Proposition 4.3.2:

Suppose that our credences are updated by means of JC, FCI, and FCU. Recall the equation (4.3.1a) appearing in the proof of 4.3.1. In the very similar way, we have, with the help of FCU and the probability calculus, that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) = \sum_{w,ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)\mathfrak{C}_{\mathbb{E}}(A|wU_{ch}E_i)ch(w|E_i).$$

This entails that when $w^* \in \Omega$, $U_{ch'} \in \mathbb{U}$, and $E_k \in \mathbb{E}$,

$$\begin{aligned}\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(w^*U_{ch'}E_k) &= \sum_{w,ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)\mathfrak{C}_{\mathbb{E}}(w^*U_{ch'}E_k|wU_{ch}E_i)ch(w|E_i)\\ &= \mathfrak{C}_{\mathbb{E}}(U_{ch}E_k)\mathfrak{C}_{\mathbb{E}}(w^*U_{ch'}E_k|w^*U_{ch'}E_k)ch'(w^*|E_k)\\ &= \mathfrak{C}_{\mathbb{E}}(U_{ch}E_k)ch'(w^*|E_k),\end{aligned}$$

and that when $U_{ch'} \in \mathbb{U}$ and $E_k \in \mathbb{E}$,

$$\begin{aligned}\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(U_{ch'}E_k) &= \sum_{w,ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)\mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k|wU_{ch}E_i)ch(w|E_i)\\ &= \sum_{w}\left(\sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)\mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k|wU_{ch}E_i)ch(w|E_i)\right)\\ &= \sum_{w} \mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k)\mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k|wU_{ch'}E_k)ch'(w|E_k)\\ &= \sum_{w} \mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k)ch'(w|E_k)\\ &= \mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k)\sum_{w} ch'(w|E_i) = \mathfrak{C}_{\mathbb{E}}(U_{ch'}E_k).\end{aligned}$$

Thus, we have that, for any $w \in \Omega$, $U_{ch} \in \mathbb{U}$, $E_i \in \mathbb{E}$,

$$\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(w|U_{ch}E_i) = ch(w|E_i),$$

and so that, for any A,

$$\begin{aligned}\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A|U_{ch}E_i) &= \sum_{w\in A}\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(w|U_{ch}E_i)\\ &= \sum_{w\in A} ch(w|E_i) = ch(A|E_i).\end{aligned} \tag{4.3.2a}$$

On the other hand, as explained in the proof of 4.3.1, it holds that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A|U_{ch}) = ch(A).$$

Then, this, together with JC*(or JC), entails that, for any $A \in \mathcal{F}$, $U_{ch} \in \mathbb{U}$, and $E_k \in \mathbb{E}$,

$$\begin{aligned}\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A|U_{ch}E_k) &= \frac{\sum_{E_i}\beta_i\mathfrak{C}_{\mathbb{CH}}(AU_{ch}E_kE_i)}{\sum_{E_i}\beta_i\mathfrak{C}_{\mathbb{CH}}(U_{ch}E_kE_i)} = \frac{\beta_k\mathfrak{C}_{\mathbb{CH}}(AU_{ch}E_k)}{\beta_k\mathfrak{C}_{\mathbb{CH}}(U_{ch}E_k)}\\ &= \frac{\mathfrak{C}_{\mathbb{CH}}(AE_k|U_{ch})}{\mathfrak{C}_{\mathbb{CH}}(E_k|U_{ch})} = ch(A|E_k),\end{aligned} \tag{4.3.2b}$$

where $\beta_i = \frac{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(E_i)/\mathfrak{C}_{\mathbb{CH}}(E_i)}{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(E_1)/\mathfrak{C}_{\mathbb{CH}}(E_1)}$. Finally, with the help of (4.3.2a) and (4.3.2b), we obtain Proposition 4.3.2. Done.

Several calculations related to Figure 1:

With the help of FCC, SC, and FCI, we have that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A) = \sum_{ch}\mathfrak{C}(U_{ch})ch(A), \tag{F1a}$$

$$\mathfrak{C}_{\mathbb{CH},E}(A) = \mathfrak{C}_{\mathbb{CH}}(A|E) = \frac{\mathfrak{C}_{\mathbb{CH}}(AE)}{\mathfrak{C}_{\mathbb{CH}}(E)} = \frac{\sum_{ch}\mathfrak{C}(U_{ch})ch(AE)}{\sum_{ch}\mathfrak{C}(U_{ch})ch(E)}, \tag{F1b}$$

$$\mathfrak{C}_E(A) = \mathfrak{C}(A|E), \text{ and} \tag{F1c}$$

$$\mathfrak{C}_{E,\mathbb{CH}}(A) = \sum_{ch}\mathfrak{C}_E(U_{ch})ch(A|E) = \sum_{ch}\mathfrak{C}(U_{ch}|E)ch(A|E) \tag{F1d}$$

As assumed in the example related to Figure 1, $\mathbb{U} = \{U_{ch'}, U_{ch^*}\}$. Then, it follows from (F1a)~ (F1d) that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A) = \mathfrak{C}(U_{ch'})ch'(A) + \mathfrak{C}(U_{ch^*})ch^*(A), \tag{F1a*}$$

$$\mathfrak{C}_{\mathbb{CH},E}(A) = \frac{\mathfrak{C}(U_{ch'})ch'(AE) + \mathfrak{C}(U_{ch^*})ch^*(AE)}{\mathfrak{C}(U_{ch'})ch'(E) + \mathfrak{C}(U_{ch^*})ch^*(E)}, \tag{F1b*}$$

$$\mathfrak{C}_E(A) = \frac{\mathfrak{C}(AE)}{\mathfrak{C}(E)}, \text{ and} \tag{F1c*}$$

$$\mathfrak{C}_{E,\mathbb{CH}}(A) = \mathfrak{C}(U_{ch'}|E)ch'(A|E) + \mathfrak{C}(U_{ch^*}|E)ch^*(A|E). \tag{F1d*}$$

Note that $E \equiv w_1 \vee w_2 \vee w_4 \vee w_5$, $U_{ch'} \equiv w_1 \vee w_2 \vee w_3$, and $U_{ch^*} \equiv w_4 \vee w_5 \vee w_6$. With the help of (F1a*)~(F1d*), the chance assignments of ch' and ch^*, and the initial credence assignment of $\mathfrak{C}$, then, we have that, for any $A \in \mathcal{F}$,

$$\begin{aligned}
\mathfrak{C}_{\mathbb{CH}}(A) &= \frac{1}{2}ch'(A) + \frac{1}{2}ch^*(A), \\
\mathfrak{C}_{\mathbb{CH},E}(A) &= \frac{12}{17}\left(ch'(AE) + ch^*(AE)\right), \\
\mathfrak{C}_E(A) &= \frac{10}{7}\mathfrak{C}(AE), \text{ and} \\
\mathfrak{C}_{E,\mathbb{CH}}(A) &= \frac{9}{14}ch'(AE) + \frac{16}{21}ch^*(AE).
\end{aligned}$$

Now, we can derive the probability assignments of $\mathfrak{C}_{\mathbb{CH}}$, $\mathfrak{C}_{\mathbb{CH},E}$, $\mathfrak{C}_E$, and $\mathfrak{C}_{E,\mathbb{CH}}$. For example,

$$\begin{aligned}
\mathfrak{C}_{\mathbb{CH}}(w_1) &= \frac{1}{2}ch'(w_1) + \frac{1}{2}ch^*(w_1) = \frac{1}{2}ch'(w_1) = \frac{1}{6}; \\
\mathfrak{C}_{\mathbb{CH},E}(w_1) &= \frac{12}{17}\left(ch'(w_1E) + ch^*(w_1E)\right) = \frac{12}{17}ch'(w_1) = \frac{4}{17}; \\
\mathfrak{C}_E(w_1) &= \frac{10}{7}\mathfrak{C}(w_1E) = \frac{10}{7}\mathfrak{C}(w_1) = \frac{1}{7}; \text{ and} \\
\mathfrak{C}_{E,\mathbb{CH}}(w_1) &= \frac{9}{14}ch'(w_1E) + \frac{16}{21}ch^*(w_1E) = \frac{9}{14}ch'(w_1) = \frac{3}{14}.
\end{aligned}$$

These results conform with the probability assignments in Figure 1.

Several calculations related to Figure 2:

Suppose that $\beta^{\mathfrak{C}_{\mathbb{E}}}_{\mathfrak{C}}(E_i, E_1) = \beta^{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}}_{\mathfrak{C}_{\mathbb{CH}}}(E_i, E_1) = \beta_i$ for any $E_i \in \mathbb{E}$. Then, we have, with the help of FCU, JC* (or JC), and FCI, that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A) = \sum_{ch} \mathfrak{C}(U_{ch})ch(A), \tag{F2a}$$

$$\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A) = \frac{\sum_{E_i} \beta_i \mathfrak{C}_{\mathbb{CH}}(AE_i)}{\sum_{E_i} \beta_i \mathfrak{C}_{\mathbb{CH}}(E_i)} = \frac{\sum_{E_i,ch} \beta_i \mathfrak{C}(U_{ch})ch(AE_i)}{\sum_{E_i,ch} \beta_i \mathfrak{C}(U_{ch})ch(E_i)}, \tag{F2b}$$

$$\mathfrak{C}_{\mathbb{E}}(A) = \frac{\sum_{E_i} \beta_i \mathfrak{C}(AE_i)}{\sum_{E_i} \beta_i \mathfrak{C}(E_i)}, \text{ and} \tag{F2c}$$

$$\begin{aligned}
\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) &= \sum_{ch,E_i} \mathfrak{C}_{\mathbb{E}}(U_{ch}E_i)ch(A|E_i) \\
&= \sum_{ch,E_i} \frac{\sum_{E_j} \beta_j \mathfrak{C}(U_{ch}E_iE_j)}{\sum_{E_j} \beta_j \mathfrak{C}(E_j)} ch(A|E_i) \\
&= \sum_{ch,E_i} \frac{\beta_i \mathfrak{C}(U_{ch}E_i)}{\sum_{E_j} \beta_j \mathfrak{C}(E_j)} ch(A|E_i) \\
&= \frac{\sum_{ch,E_i} \beta_i \mathfrak{C}(U_{ch}E_i)ch(A|E_i)}{\sum_{E_j} \beta_j \mathfrak{C}(E_j)}
\end{aligned} \tag{F2d}$$

As assumed in the example related to Figure 2, $\mathbb{E} = \{E, \neg E\}$. Suppose that $\beta^{\mathfrak{C}_{\mathbb{E}}}_{\mathfrak{C}}(E, \neg E) = \beta^{\mathfrak{C}_{\mathbb{CH},\mathbb{E}}}_{\mathfrak{C}_{\mathbb{CH}}}(E, \neg E) = \beta$. Then, it follows from (F2a)~(F2d) that, for any $A \in \mathcal{F}$,

$$\mathfrak{C}_{\mathbb{CH}}(A) = \mathfrak{C}(U_{ch'})ch'(A) + \mathfrak{C}(U_{ch^*})ch^*(A), \tag{F2a*}$$

$$\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A) = \frac{\sum_{ch} (\beta\mathfrak{C}(U_{ch})ch(AE) + \mathfrak{C}(U_{ch})ch(A\neg E))}{\sum_{ch} (\beta\mathfrak{C}(U_{ch})ch(E) + \mathfrak{C}(U_{ch})ch(\neg E))}, \tag{F2b*}$$

$$\mathfrak{C}_{\mathbb{E}}(A) = \frac{\beta\mathfrak{C}(AE) + \mathfrak{C}(A\neg E)}{\beta\mathfrak{C}(E) + \mathfrak{C}(\neg E)}, \text{ and} \tag{F2c*}$$

$$\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) = \frac{\sum_{ch} (\beta\mathfrak{C}(U_{ch}E)ch(A|E) + \mathfrak{C}(U_{ch}\neg E)ch(A|\neg E))}{\beta\mathfrak{C}(E) + \mathfrak{C}(\neg E)}. \tag{F2d*}$$

Note that $\mathbb{U} = \{U_{ch'}, U_{ch^*}\}$, $E \equiv w_1 \vee w_2 \vee w_4 \vee w_5$, $U_{ch'} \equiv w_1 \vee w_2 \vee w_3$, and $U_{ch^*} \equiv w_4 \vee w_5 \vee w_6$. Note also that it is assumed that $\beta = 2$. Then, we have, with the help of (F2a*)~(F2d*), the chance assignments of ch' and ch^*, and the initial credence assignment of $\mathfrak{C}$, that, for any $A \in \mathcal{F}$,

$$\begin{aligned}
\mathfrak{C}_{\mathbb{CH}}(A) &= \frac{1}{2}ch'(A) + \frac{1}{2}ch^*(A), \\
\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(A) &= \frac{24}{41}ch'(AE) + \frac{12}{41}ch'(A\neg E) + \frac{24}{41}ch^*(AE) + \frac{12}{41}ch^*(A\neg E), \\
\mathfrak{C}_{\mathbb{E}}(A) &= \frac{20}{17}\mathfrak{C}(AE) + \frac{10}{17}\mathfrak{C}(A\neg E), \text{ and} \\
\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(A) &= \frac{9}{17}ch'(AE) + \frac{6}{17}ch'(A\neg E) + \frac{32}{51}ch^*(AE) + \frac{4}{17}ch^*(A\neg E).
\end{aligned}$$

Now, we can derive the probability assignments of $\mathfrak{C}_{\mathbb{CH}}$, $\mathfrak{C}_{\mathbb{CH},\mathbb{E}}$, $\mathfrak{C}_{\mathbb{E}}$, and $\mathfrak{C}_{\mathbb{E},\mathbb{CH}}$. For example,

$$\begin{aligned}
\mathfrak{C}_{\mathbb{CH}}(w_1) &= \frac{1}{2}ch'(w_1) + \frac{1}{2}ch^*(w_1) = \frac{1}{2}ch'(w_1) = \frac{1}{6}; \\
\mathfrak{C}_{\mathbb{CH},\mathbb{E}}(w_1) &= \frac{24}{41}ch'(w_1) + \frac{24}{41}ch^*(w_1) = \frac{24}{41}ch'(w_1) = \frac{8}{41}; \\
\mathfrak{C}_{\mathbb{E}}(w_1) &= \frac{20}{17}\mathfrak{C}(w_1E) + \frac{10}{17}\mathfrak{C}(w_1\neg E) = \frac{20}{17}\mathfrak{C}(w_1) = \frac{2}{17}; \textit{ and} \\
\mathfrak{C}_{\mathbb{E},\mathbb{CH}}(w_1) &= \frac{9}{17}ch'(w_1E) + \frac{32}{51}ch^*(w_1E) = \frac{9}{17}ch'(w_1) = \frac{3}{17}.
\end{aligned}$$

These results conform with the probability assignments in Figure 2.

References

Bradley, Richard (2005). Radical probabilism and bayesian conditioning. *Philosophy of Science* 72 (2):342-364.

Douven, Igor & Romeijn, Jan-Willem (2011). A new resolution of the Judy Benjamin Problem. *Mind* 120 (479):637-670.

Field, Hartry (1978). A note on Jeffrey conditionalization. *Philosophy of Science* 45 (3):361-367.

Hall, Ned (1994). Correcting the guide to objective chance. *Mind* 103 (412):505-518.

Hall, Ned (2004). Two mistakes about credence and chance. *Australasian Journal of Philosophy* 82 (1):93 – 111.

Harper, William L. & Kyburg, Henry E. (1968). The Jones case. *British Journal for*

the Philosophy of Science 19 (3):247-251.
Ismael, Jenann (2008). Raid! Dissolving the big, bad bug. *Noûs* 42 (2):292–307.
Jeffrey, Richard C. (1970). Dracula Meets Wolfman: Acceptance vs. Partial Belief. In Marshall Swain (ed.), *Induction, Acceptance, and Rational Belief*. D. Reidel. pp. 157-185.
Jeffrey, Richard (2004). *Subjective Probability: The Real Thing*. Cambridge University Press.
Levi, Isaac (1967). Probability kinematics. *British Journal for the Philosophy of Science* 18 (3):197-209.
Lewis, David (1980). A subjectivist's guide to objective chance. In Richard C. Jeffrey (ed.), *Studies in Inductive Logic and Probability*. University of California Press. pp. 83–132.
Lewis, David (1994). Humean Supervenience Debugged. *Mind* 103 (412):473-490.
Meacham, Christopher J. G. (2010). Two mistakes regarding the principal principle. *British Journal for the Philosophy of Science* 61 (2):407-431.
Nissan-Rozen, Ittay (2013). Jeffrey Conditionalization, the Principal Principle, the Desire as Belief Thesis, and Adams's Thesis. *British Journal for the Philosophy of Science* 64 (4):837-850.
Pettigrew, Richard (2013). What Chance-Credence Norms Should Not Be. *Noûs* 47 (3):177-196.
Thau, Michael (1994). Undermining and admissibility. *Mind* 103 (412):491-504.
Wagner, Carl G. (2003). Commuting probability revisions: The uniformity rule. *Erkenntnis* 59 (3):349-364.

7th Asia Pacific Conference on the Philosophy of Science, Chiayi, Taiwan (15-16 Dec. , 2017)

The Engagement of Kalam in Modern Science: A physicist view

Shahidan Radiman, Phd

School of Physics an Novel Materials,

Faculty of Science and Technology

Universiti Kebangsaan Malaysia

Bangi 43600 , Selangor DE

MALAYSIA

E-mail: shahidan@ukm.edu.my

"The concept of a law of Nature cannot be made sense of without God "- Nancy Cartwright in her paper entitled "No God, no laws".

1. Introduction

Kalam or Islamic Rational Science was established during the Umayyad (8th C AD) and reached its height during the Abbasid Caliphate period (10th C AD) in an attempt to understand aspects of the Islamic faith by logical reasoning in main due to the influence of Greek rational thought on Muslim philosophers especially in Basra and Baghdad, both in Iraq . Two main branch of Kalam were developed namely Jalil al-Kalam (Divine attributes and actions) and Daqiq al-Kalam (rational science). Two main school of Kalam emerged namely the Muktazilites and the Asharites. Despite the different views by the Kalam philosophers (Mutakallimun) they all subscribed to some common basic principles in understanding nature. Here we discussed 5 of these principles which influenced the philosophy of modern science, especially reductionism. The dichotomy of jawhar (substance) and a'radh (accident) can be seen to perpetuate into the dichotomy of ontic and epistemic modalities found in modern quantum theory. Discreteness of natural structures including space and time which is one of the main Kalam principle played important role in quantum gravity theories either via causal sets or pregeometry and braneworld models. Whereas the Mutakallimun were once divided about the existence of the Universe into ex niliho vs pre-existing , current cosmologists are equally divided into this category as well with new understanding on Multiverse

scenario. Some aspects of modern physics concepts had already been proposed the Mutakallimun some of which will be discussed in this paper.

Looking back in history, Kalam Jadid or the New Dialectics intellectual movement was initiated by Al Ghazali and matured by Fakhr al Din al-Razi. As Adisetia (1) mentioned " This historical success provides pertinent lesson for Muslim scholars and intellectuals today to formulate what can be called Kalam al 'Asr or Dialectics of the Age to take on the challenge and allure of modern science ". The Quran exhorts believers to look into the cosmic horizon (al afaq) and into their own selves (al anfus) for empirical and experiental evidence revealing the truth (al haqq) – see Chapter Fussilat (41): 53 in the Quran) . Traditionally the Kalam approach took the first part (studying the Universe or macrocosmos) whereas the Sufi approach (as exemplified by the Sufi master, Ibn Arabi) "look into themselves" (man as insanul kamil and microcosmos). In Kalam , the idea that God's creative activity is continuous in the Universe is known as "continuous creation". There are two main views about this. First, God still play an active part in the dynamics and ever changing course of the Universe and in the second view , after creating the Universe and its physical laws , God take rest and led the Universe evolved on its own , into a fated future which only God knows (via His programme). In fact , some people believe that after creation of the Universe in six days (as mentioned in the Quran) God "take a rest or holiday" (just observe it) after He had done all the creation ! Of course Muslims take the first view , because from the Quran we know that God play an active role especially in giving "miracles" to Prophets and even to Friends of God (waliyy). The fact that God is active but must "go against" His own Physical Laws during the occurrence of "miracles" led Al Ghazali to propose the doctrine of Occasionalism. There is a "weaker" version to this called "concurrentism" which states that there are particular events or phenomenon that are produced both by Divine intervention as well as power of finite beings (physical laws are still followed) put together. In this weaker version, causality is still being followed but the result is still a "miracle" (one possibility is called "freak" event). In the strong version of Occasionalism no finite being has a causal role in creation. This doctrine was first formulated by the Asharite in Islamic Kalam which was later echoed by the Cartesians (the philosophers who follow the ideas of Descartes) and famously articulated by Malebranche. In this Asharite cosmology the physical Universe can be categorised into essence **(jawhar**) and accident (**a'rad**) . Jawhar and a'rad form an atom . The accidents can be seen as properties whereas essence can be considered as matter . Classical physics hold to the notion that matter and properties come together

but quantum physics via the Quantum Cheshire Cat effect showed that properties can persist even when the matter (say neutrons) has disintegrated or disappeared. This is the analogue of the Cheshire cat in "Alice in Wonderland" whose grin persists even when the cat is gone. We will see that in the Copenhagen Interpretation of Quantum Mechanics an even stronger version of this was held by Bohr and colleagues called positivism , whereby an object only exist when you measured it. This is in contrast to Einstein (realist) who believed that "the moon is there even though you did not look at it ".

In Asharite Kalam , all essences and accidents (together forming "atoms") need the power of God in order to exist and subsists over time - then, it seems only Occasionalism is the "mechanism" for the Universe to continually exist. There is a strange and mysterious analogy between substance and accidents being separated but entangled particles. This has been demonstrated in the quantum weak measurements whereby it was experimentally proven that the past can affect the future (and vice versa) via time entanglement. The infinitely important role of God (He is All Emcompassing) was pointed out by Al Ghazali by emphasizing that correlations do not imply causal relations . After all , God is Time (a saying in a hadith) and so you don't need to impose "time ordering" in His actions. A curious implication of this is the violation of time reversal operation , T in CPT theorem which says that CPT is conserved and since CP is violated this merely imply that T is violated too. So there are "causal structures" which we do not know now which led to CP violation. In the Asharite Kalam human being acquire their acts (kasb) while God created these acts. Since we know that this is human volition (free will) and human are responsible for their own acts and choices , this led to sin and reward . In this way the acts goes back to God and human are free from any judgement – this is contrary to the purpose of Heaven and Hell. Later, Ibn Humam (an Asharite) deny that human choice falls under the scope of Divine power. After all, in Islam, mankind is the viceregent or deputy of God – he is free to make his own choice at his own peril ! In other words, human choices are relational matters not under the purview of any kind of power. Man is given a free will and God has given him the power of reasoning to guide him. Malebranche once concluded that a belief in secondary causality that is , ascribing causal power to beings other than God leads to paganism. In Islam this is also called shirk khafi (hidden shirk or hidden polytheism). So , occasionalism where God continuously create and destroy and recreate again led to permanency of things giving "occasional causes" which allows events that can violate causal and physical laws in between. In Mulla Sadra (a Muslim philosopher)

"gradation of beings" scenario, man being in the highest hierarchy is given the total free will and choice to act , unlike animals which are given limited choices and thus not accountable for their actions. On the other hand, man's intention is also counted as action such that a good intention alone will be rewarded by God but bad intentions will be counted only after an action has been executed .

2. Philosophy , Kalam and Modern Science

Asharite and Maturidites mutakallims are mediating schools between the rationalism of the Muktazilites and the extreme literalism of the Hanbalites. In Iraq the mediating school was represented by Abu Hassan al Shaari (d.324H) and in Samarqand by Abu Mansur al Maturidi (d. 333 H). Later scholars in the Assharite school were al Baqillani (d. 403 H) , al Juwayni (d), al Ghazali (d.505 H) and al Sanusi , whereas Maturidites were given new vitality in the work of Al Bazdawi (d.493 H). In **Jalil al Kalam** (knowledge of God) which explain the attributes of God , both Sunni schools introduced several common doctrines e.g the doctrines of absolute difference (mukhalafah) , without any quality (bila kayfa) and without drawing any anthropomorphization (bila tasbih) as protective principles against hereticism. Between the two schools there are some controversial problems regarding the nature of God's attributes – further studies showed that the Sufi master, Ibn Arabi had beautifully resolved these problems in many of his major works (but will not be discussed further here).

Historically, Sayf al Din al Amidi was the first to extensively used philosophical arguments and concepts (traditionally laying outside the realm of Kalam) in discussing about the atom in his kitab **Abkar al Afkar** . As Laura Hassan (2) rightly put it " It is increasingly understood , contra the longstanding myth that Ibn Sina (d 428H/1037) islamicised Neoplatonism died at the pen of Al-Ghazali (d.505H/1111) that his philosophy in fact became the subject of analysis and appropriation in the following centuries". So, the new Kalam Jadid has rightly put a roadmap for Kalam al 'Asr by pursuing into the realm of modern scientific philosophy and contributing to Modern science. But first let us have a look at the main principles laid out in Daqiq al- Kalam (3) of the Asharite Mutakallimun :

1. **The creation of the world** – according to the mutakallimun , the world is not eternal but was created at some finite point in past time . This put modern

theories of cyclic cosmology (including ekpyrotic cosmology) to be out of question and basically upholds creation of the Universe ex nihilo or in Andrei Linde's word " tunnelling out of nothing". Previously we have already highlighted the work of Mir Damad (4) who proposed the creation of the Universe out of time i.e wujud dahr.

2. **Discretenesss of natural structures**. The Mutakallimun believed that all entities in the Universe are composed of finite number of fundamental components called **jawhar** ("substance") that is indivisible and has no parts. This seems to be in direct correspondent with Standard Model of physics where these fundamental particles consists of 3 families of lepton-quarks. The jawhar was thought to be an abstract entity that acquires its physical properties and value when occupied by a character called 'ard (accident). Again, this agrees with the Standard Model of physics where "ard is supplied by the Higgs field which gave "mass" to all the particles interacting with it, with the mass "given" to the particles being proportional to the coupling constants of the interactions. In fact , discreteness according to the Mutakallimun applied not only to material bodies but also to space and time (this is in total agreement with quantum theory , where we need to modify the Uncertainty Principle due to discreteness at the Planckscale) including motion and energy (being both quantised). In fact , there has been some proposal that gravity consists of space-time "atomic" condensate .

   In the past centuries the translation of the word " rabbil falaq " occurring in the Quran – Lord of the Daybreak should instead be translated (contextually) as Lord of the orbits (occurring in Quran , Chapter 113). Why? Because we see order and forces in orbits – in galaxies and stars and planetary systems. But more than that orbits of electron in atoms and shell models in nuclei – they are all quantised (both motion and energy) and shell model determined the stability of nuclei (up to the superheavy nuclei recently confirmed by experiments). So, Lord of the Orbits showed that He is active and busy and yet probabilistic too . Discreteness also imply the entropic character of existence, example gravity. There has been several proposals in the literature on entropic and emergent gravitational fields.

3. **Continual re-creation of an ever changing world**. Because God is the Absolute Creator, Ever Living and Ever-Acting but also the Mighty Destroyer,

to sustain the Universe the particles (that form the Universe) have to be destroyed and re-created at every moment . This was later developed into a viable theory of Occasionalism by Al Ghazali and others (already mentioned in previous discussion). The continuous process thus gives a continuous presence of the Universe. Surprisingly, there is a similar idea proposed by GRW (Ghirardi-Rimini-Weber) interpretation of quantum mechanics , also called the flash ontology whereby wavefunctions are collapsed due to observations , but with no mechanism for "re-creation " of the wavefunction .

4. **Indeterminacy of the Universe**. Since God posseses absolute free will and since He is the personal creator and sustainer of the Universe, He is at liberty to take any action He wishes- consequently, the laws of nature has to be probabalisitic (defining His habits) rather than deterministic so that physical values are to be contingent and undetermined. This "indeterminism" however is only from human point of view, not God's. This indeterminism is exactly in full agreement with Quantum theory. In fact, due to "secondary causes" chaotic regimes and deterministic character can also be realised e.g as shown by Bohmian mechanics trajectory-based methods. Base on this indeterminism, Mutakallimun reject the existence of natural causality , a basic assumption of Occasionalism.

5. **Integrity of space and time**. The Mutakallimun has the understanding that space has no meaning of its own without there being a body to conceive the existence of a space. This is in line with Mach Principle which is still being studied classically as well as its quantum gravity version. The existence of space affecting motions of large bodies led to Thirring and Sagnac effects which are significant for spinning blackholes. The connection between space and time is deeply rooted in the Arabic language itself –therefore, neither absolute space nor absolute time in fact exists. Again, this is agreement with modern physics idea that time is emergent (e.g via quantum entanglement between bodies or particles) or that the Universe can in fact exists in higher dimensions (3 space and 1 time dimensions and the rest are compactified), so that what we have (the uncompactified dimensions) are in fact emergent properties of the whole manifold.

So, what we find in these five principles of **Daqiq al-Kalam** are the precursors to basic assumptions in the theories of modern physics , stretching from quantum theory to gravity and particle physics .

From philosophical viewpoint, we saw in old Kalam that their atomism imply reductionism – meaning that we need to investigate the jawhar to know the basic constituents of nature (or the Universe) and reconstructing them back to understand the Universe in a unified (tawhidic) manner since the Creator is afterall only Single. Thus reductionism has been the main philosophy employed in understanding physics, biology and chemistry. It seems now these three major fields of science has converged in the field of Nanotechnology and hard science is slowly merging with soft sciences e.g economy, sociology and arts via the science of Complexity (giving new fields such as Econophysics etc). This is one area of "synthesis" which Kalam can be engaged on within the new enquiry (Kalam Al 'Asr).

3. New Enquiry in the Science of Kalam

We can now agree why Kalam is also a science, since a lot of its assumptions and ideas coincided with modern scientific enquiry both from its methodology as well as conclusions. Kalam need to enquire into the areas that current philosophy of science failed to direct e.g idea of "naturalness" with regards to mass hierarchies in High Energy Physics or ideas of Multiverses (first proposed by Max Tegmark) as well as quantum gravity and dark matter searches. In fact late Asharite, Al-Ghazali is the strongest proponent of the "best possible worlds" created by God (see the book by Ormsby (5).) Harry Cliff (6) in one of his TEDX talk remarked ""In reality, the Higgs field is just slightly on. It's not zero, but it's ten-thousand-trillion times weaker than it's fully on value – a bit like a light switch that got stuck just before the 'off' position. And this value is crucial. If it were a tiny bit different, then there would be no physical structure in the universe. Why the strength of the Higgs field is so ridiculously weak defies understanding".

Also, reductionism has already reached its limits – one of this is the idea that Large Hadron Collider can form miniblackholes by colliding particles. If so, then no more small distance physics or high energy physics can be done in an LHC because all of these particles will collapse into a miniblackhole. A similar question arise when we reach the Planckscale – of course Big Bang never starts at r=0 , but somewhere just

above the Planckscale , but how ? How can Kalam probed and give an answer to this mystery?

More recently, topology has become an important buzzword, not only because of the discovery of many novel materials that are topological (topological insulators , topological superconductors , metamaterials) but also its role in quantum theory (knot theory for example, Calabi-Yau manifolds in superstings and membranes) , structures of the Universe and more recently "topological entanglement". It looks like Kalam has a long way to go into probing these questions but remember, the seed of that enquiry had already been sown. The Muktazilite al-Nazzam had already proposed discrete jumps for microscopic motions which he called tafra' (discrete jump) – the forerunner of quantised motion and this include Majorana fermions and skyrmions that move in specific orbits (quantised energy). God's creation-destruction of the Universe components are reflected in creation and annihilation operators in Quantum Mechanics. In fact one can say analogically that God is also doing Quantum Mechanics of the Universe by solving the Wheeler-De Witt equation for the whole Universe in analogy with Schrodinger equation for a single particle. , perhaps in a Bohmian way !

Altaie (3) believed that " God has not only created this world justly and for a purpose , but has set some built-in mechanism to safeguard that the world remain comprehensible ". Perhaps this was just echoing what has been the main motto of theoretical physics research "the unreasonable effectiveness of mathematics" in physics i.e mathematics must be the tool to understand all the laws of physics. This is also supported by one of the main Kalam proponent, Fakhr al Din al-Razi who emphasized the priority of reason (logic , inference) in all of his Kalam and Usul al-Fiqh works (7). Kalam has a role to play in **tafsir ilmi** (see Osman Bakar's new book in this topic (8)) including quranic cosmogony (9) . Contextuality , the impossibility of assigning a single random variable to represent the outcomes of the same measurement procedure under different experimental conditions can be scrutinised via the principles of Kalam (I have only discuss five but in some studies, there were twelves principles) . In fact, Barros and Oas (10) recently proposed that this shortcoming of using quantum probabilities can be solved by introducing negative quasi-probability distributions. There are several ongoing challenges, for example recently Feintzeig proved that ontological model framework failed to represent even the most well known interpretation of Quantum Mechanics (11) and Pienaar (12) showed that there is a lot more to do on causal structure studies using graphs. Of course then, up to now no theory can explain the detail structure of vacuum.

The physics of life via say Quantum Biology is another area where Kalam can work alongside modern physics to explore at least starting with some basic principles. We have also recently work in this field using a Bohmian mechanics approach (13). The epistemological aspect of Ibn Haytham's (another late Mutakallim) scientific thought was only recently explored (14) which add to the credibility of using Kalam approach for synthesising novel scientific knowledge with that of traditional enquiry.

4. Conclusions

In Al-Ghazali's view (as mentioned in the **Ihya'** and **Jawhar al Quran –** see for example Treiger (15) , the highest of the theoretical science is the "science of the unveiling". In modern terms, we can ask the question "what is the purpose of studying science if it doesn't help you in knowing God? "

There are many challenges in the frontier areas of physics such as unifying General relativity of Einstein with quantum theory which allows us to understand the Universe and its evolution. Then there are challenges in the material physics especially at the nanoscale- this is also an area where chemistry and biology meet and a possible emergence of Quantum Biology. There is then a new science born from physics called Complexity which tries to understand the soft sciences using quantitative tools –like internet traffic, macro and microeconomics and financial systems as well as human social and economic activities. Kalam has indeed some philosophical and guiding principles to offer in many of these areas of research and fundamental enquiry.

Kalam's new enquiry will need to explore the human mind and their psychologies as well as to probe deep into fundamental language (semiotics) that underlie common programmes that indeed programmed the Universe, the DNA and social biota . But even, the studies on Kalam (both old and new) have not been wide and deep enough (many Kalam writings are still in manuscript forms) for anyone to extract some of their important theoretical findings. In fact, as a matter of principle we can flashback and later move forward as proposed by Adisetia (16) to study Al-Ghazali and Fakhr al Din al-Razi both for inspiration and systematics that has splendoured Kalam science in the age of Islamic Science which later inspired Europe to transform itself into the Renaisance period.

5. References

1. Adisetia Md Dom , 2012. Kalam Jadid , Islamisation and the Worldview of Islam: operationalising the neo-Ghazalian-Attasian vision , Islam and Science (Summer 2012) (pp 1-23)
2. Laura Hassan , 2014. The encounter of Falsafa and Kalam in Sayf al Din al Amidi's discussion of the atom : Asserting traditional boundaries , questioning traditional doctrines, SOAS Journal of Postgraduate Research Vol (6), 77- 96.
3. Basil Altaie . 2017. God , Nature and the Cause , Pub. (ISSI , Putrajaya, Malaysia)
4. Shahidan Radiman, 2007. Some Aspects of Islamic Cosmology and the current state of Physics Shahidan Radiman , Proceeding of the Int. Conf. on Mathematical Sc. ICMS 2007 (29 Nov.2007)
5. E.L Ormsby , Theodicity in Islamic Thought , Princeton Univ.Press (1984)
6. Quoted in http://www.businessinsider.my/the-end-of-physics-as-we-know-it-2016-1/?r=US&IR=T
7. Mohd Farid Mohd Shahran , 2015. The priority of rational proof in Islam : The view of Fakhr al Din al Razi , TAFHIM (IKIM Journal of Islam and the Contemporary World 8, 1-18.
8. Osman Bakar , 2017. Quranic Pictures of the Universe , Islamic Book Trust and UBD Press.
9. Haslin Hasan and Ab. Hafiz Mat Tuah, 2014. Quranic cosmogony : Impact of Contemporary Cosmology on the Interpretation of Quranic Passages Relating to the origin of the Universe , Kyoto Bulletin of Islamic Area studies , 7 (March 2014), 124-140.
10. J.A de Barros and G. Oas , 2015. Some examples of contextuality in physics : Implications to quantum cognition , World Scientific Review (May 2015) , Chapter 1 (pp 1 -25)
11. B. Feintzeig , 2014. Can the ontological models framework accommodate Bohmian Mechanics?
12. J.Pienaar , 2017. Which causal structures might support a quantum-classical gap? , New J of Physics 19 , 043021
13. Wan Qashishah Akmal W.Razali , Shahidan Radiman , Hishamuddin Zainuddin and Siti Norafidah M.Ramli, 2017. Stochastic Process description via interpretation to a single molecular rotor, Int J Adv Res. 5 (5) , 389-395 , http://dx.doi.org/10.21474/IJAR01/4123
14. M. Syamir Alias and M. Syukri Hanapi , 2016. The epistemological aspect of Ibn Haytham's scientific thought, Sains Humanika 8 (2-3) , 87-92 .

15. A.Treiger , 2011. Al Ghazali's Classification of the Sciences and description of the Higest Theroretical Science , Divan Disiplilerarasi Calismar Dergisi (2011/1) 1-15 .
16. Adiasetia Mohd Dom , 2011. Reviving Kalam Jadid in the Modern Age : The perpetual relevance of al Ghazali and Fakhr al-Din al-Razi , TAFHIM (IKIM Journal of Islam and the Contemporary World 4 , 107-157 .

New data on the linguistic diversity of authorship in philosophy journals

Chun-Ping Yen* and Tzu-Wei Hung**

Abstract: This paper investigates the representation of authors with different linguistic backgrounds in academic publishing. We first review some common rebuttals of concerns about linguistic injustice. We then analyze 1,039 authors of philosophy journals, primarily selected from the 2015 Leiter Report. While our data show that Anglophones dominate the output of philosophy papers, this unequal distribution cannot be solely attributed to language capacities. We also discover that ethics journals have more Anglophone authors than logic journals and that most authors (73.40%) are affiliated with English-speaking universities, suggesting other factors (e.g. philosophical areas and academic resources) may also play significant roles. Moreover, some interesting results are revealed when we combine the factor of sex with place of affiliation and linguistic background. It indicates that while certain linguistic injustice is inevitable in academic publishing, it may be more complex than thought. We next introduce Broadbent's (2009a, 2009b, 2012, 2013, 2014) contrastive account of causation to give a causal explanation of our findings. Broadbent's account not only well characterizes the multifaceted causality in academic publishing but also provides a methodological guideline for further investigation.

Keywords: *Linguistic injustice; Lingua franca; Philosophy journals; Linguistic privilege, Causality*

1 Introduction

Inequality in academic publishing is an emerging issue with researchers having recently identified various disproportionate distributions of authorship. Wellmon and Piper (2017) disclose that 86% articles of a number of top humanities journals from 1969 through to 2015 were published by authors working at, or graduated from, 20 elite universities. Wihelm, Conklin, and Hassoun (2017) expose the underrepresentation of female scholars in publishing, in which 85% of all articles in top 25 philosophy journals are produced by male authors. Politzer-Ahles, Holliday, Girolamo, Spychalska, and

* Graduate Institute of Philosophy, National Tsing Hua University; chunping.yen@gmail.com.
** Institute of European and American Studies, Academia Sinica; htw@gate.sinica.edu.tw.

Berkson (2016) also argue how *lingua franca* may affect the justice of academic publishing, and call for an empirical survey of the issue.

In this paper, we investigate the issue of linguistic justice in academic publishing. Linguistic injustice, generally, refers to the disadvantage imposed on non-native speakers of a *lingua franca*, which usually denotes English language. Linguistic injustice in academic publishing has been recently and intensively discussed by researchers in the sciences (Bortolus 2012; Clavero 2011; Corcoran 2015; Di Bitetti & Ferreras 2017; Ferguson 2007; Guariguata et al. 2011; Primack 2009; Mori et al. 2015) and the humanities (Langum & Sullivan 2017; Hyland 2016a; Radder 2015; Wolters 2015; De Shutter & Robichaud 2015). However, whether and to what extent linguistic justice does exist, as well as whether the injustice reported by academics can be attributed to language dominance, remains a debate.

For example, Philippe van Parijs (2011, 2015) argues that while the spreading of the English language is somewhat desirable, it unavoidably results in injustice in three senses. First, although English as a global *lingua franca* provides a public good to all users through communication, it costs non-Anglophones more to participate in the communication, which leads to *unfair cooperation*. Besides, a *lingua franca* creates *unequal opportunities* among English speakers (native vs. non-native) and between English speakers and those who cannot master it. Neither access to opportunities to learn English, nor to related employment, are alike. Third, as languages are identity markers, the domination of English causes unequal recognition and a *disparity of esteem* in non-Anglophones. To reduce injustice, van Parijs encourages all language communities to "impose their language in public education and public communication within some territorial boundaries" (2015, p.224).

Conversely, Stephen May (2015) and Sue Wright (2015) both question van Parijs' idea of 'language' because it shares an outdated view with the historical builders of nation-state; i.e., a language (e.g., German, French and Spanish) is an indicator representing the national identity, singular, and sharply differentiated from other languages. But in fact, sociolinguistic literature has shown not only that human communication is heterogeneous and variable, but also that migrations and diasporas remain tied up to their roots even in this time of globalization. Thus van Parijs fails to recognize the diversity of English languages and overlooks factors that may lead to injustice. Their criticism echoes De Schutter's (2014, 2017a, 2017b) view that injustice could occur within a language and between languages.

Likewise, Gibson Ferguson (2007) and Ken Hyland (2016a) review the alleged evidence for linguistic injustice in academic publishing and argue that there is little support for the claim of systematic and widespread bias against non-native researchers. Framing the issue as an Anglophone vs. non-Anglophone dualism is even criticized as

oversimplified and counterproductive (Guariguata et al. 2011; Hyland 2016a). While admitting that language can be a barrier for scholars from non-Anglophone backgrounds, they deny that such a linguistic factor is responsible for publication success. On the contrary, non-language factors, such as degree of training and experience in academic writing, geographical location, and financial support, are more powerful determinants of publication success than language (Hyland 2016a, p.64; Guariguata et al. 2011, p.59; Ferguson 2007, pp.29-31). However, Stephen Politzer-Ahles and colleagues (Politzer-Ahles et al. 2016) maintain that this line of thinking underestimates the effect of linguistic privilege—certain social benefits that a native speaker of a language has not earned but enjoys. They contend that, as academic publishing may require less effort for Anglophones and editors prefer native-like English, injustice occurs.

In this paper, we aim to reveal the unequal distribution of philosophy authors in terms of their mother tongues, as well as to discuss in what sense the unequal distribution can be causally attributed to linguistic privilege in academic publishing. We first present a critical review of the dismissal of concerns about linguistic injustice (Section 2) with a novel survey on the representation of authors with diverse backgrounds (Section 3). While our study supports the concerns about linguistic injustice, it also indicates that the interplay of different factors may be more complex than both the believers and the deniers of linguistic injustice usually think. A multifactorial model of causal attribution is called for. To this end, in Section 4 we introduce Broadbent's (2009a, 2009b, 2012, 2013, 2014) contrastive account of causation. We also briefly discuss how such a multifaceted causal model can provides a methodological guideline for further investigation.

2 Making sense of linguistic privilege

The often cited evidence for the dismissal of concerns about linguistic injustice suggests that non-native English researchers are writing successfully for publication, that their native English counterparts also feel challenged and intimidated by writing for publication, and that the alleged inequalities in research output between native English and non-native English researchers closely mirror global socio-economic inequalities between developed and developing countries (Ferguson 2007; Hyland 2016a). Even if English remains a barrier to publication for some researchers, it is not proven that it is the most powerful determinant of publication success. According to those deniers, academic publishing is an enterprise more likely about one's experience and

connectivity, as well as the resources and supports one can access, than about one's linguistic background. It is not determined by "whether papers are written in a second language or not but [by] whether they have something to say and are written in a way the target community expects things to be said" (Hyland 2016b, p.10; see also Ferguson 2007; Guariguata et al. 2011). Thus, they argue that there is a need for more evidence for systematic disadvantages or prejudice against non-native English researchers in academic publishing before we entertain claims that native English researchers are linguistically privileged in placing their research results in high prestige international journals. Failing to do so, says Hyland, is holding the "orthodox, and unexamined, belief that life is much easier for such individuals when they seek to publish academic papers simply because they are *Native English Speakers*" (Hyland 2016b, p.9, italics in original).

We agree that a successful publishing career requires academic writing literacy, which is surrounded by challenges in addition to simple linguistic proficiency. Nonetheless, the argument that issues of linguistic disadvantage are largely irrelevant to the issue of academic publishing is not thus convincing. Acknowledging that there are more factors, other than language, involved in academic writing, by itself does not dismiss the issue of linguistic privilege. It is important to note that most people hold a combination of membership statuses. One can be a native English speaker and a novice academic writer at the same time, and the evidence that native speakers struggle as novice writers, for example, is not evidence that they would not struggle yet more were they non-native speakers (Politzer-Ahles et al. 2016; Subtirelu 2016). Moreover, it is inevitable that different factors intersect and interact with each other, resulting in a more complex influence on publication success. Whether one's research is well supported, and whether the research output is presented in near-native English, for example, can both play significant roles in the probability of a manuscript being accepted in academic journals.

It seems puzzling if one rejects the relevance of linguistic disadvantage in academic writing but explicitly acknowledges that "the need for a certain proficiency in a foreign language inevitably creates an added burden for authors" (Hyland 2016a, p.61; see also Ferguson 2007; Guariguata et al. 2011). Since such a burden presents a challenge to non-native English speakers relative to native English speakers in academic publishing, and as this is due to nothing they have themselves done, there seems no grounds for dismissing the role of linguistic privilege in academic publishing. Moreover, a certain language proficiency seems too important to not be taken into consideration for publication success when it is crucial to deliver the manuscripts "in a way the target community expects things to be said" (Hyland 2016b, p.10).

Those who deny linguistic injustice do not seem to think that one can simultaneously hold both that the native English speakers have a linguistic advantage over the non-native speakers in academic writing, and that academic writing presents considerable challenges to all scholars no matter what their first language is. It is obviously seen in, for example, Hyland's response to Politzer-Ahles et al. (2016) where he states that once we think that native English speakers enjoy a linguistic advantage over the non-native ones in academic writing, we presumably think that the native English speakers' "ability to write academic prose is not earned but conferred on them by virtue of their membership of the privileged native English group" (Hyland 2016b, p.9). As a consequence, they worry that the current discussion of linguistic disadvantage in academic writing is "damaging" and "discouraging" to the non-native English speaking writers as it "tells them to look for prejudice rather than revision" (Hyland 2016a, p.66; see also Guariguata et al. 2011; Ferguson 2007).

We think that the key to dissolving the cause of such worry is to acknowledge that there are factors other than language proficiency to effect on scholars' ability to produce quality manuscripts. Since there are other factors involved in academic writing, there is no grounds to think that one's success or failure in a publishing career is due to one's mother tongue alone. It follows that non-native English researchers may have their manuscripts accepted and native English researchers may have their manuscripts rejected in English journals. Moreover, as Politzer-Ahles and colleagues (2016, p.4) point out, it does not entail that whenever a privileged individual accomplishes in academic publishing, it is due to the fact that she is a member of the privileged group. Nor does it ascribe to us the belief that there is no difficulty of getting published in academia for the native English scholars (Politzer-Ahles et al. 2016).

In his response to Politzer-Ahles et al (2016), Hyland protests that the idea of linguistic injustice in this sense, if not an illusion, is "uncertain enough not to warrant categorical assertions" (Hyland 2016b, p.10). While Hyland may have reason to be disappointed about the uncertainty involved, such uncertainty is not uncommon in our everyday causal claims. Consider the case of smoking. While the significant positive association between cigarette smoking and lung cancer was observed in epidemiological studies since the mid 1950s, solid inferences were difficult to be drawn from such statistical associations. There are people who develop lung cancer even if they have never smoked, and there are people who smoke without developing lung cancer. The association between smoking and lung cancer is neither necessary nor sufficient. Moreover, smoking seems associated with other diseases as well.

[1]

In the report on Smoking and Health published in 1964, the Advisory Committee of the U.S. Surgeon General to the Public Health Service stated, in regard to that the causal status of associations, that "[the] causal significance of an association is a matter of judgment which goes beyond any statement of statistical probability." Nonetheless, the report also granted that "[w]hen coupled with the other data, results from the epidemiologic studies can provide the basis upon which judgments of causality may be made." Five criteria were suggested as informal guidelines to help assess the causal significance of an association in the very report.[2] A major goal of epidemiology is to prevent and control disease so these criteria are inevitably influenced greatly by the practical requirements of disease prevention.[3] As far as prevention is concerned, knowledge of causal mechanisms in their entirety is unnecessary to justify a preventive intervention (MacMahon & Pugh 1970, pp.23-24). Epidemiologic studies have shown that smoking is linked to more than 80% of lung cancers and that stopping smoking can greatly reduce the risk of lung cancer (CRUK). Applying the criteria of causal judgment to the case of smoking and lung cancer, such evidence is considered sufficient to justify a preventive intervention against smoking in epidemiology, given that there is no strong substantial evidence for a confounding factor that could explain the strong association between smoking and lung cancer.

Similarly, if there is strong substantial evidence for the association between linguistic injustice and academic publishing at one hand, and there is no better alternative explanation on the other, then we will have justification to intervene against linguistic injustice in academic publishing. For linguistic privilege to be a factor in publishing success, it does not require attributing one's success or failure in publishing to one's membership in the privileged or underprivileged group, as Hyland seems to think. The full knowledge of the underlying causal mechanisms is unnecessary for our intervention in the system to compensate for linguistic injustice (if any).

3 Data

[1] Likewise situations are rather common in the area of chronic diseases, primarily cancer and cardiovascular disease. There is usually a large number of factors involved in the development of these diseases, but presence of those factors might not always result in the expected disease.

[2] They are consistency, strength, specificity, temporal relation, and coherence of the association.

[3] For examples of the causal criteria in Epidemiology, see Hill (1965) and Susser (1991).

We have explained above why some common rebuttals are conceptually untenable. This section then presents new data on the disproportionate representation of authors with different language backgrounds by surveying 22 philosophy journals in the analytic traditions.

3.1 Methods

We reviewed 1,039 authors who published philosophy papers in the analytic traditions between 2013 and 2017, in which 703 authors belong to journals in English and 336 from those in Mandarin Chinese. The English journals fall into two categories: general philosophy and specialist journals. The former includes 10 journals selected from the Leiter Report (2015) to cover different countries in Anglosphere (see Table 1).[4] These countries are identified according to the SCImago Journal & Country Rank (SJR). We used journals of a non-Anglosphere country as control group to see whether they have fewer Anglophone authors. In each journal, about 50 authors are recorded. The latter category contains eight specialist journals: half of them belong to the area of logic and mathematic philosophy, and the other half belong to ethics and legal philosophy. The two areas are carefully chosen to decide whether formal languages may reduce the advantage of English, in which 25 authors are recorded in each specialist journal. In both categories, we only recorded original articles and discussion articles (i.e. pieces that are neither original articles nor book reviews). When an article has more than one author, each is recorded equally. For authors who appeared more than once in the same issue (e.g., reply to comments), the author is only counted once.

By Anglophone, we mean people who speak English as their native or first language. Although English varies from place to place (Hyland 2016a), the variants are not regarded as different languages in this study. Moreover, not all countries independent from the British Empire which retain official status of English are considered a part of the Anglosphere unless they have a high education penetration rate and English is the language of instruction in all public schools. In this sense, Singapore, but not India, is regarded as a part of the Anglosphere.

Whether an individual author is Anglophone is primarily determined by the open data on the Internet (e.g., CV, university website, or social media, etc.). If the direct information is unavailable, other clues are referred in a synthesized manner.[5] For example, an author born in Taipei who received a BA in Taiwan but works in the US is

[4] In this study, Anglosphere refers to English-speaking nations including the US, the UK, Australia, Canada, Ireland, Singapore, and New Zealand. All have similar cultural roots back to the British Empire.

[5] These clues include names, other languages of research, the match between the place of birth and that of receiving first degree, etc.

judged to be a native Mandarin speaker. Other traces (e.g. the ascent of an author's talk on TED or YouTube) are also considered. Although the above methods are useful in distinguishing Anglophones from those who are not, they are not always helpful in identifying an author's mother tongue. Nor can they reveal whether an author is bilingual or multilingual. Besides linguistic background, we also recorded authors' sex and affiliation for cross comparison. We divided *sex* into male and female and *affiliation* into universities in Anglosphere nations and those which are not. These variables are used to recognize possible confounders of the target phenomena.

Anglophone Journals (703 authors)		
General philosophy (501)	*Philosophical Review* (Duke, US)	2013(3)-2017(2)
	Journal of Philosophy (JP Inc. US)	2015(5)-2016(11)
	Nous (Wiley, UK)	2016(2)-2017(2)
	Mind (OUP, UK)	2016(3)-2017(2)
	Analysis (OUP, UK)	2016(1)-2016(4)
	Synthese (Springer, Netherlands)	2017(4)-2017(5)
	Erkenntnis (Springer, Netherlands)	2016(6)-2017(3)
	Australasian Journal of Philosophy (Taylor & Francis, UK)	2016(3)-2017(2)
	Canadian Journal of Philosophy (Taylor & Francis, UK)	2016(3)-2017(4)
	European Journal of Philosophy (Wiley, UK)	2015(4)-2017(1)
Logic (101)	*Journal of Philosophical Logic* (Springer, Netherlands)	2016(6)-2017(3)
	Journal of Formal Logic (Notre Dame-Duke, US)	2016(4)-2017(2)
	Philosophia Mathematica (OUP, UK)	2016(2)-2017(1)
	Mathematical Logic Quarterly (Springer, Netherlands)	2016(6)-2017(2)
Ethics (102)	*Philosophy & Public Affairs* (Wiley, UK)	2014(4)-2017(1)
	Ethics (Chicago, US)	2016(3)-2017(2)
	Law and Philosophy (Springer, Netherlands)	2016(5)-2017(3)
	Neuroethics (Springer, Netherlands)	2016(3)-2017(1)
Sinophone Journals (336)		
Logic (100)	《逻辑学研究》(*Journal of Studies in Logic*)	2014(4)-2017(1)
Ethics (236)	《伦理学研究》(*Journal of Studies in Ethics*)	2016(1)-2017(1)

Table 1 List of journals and their issues.

3.2 Results

Our survey reveals several noteworthy results. First, 68.8% of all authors are Anglophone, followed by German (6.54%) and Dutch speakers (3.55%). Where language family is concerned, more than 92.49% authors are the speakers of Indo-European languages (in which 93.61% are Germanic languages) (Section 3.2.1). Second, when a formal language is used in addition to English in a journal, the diversity of authors' linguistic background dramatically increases. Logic journals have more non-Anglophone authors than ethics journals. This effect is also reported in logic journals published in Mandarin Chinese (Section 3.2.2). Third, most authors, regardless of their native language, are working in English-speaking universities or countries. This fact suggests that academic advantage (e.g., resources and connections) may outweigh linguistic advantage in certain circumstances. We also discovered that journals registered in English-speaking countries have more Anglophone authors than their non-Anglosphere counterparts (Section 3.2.3). Finally, when the factor of sex is added for cross-comparing, non-Anglophone female authors affiliated with non-English-speaking universities are outnumbered by Anglophone male authors affiliated with English-speaking ones (Section 3.2.4).

3.2.1 Linguistic backgrounds

We sorted 703 authors into different groups by their mother tongue and arranged the groups by the number of speakers in descending order. We found that English, as expected, is the group with the most journal authors. 484 out of 703 authors are native English speakers.[6] German and Dutch take 2nd and 3rd places in the ranking with 46 and 25 authors respectively, but the gap between the first and second most widely used languages is huge. Next to Dutch is Italian (24), French (19), Swedish (19), Hebrew (12), and Norwegian (11). Hungarian, Chinese, and Korean share the 9th place with five authors each. In other words, Anglophones enjoy the absolute majority among all authors, with even the sum of the other 10 languages falls short of the number of Anglophones by 313 authors. The ranking is expressed in percentages as Fig. 1.

When the focus is not individual language but language family, Indo-European speakers reach 92.49% (despite that only 46.32% of all humans are the of Indo-European speakers). When broken down further, all the top three languages belong to North-Germanic branches. Besides, 83.58% of all authors are the speakers of Germanic

[6] As estimated in 2017, only about 5% of the global population (estimated 7.5 billions) are native English speakers (estimated 372 millions).

languages (including English), followed by Romanian languages (6.95%), Balto-Slavic languages (1.12%), and other Indo-European languages (0.84%). Sino-Tibetan and Austronesian languages only have 0.71% and 0.14% of all authors respectively. In other words, the longer distance between a language and English in the family tree, the fewer authors of that language are reported. This result is compatible with Poilzer et al.'s (2016) study that Anglophone authors have a certain linguistic privilege in academic publishing.

The above data show that, regarding to speaker population, the distribution of authors' linguistic background is highly disproportionate. However, the data by no means indicate that this unequal distribution can be solely attributed to the linguistic advantages of Anglophones. Other factors (e.g., whether authors are from developed countries) may also take part. To identify these possible factors, further analysis is offered in the next sections.

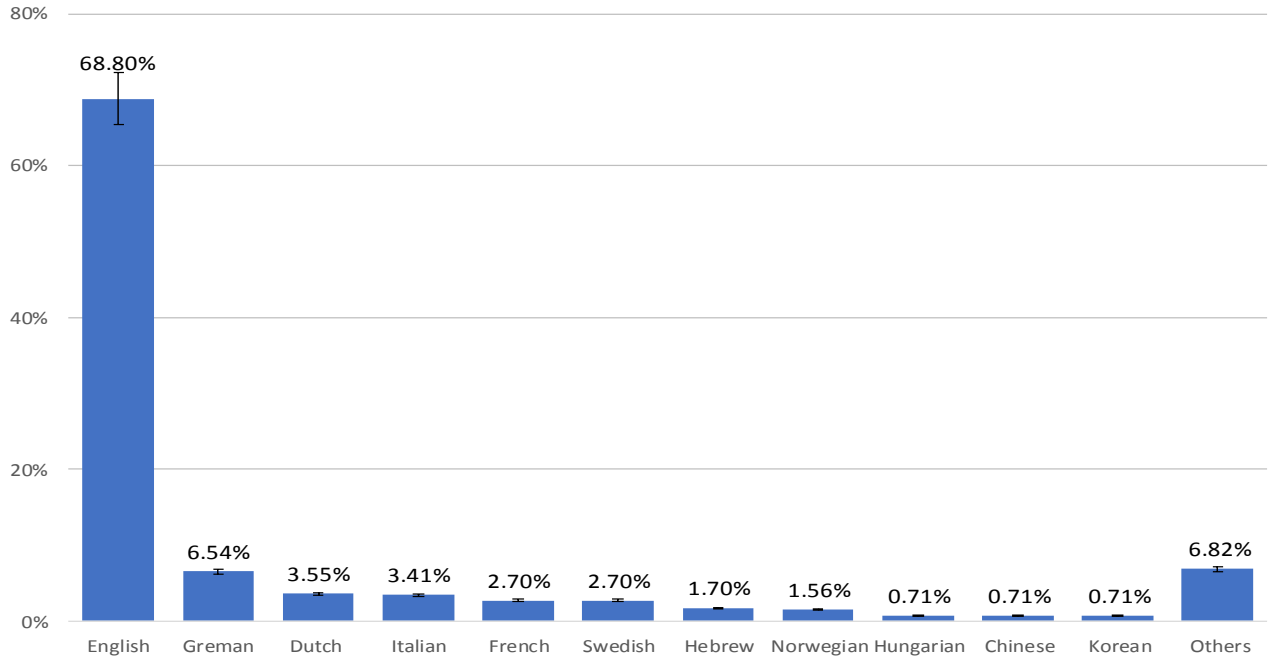


Fig. 1 Top 11 languages of 703 authors

Indo-European languages	English 68.8% Germanic (excl. English) 14.78% Romance 6.95% Balto-Slavic 1.12% Other Indo-European languages 0.84%	92.49%

Afro-asiatic languages	Modern Hebrew 1.70%	1.84%
	Arabic 0.14%	
Uralic languages	Hungarian 0.71%	1.13%
	Finnish 0.28%	
	Estonian 0.14%	
Sino-Tibetan languages	Mandarin Chinese 0.71%	0.71%
Korean	Korean 0.71%	0.71%
Turkic	Turkish 0.28%	0.28%
Austronesian	Malay 0.14%	0.14%
Other language families		2.70%

Table 2 Linguistic backgrounds of authors from 18 philosophy journals published in English

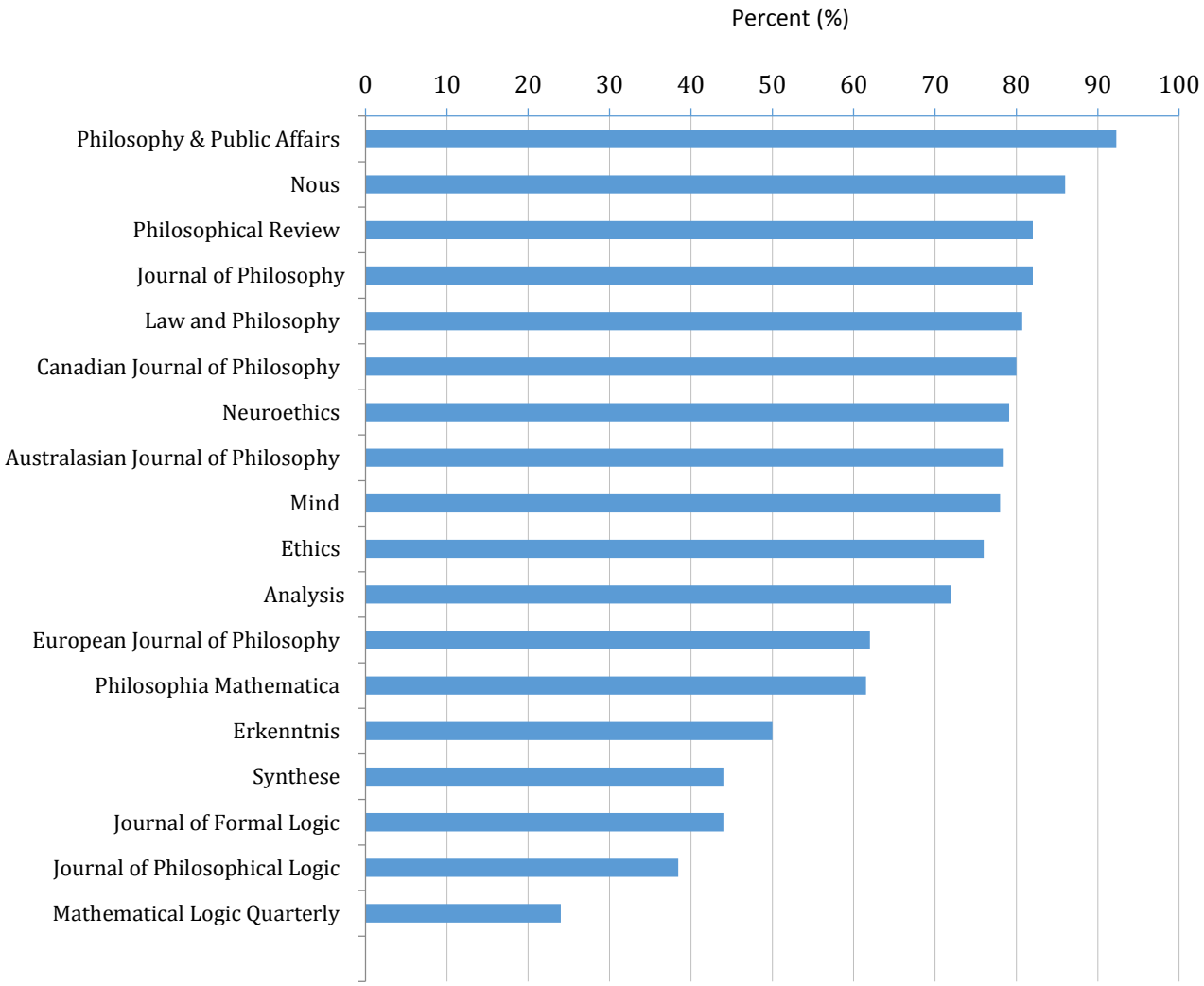


Fig. 2 Percent Anglophone authors of all articles by journals.

3.2.2 Formal languages vs. lingua franca

Before the survey, we suspected that the number of Anglophone authors might be reduced in the fields of philosophy where formal languages are also widely used. Despite variation across journals, the data confirmed our suspicion that journals in logic and mathematic philosophy (logic journals for short) have significantly fewer Anglophone authors than those of ethics and law philosophy (ethics journals for short). 82.02% of all authors of ethics journal are native English speakers, and the percentage drops to 41.90% in logic journals. When breaking down by individual journals, we have similar results. Every ethics journal has a higher percentage of Anglophone authors than every logic journal (Fig.3).

Similar difference can be found between two top journals published in Chinese Mandarin; *the Journal of Studies in Logic* (逻辑学研究) and *the Journal of Studies in Ethics* (伦理学研究). We found that although Chinese is the lingua franca of the journals, it is affected by formal languages as well. Among all authors of the ethics journal published in Chinese, 99.15% (234 out of 236 authors) are native Mandarin speakers. However, the rate drops to 88% (88 out of 100 authors) in the logic journal (Fig.4). All these results suggest that the dependency of formal language (logic or mathematics) decreases the influence of a lingua franca in academic publishing.

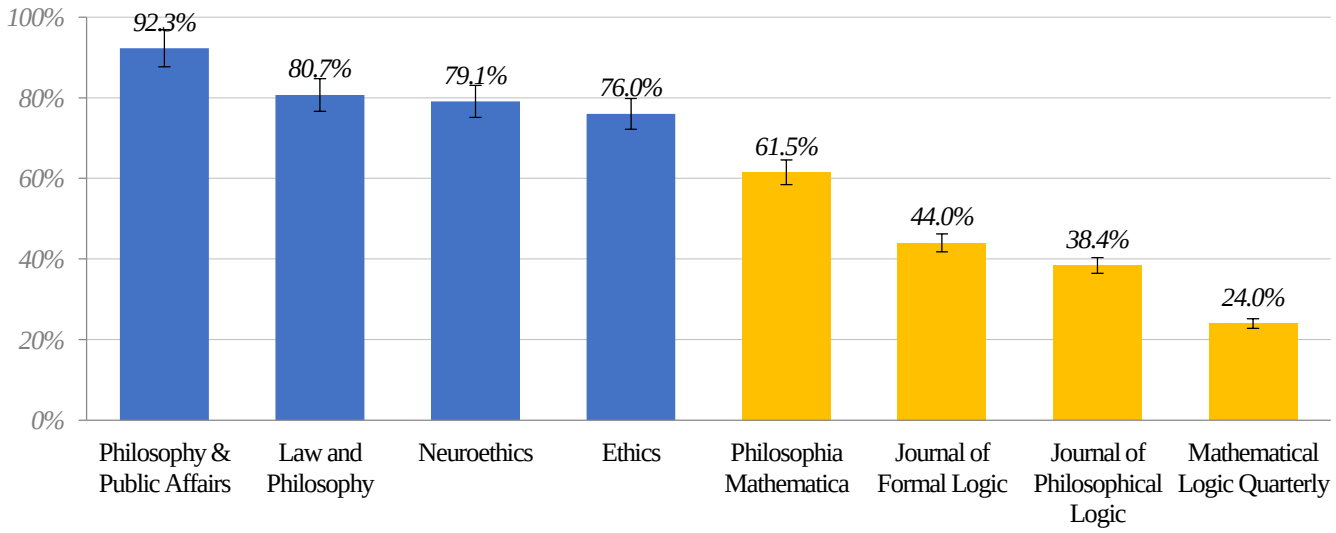


Fig. 3 Percent Anglophone authors of logic (yellow) and ethics journals (blue) published in English

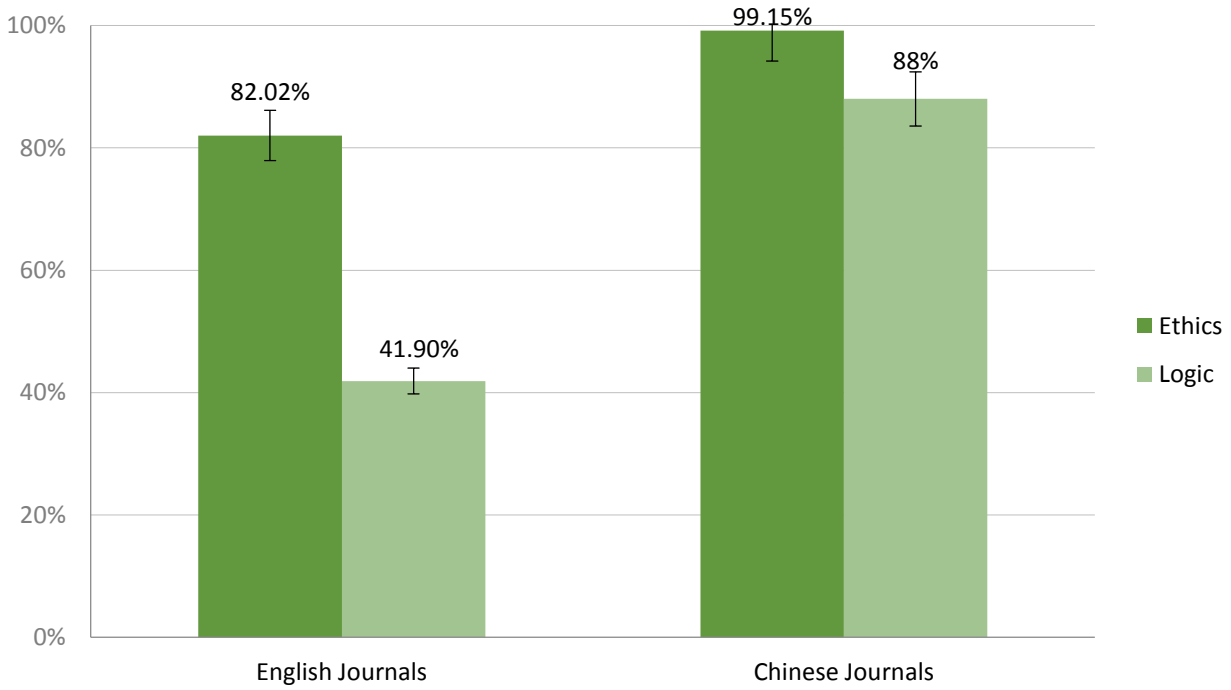


Fig. 4 Percent Anglophone authors (left) and Sinophone authors in ethics (dark green) and logic journals (light green).

3.2.3 Hereditary advantages of Anglosphere

Those who question linguistic injustice often argue that the low rate of representation of non-Anglophone scholars in academic journals is not due to linguistic disadvantage, but because they often suffer from various forms of isolation when returning to own countries. [7] Our data show that this account rightly identifies one factor of

[7] Examples include insufficient funding and peers to discuss with, unable to attend conferences, disconnect to the previous academic networks, or unfamiliar with journals' rules of game. The scholars may be isolated from the mainstream of professional community and resources, affecting the quality and

disproportionate representation of Anglophone authors, although there are other factors affecting the phenomenon. First, of all 703 authors, 73% are affiliated with at least one university/institute in Anglosphere nations. Among them, 85.07% works in the US and UK while 14.92% in other English-speaking states. These countries not only share similar analytic traditions but also have more research investments than their developing competitors. We also found that almost 1/3 of non-Anglophone authors work in the Anglosphere. Historically, Anglophone nations have deep connections with philosophy in the analytic traditions. All these nations are also economically developed and have relatively stable funding sources to facilitate high-quality studies. These high-quality studies usually attract more research investment, resulting in the circularity of academic lead of the English-speaking world. When this economic-academic advantage combines with linguistic advantage, the gap between Anglophone and the other authors is inevitably widening. Our data show that the US and UK universities remain producing the most papers, accounting for 62.4% of all published papers. Our survey also conforms to Wellmon and Piper's (2017) data that 86% articles of humanities journals are published by authors working at or graduated from top 20 elite universities. All of these universities are either US- or UK-based; about 51% of all articles in humanities journals are produced by 10 universities, including Yale, Harvard, Berkeley, Columbia, Princeton, Stanford, Oxford, Cornell, Chicago, and John Hopkins.

Conversely, although China has huge investment on academic research and is now the second largest country (just behind the US) in manuscript submissions (SCImago 2014), China only has 6% of the total number of published papers in all academic disciplines, loses to the US (29%) and Japan (8%), and shares the 3rd place with the UK and Germany (World Bank, 2012). Besides, most of China and Japan's published papers are scientific. Our data show that only 0.71% and 0.14% of all authors of philosophy journals published in English are native Chinese and Japanese speakers respectively. In other words, the Anglophone nations continue to dominate the world output of philosophy papers.

Moreover, we also found the 'Home advantage' effect of Anglosphere journals (Fig. 6). That is, philosophy journals registered in the Anglosphere countries (i.e., the US and the UK.) normally have higher representation of native English-speaking scholars than those registered in other regions (e.g., the Netherlands). In general philosophy, 71.4% of all authors in journals registered in Anglosphere are Anglophones, but only 47% are so in journals registered in other regions. Similar result can be found in ethics and logic journal too. While there are 21.94% more logic journals in the Anglosphere than in the non-Anglosphere, there are only 4% more ethics journals in the Anglosphere than in the non-Anglosphere. We suspect that this effect might be a result of the possibility that

quantity of their research (Hyland 2016a; Ferguson 2007).

editors working in English-speaking countries are more sensitive to non-native English. Although the actual cause remains unclear, the difference is significant. Our survey conforms to previous studies which found that journals favor researchers located in the same country as the journal (Daniel 1993; Ernst & Kienbacher 1991; Link 1998). Interestingly, such preference is often considered relating to factors other than authors' institutional affiliations, including the differences in scientific rigor, writing proficiency, or writing style between various institutions across countries as well (Lee et al. 2013; Resnik & Elmore 2016). It suggests that future studies should not ignore the role of language factors.

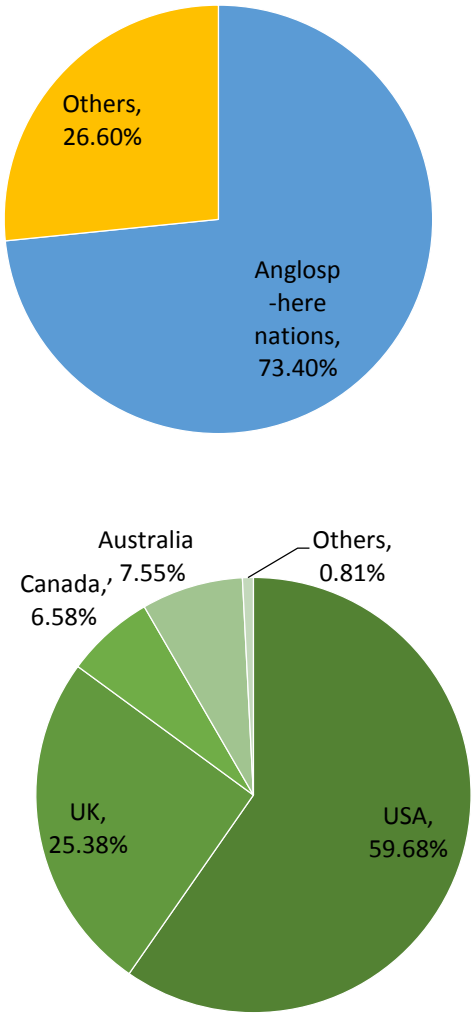


Fig. 5 The left pie charts summarizes the distribution of authors' regions of affiliation, while the right pie chart summarizes the distribution of nations within Anglosphere.

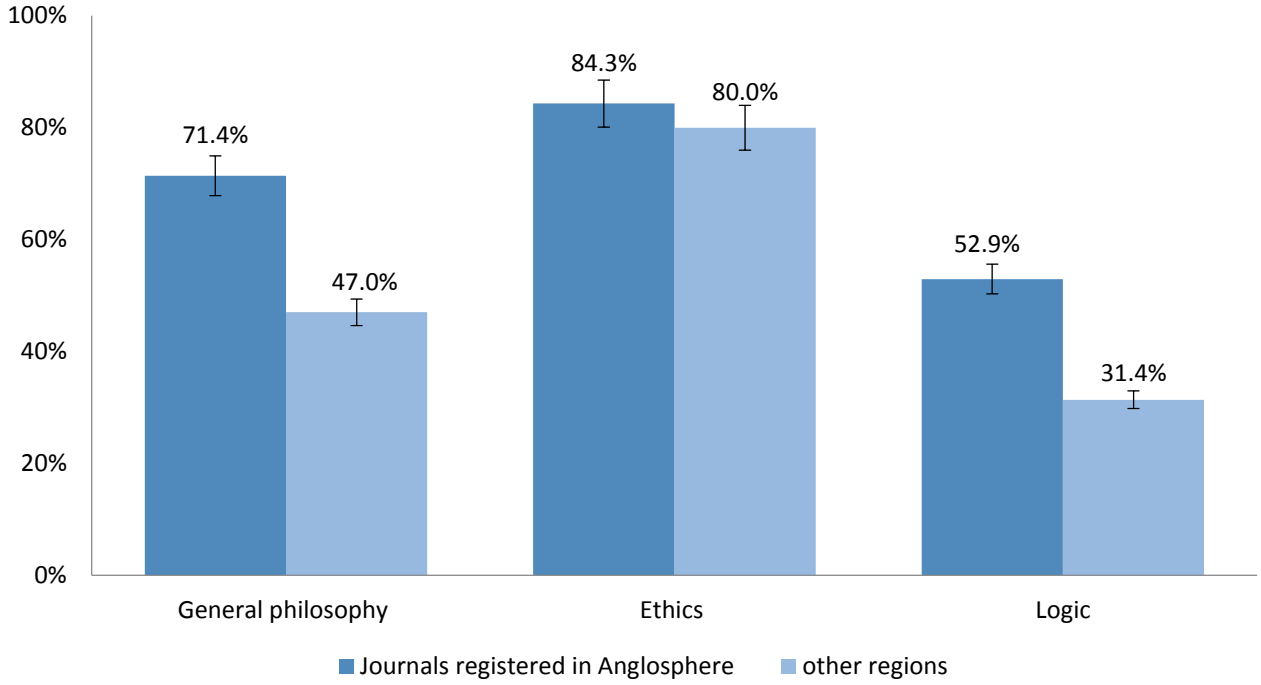


Fig. 6 Percent Anglophone authors of general philosophy journals (left), ethics journals (centre), and logic journal (right).

3.2.4 Cross comparison of sex, affiliation, and mother tongue

When taking author's sex into account, we found that 85.4% of all authors are male. The ratio between female and male philosophers is 1:5.8. This result is close to Wilhelm et al.'s (2017) survey on the representation of women in philosophy, in which men (83.7%) still dominated philosophy journals in 2015.[8]

Interesting result occurs when we combine the factor of sex with affiliation place and linguistic backgrounds (see Fig. 7). We found that most authors are from the dominant group of Anglophone males affiliated with Anglosphere universities (57.6%), followed by non-Anglophone males affiliated with non-Anglosphere universities (17.6%), Anglophone females affiliated with Anglosphere universities (8.67%), and non-Anglophone males affiliated with Anglosphere universities (8.1%). Moreover, authors who are Anglophone, female, and from Anglosphere universities (8.67%) are fewer than those who are non-Anglophone, male and from non-Anglosphere universities (17.6%), indicating that sex is a strong factor that may defeat the disadvantage of language and affiliation. However, the authors of Anglophone, female, and

[8] Beyond that, current evidence for gender bias in journal peer review is weak (Lee et al. 2013; Resnik & Elmore 2016).

Anglosphere universities (8.67%) also outnumber those of non-Anglophone, male and Anglosphere universities (8.1%), suggesting that linguistic advantage is more crucial than sex factor in certain circumstances. Finally, there are more authors who are non-Anglophone, female, and from non-Anglosphere universities (2.7%) than those who are Anglophone, male, and from non-Anglosphere universities (2.13%), showing that academic resources and connections may outweigh other factors as well. As Anglophone authors working at oversea institutes may suffer from insufficient academic resources and connections (or even greater isolation) than local researchers. In consequence, while language may affect the authors' representation, its effect varies when coupled with other factors in different situations.

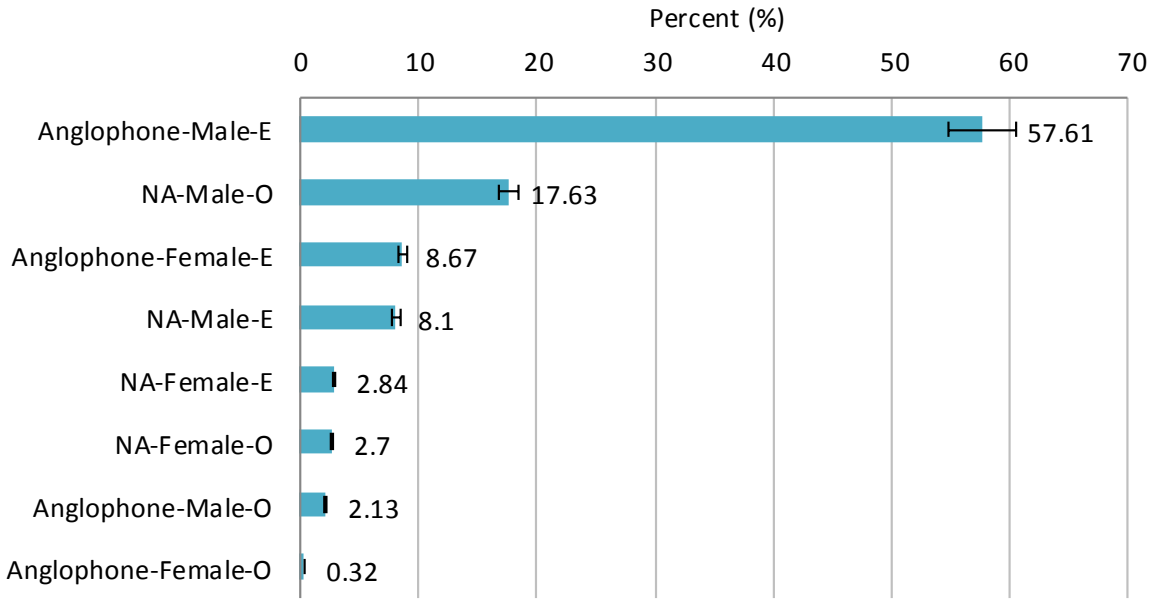


Fig. 7 Percent authors of different groups divided in terms of linguistic background, sex, and the place of affiliation. NA=non-Anglophones, E=working at universities or institutes in English-speaking countries, and O=working in other countries

4 Discussion

The above data reveal that the distribution of authors with different linguistic backgrounds is disproportionate, but skeptics may argue that, while the data show that linguistic factor and unequal distribution are *positively correlated*, it does not imply that they are *causally related*. Thus, in what sense can one justifiably claim that the underrepresentation of certain authors does causally result from linguistic disadvantage? As discussed in Section 2, different factors can intersect and interact with each other to result in a more complex influence on publication success. In addition to language proficiency, non-language factors, such as one's experience, connectivity, and the resources and supports one can access, are also likely factors behind publication success. Our data confirm this complexity. Publication success is due to the interaction of these various factors, not any one of them alone. Studies of journal peer review also indicates the interaction between the language factors and the non-language ones, say, affiliation bias, for example (Lee et al. 2013; Resnik & Elmore 2016). We need a causal framework to cope with the complex causal interaction involved in academic publishing. We first give a causal explanation of our findings in Section 3 (Section 4.1). We then respond to the concern about possible implicit bias in academic publishing (Section 4.2). In closing, we consider the question whether English is the specialist language of analytic philosophy (Section 4.3).

4.1 Causality

We employ Alex Broadbent's (2009a, 2009b, 2012, 2013, 2014) contrastive approach of causation to interpret our data. There are two things about this approach which makes it preferable for the present context. First, it does not appeal to a complete theory of the nature of causation which we do not have to give an account of how we can use measures of association to say something about causal facts.[9] Second, it is a multifactorial model which offers a general explanation for the difference between cases with and without the effect brought about by the causes. With these two features, it allows us to give a valid causal explanation of our findings in Section 3.

The key notion of Broadbent's account is *contrastive explanation* which concerns explaining differences. It is commonly considered cause as something that makes

[9] Take cigarette smoking and lung cancer for example. On the one hand, there is a significant positive association between the two. On the other hand, such measure of strength of association is causally neutral. When it is used to make about some causal claim, we need to give an account of how a causal interpretation of the measure of association works. See Broadbent (2013, pp.34-50) for his diagnosis of how the rival approaches, the probabilistic approach and the counterfactual approach, fail the task.

difference. But which kind of difference-making are we thinking about? Broadbent argues that the right kind of difference-making for causation is the difference between fact and foil:[10]

A cause makes a difference in that it is a difference between the effect being as it is and the effect being different or absent. (Broadbent 2013, p.52)

This kind of "effect-led difference making" is entirely different from "cause-led difference making" invoked by the counterfactual approach of causation, according to which, "[a] cause makes a difference in that had it been different or absent, then its effect would have been different or absent" (Broadbent 2013, p.52).

As Broadbent points out, oftentimes we do not ask simply 'Why *P*?' but instead 'Why *P* rather than *Q*?' Making a difference in the cause-led sense is not enough for a good causal explanation. Instead, we must mention a causal difference between the fact, *P*, and the foil, *Q*. For example, if you ask why a philosopher arrived late to a meeting rather than on time. It will not be helpful to mention the presence of oxygen or the step the philosopher took as she came through the door as part of the answer even though they both make a difference to her late arrival in the cause-led sense. Without the presence of oxygen or the step she took as she came through the door, she would not have arrived at all. On the contrary, that her train was late can explain her late arrival because in the case where the train were not late, she would have arrived on time.

Understanding causation as effect-led difference making, Broadbent's view allows us to say that "[a] measure of strength of association is a measure of causal strength if and only if the exposure explains the measured net difference in outcome" (Broadbent 2013, p.50). He then defines measures of causal strength in terms of the contrastive conception of causation. Accordingly, a measure of causal strength is "a measure of the net difference in outcome explained by an exposure" (Broadbent 2013, p.50).

Suppose that the relative risk (*RR*, which is a measure of strength of association) of lung cancer among smokers compared to non-smokers is 20. It means that 20 times as many smokers develop lung cancer during their lives as non-smokers. A causal interpretation of this claim "says that the *RR* of 20 is explained by the difference in smoking status between the smokers and non-smokers" (Broadbent 2013, p.51). Accordingly, the *RR* of 20 provides a measure of the causal strength of cigarette smoking with respect to lung cancer. Cigarette smoking causes the *RR* of 20 for lung cancer.

[10] See also Jonathan Schaffer (2005, pp.327-329, 2010). He too also holds a contrastive view of causation but with a contrast for both cause and effect.

Note that on this account, cigarette smoking need only be a cause, rather than the only cause, of lung cancer among smokers (Broadbent 2013, p.54). Broadbent criticizes that the monocausal conception of causation is too strict and too permissive. It is too strict because there are events which do not fit the monocausal model. It is too permissive because it does not exclude events which are common to the causal history of many events. Big Bang, for example, satisfied as the cause of all events, according to the monocausal account (Broadbent 2013, pp.153-154).

As Broadband points out, everything is multifactorial in the sense that it takes the operation of multiple causal factors to give rise to it. A useful or even nontrivial model of causal structure cannot simply permit the cataloguing of multiple causal factors without offering a way of discriminating among the causes. In order to preserve the idea that effects are caused by causes in a certain way, the contrastive model consists in citing explanatory causes which are present in cases of the effect and absent otherwise. Following Peter Lipton (2004), Broadbent proposes the contrastive model to character the causal structure of disease. "[T]o call a set of symptoms a disease," says Broadbent, "is to say that we have a causal explanation of the difference between people with these symptoms, and healthy controls who lack these symptoms" (Broadbent 2014, p.253). According to this model, in order for *D* to count as a disease, it is necessary for the following three conditions to be satisfied (Broadbent 2013, p.158, 2014, p.256).

SYMPTOMS
Cases of *D* exhibit symptoms, which are absent from controls.[11]

CASES
These symptoms are caused by $C_1, \ldots C_n$ together.

CONTROLS
At least one of $C_1, \ldots C_n$ is absent from controls.

Similarly, we can construct a contrastive model for our current concern. Our data reveal a highly disproportionate distribution with respect to authors' linguistic background which we think is effected by not only the language factors but also the non-language ones. To say that a person *p* who has to make an extra effort to get her work published due to these factors is in an underprivileged position, compared with her peers, is to say that we have a causal explanation of the difference between people who have to make an extra effort to disseminate their research due to the listed factors and people who do

[11] A case of *D* exhibits symptoms that distinguish it from controls.

not. A contrastive model is to give a causal explanation as such. For a person *p* to be in an underprivileged position *U* in publication, it is necessary for the following conditions to be satisfied.

1. Cases of *U* exhibit the SYMPTOM that it takes *p* an extra effort to disseminate her research,[12] and this SYMPTOM is absent from controls.[13]
2. The SYMPTOM that it takes *p* an extra effort to disseminate her research is caused by factors including her being a non-native speaker of a *lingua franca*, a novice, in a less advantaged research environment together.[14]
3. At least one of the aforementioned causes is absent from a set of controls which are taken as the normality of the business of publication.[15]

We want to know whether the linguistic privilege constitutes a cause of the distribution of authorship in publication in the sense that writing in English puts a non-native English speaker's work in an underprivileged position. Politzer-Ahles and colleagues (2016) suggests two places to look into for a start: the effort required to disseminate research and publishing bias. In Section 2, we argue that given the acknowledgement of the proficiency in a foreign language as an extra burden for the writers, it strikes us as odd that Hyland would reject the effort required for the non-native English speakers to be competent in writing English. Our data in Section 3 show that Anglophone authors share a certain linguistic privilege in academic publishing and the very privilege together with other, also significant, factors is in a strong association with the low rate of representation of non-Anglophone scholars in philosophy journals. In this subsection, we give an account that these factors, linguistic or not, jointly constitute the cause of the distribution of authorship in publication. Now let's move on to the implicit publishing bias.

---

12 That is, it is more challenging for *p* to disseminate her work due to the difficulty getting her work published.

13 A SYMPTOM is a difference between p and a contrast class.

14 These factors are chosen based on our data in Section 3.

15 So the controls may include, say, native English speakers who are also novices in academia and do not have to make an extra effort (as *p* does) to disseminate their work, senior non-native English scholars who do not have to make an extra effort (as *p* does) to disseminate their work, and so forth. People who have to make an extra effort (as *p* does) to disseminate their work are excluded from the controls due to the first condition. People who are with every cause listed in the second condition are excluded from the controls too. On this approach, the exclusion of cases from the controls can itself be analyzed contrastively further. See Broadbent (2013, pp.158-159, 2014).

4.2 Implicit Bias in Academic Publishing

Over the last few decades, social and cognitive psychologists have established that human beings are prone to a range of such implicit associations which contribute to patterns of discriminatory behavior.[16] According to a working definition given by Jules Holroyd (2012, pp.274-275), an individual A harbors an implicit bias against some group (G) when A has automatic cognitive or affective associations between (A's concept of) G and some negative property (P) or stereotypic trait (T), which are accessible and can be operative in influencing judgment and behavior without the conscious awareness of A. Since the associations involved here are automatic, it is not something reflected by one's self-report whether one harbors certain implicit biases about certain groups.[17] As Politzer-Ahles and colleagues (2016, p.5) point out, linguistic injustice, if existing, involves with implicit association between native-like English and the quality of research which the reviewers and editors may not be able to acknowledge or even recognize. In other words, the existence of linguistic injustice is compatible with Hyland's (2016a, p.65) observation that "the interviews with editors and studies of reviewers' comments ... tend to find no evidence to support claims of prejudicial treatment or undue attention to language in editorial decisions." Even if Hyland's observation is true, it does not eliminate the possibility of linguistic injustice. On the other hand, we think that Hyland's response that "[t]here is, however, little evidence to support the idea that there is a widespread and systematic bias against writers whose first language is not English" (Hyland 2016a, p.66) also gets to the point, especially from the perspective of the contrastive account we follow. On this account, causal strength is measured in terms of the measurement of association. It is only reasonable to count language-related bias as a cause of linguistic injustice when there is strong substantial evidence for the association between the two. It is not enough to simply point to the presence of such bias, but requires its existence in a larger scale. Such evidence is yet to be collected.

4.3 Specialist Language?

[16] For a review of this literature, see Jost et al. (2009).

[17] While the discrimination led by implicit biases is often unintentional, unendorsed, and perpetrated without the agent's awareness, it does not follow that the agent is without control of it (Suhler & Churchland, 2009).

We have identified linguistic privilege as a cause of the distribution of authorship in publication. However, is there any further consideration in which the data of author's unequal representation can be exempted from injustice?

We called a *lingua franca* a *specialist language* if the language per se is part of the research object as well. For example, Egyptian language is the specialist language for historians and archaeologists who devoted themselves to the ancient civilization. English is also the specialist language of English phonetics and philology likewise. A scholar can hardly contribute to the field without being (or cooperating with) a competence speaker of that language. Thus, if a *lingua franca* is also a specialist language, then the disproportionate representation in academic publishing is not injustice but indispensable. However, is English the specialist language of analytic philosophy?

Although analytic philosophy was inspired by many native German speakers (e.g., Gottlob Frege, Kurt Gödel, Ludwig Wittgenstein, Rudolf Carnap, and Karl Popper, etc), it was primarily developed in the Anglo-America world, in contrast to 'continental' philosophy based in Europe. It is also undeniable that most philosophers of language in the 20th century used English sentences as their target of study. However, these facts do not imply that English is the specialist language because analytic philosophy is not area studies but philosophical studies. Due to its character of universality, the validity of analytic philosophy is not confined to Anglosphere but open to all humans. This explains why Grice's analysis of implicature can also be extended to Taiwanese (Li 1997) and Korean (Lee 2010). Conversely, unless in comparative studies, the phonetic rules abstract from English is unlikely apply to Japanese, and the theological concepts derived from Egyptian pictogram can hardly be used to understand religious concepts of ancient Chinese speakers. Hence, in analytic philosophy, English is not the specialist language and its status of *lingua franca* results from historical contingency.

Similar situations happen in Chinese philosophy as well. Mandarin is not the specialist language but *lingua franca* of Chinese philosophy. Most American scholars of Chinese philosophy are not native Chinese speakers, and some do not even speak Mandarin at all. However, they can nevertheless publish quality papers on good journals of Chinese philosophy in English with referring to translated texts. We investigated into *The Journal of Chinese Philosophy* (Wiley) in 2015 (Issues 1-4) and found that only 12% of all authors are native Chinese speakers. This is because the enquiry of human nature is not limited to Chinese people but to all humans, despite the answers Chinese philosophers offered may reflect different cultural presumptions. Therefore, if the *lingua franca* of a discipline is not its specialist language, then the unequal representation of authors in that discipline is unlikely to be justified.

5. Conclusions

To summarize, after a brief review of the debate (Section 1), we explained why some rebuttals of linguistic injustice is conceptually untenable (Section 2). Our empirical survey also showed that the distribution of authors is highly disproportionate (Section 3). A causal model is then employed to clarify why these data constitute the evidence for linguistic injustice in academic publishing, as well as how other factors may also involve in the injustice (Section 4).

Nonetheless, skeptics may still argue that the issue here is not about authorship *distribution* but about author's *contribution*. After all, academic research is for revealing the *truth*, not for seeking a *balance* among author's representation. A quick reply is that, while we agree that the goal of academic research is to discover the truth, what counts as truth or the criteria of truth is not always agreed, especially when researchers of different linguistic backgrounds have dissimilar intuitions and world views. Likewise, contribution is not necessarily incompatible with distribution. While author's contribution depends on the content of paper, language is a key vehicle for that content. We should be mindful of not to weigh our decisions associate with the comparison between native and non-native English submissions. It might be a great loss to a philosophical society if a paper is discarded at the very beginning merely because of linguistic expressions. Thus, the reason why we should care about linguistic injustice lies in the value of diversity. Increasing the diversity of authorship is to bring in fundamentally different thoughts and novel perspectives to the fields. It may also raise doubts on what is assumed as a matter of course by a group of language users. Therefore, the issue of linguistic diversity is not merely about ethics but also epistemology, which hence has to be faced up and dealt with.

In short, in this paper we exposed the data of unequal distribution of philosophy authors in terms of their mother tongues. We conclude that, while linguistic injustice is inevitable in philosophy journals, it may be more complex than thought. More attention on the multifaceted causality in academic publishing is needed in future studies.

Reference

1. Advisory Committee to the Surgeon General of the Public Health Service. (1964). Smoking and health: report of the Advisory Committee to the Surgeon General. Atlanta: Public Health Service, US Department of Health, Education and Welfare.

2. Bortolus, A. (2012). Running like Alice and losing good ideas: On the quasi-compulsive use of English by non-native English speaking scientists. *AMBIO: A Journal of the Human Environment*, 1-4.

3. Broadbent, A. (2009a). Causation and models of disease in epidemiology. *Studies in History and Philosophy of Biological and Biomedical Sciences, 40,* 302-311.

4. Broadbent, A. (2009b). Fact and law in the causal inquiry. *Legal Theory, 15*, 173-191.

5. Broadbent, A. (2012). Causes of causes. *Philosophical Studies, 158*(3), 457-476.

6. Broadbent, A. (2013). *Philosophy of Epidemiology*. New York: Palgrave Macmillan.

7. Braodbent, A. (2014). Disease as a theoretical concept: The case of "HPV-it is." *Studies in History and Philosophy of Biological and Biomedical Sciences, 48*, 250-257.

8. Clavero, M. (2011). Language bias in ecological journals. *Frontiers in Ecology and the Environment*, *9*(2), 93-94.

9. Corcoran, J. N. (2015). *English as the International Language of Science: A Case Study of Mexican Scientists' Writing for Publication* (Doctoral dissertation, University of Toronto).

10. Cancer Research UK (CRUK). Smoking facts and evidence. Web. Retrieved from http://www.cancerresearchuk.org/about-cancer/causes-of-cancer/smoking-and-cancer/smoking-facts-and-evidence.

11. Daniel, H.-D. (1993). *Guardians of science: Fairness and reliability of peer review*. Weinheim, Germany: Wiley-VCH.

12. De Schutter, H., & Robichaud, D. (2015). Van Parijsian linguistic justice–context, analysis, and critiques. *Critical Review of International Social and Political Philosophy*, *18*(2), 87-112.

13. De Schutter, H. (2014). Testing for linguistic injustice: Territoriality and pluralism. *Nationalities Papers*, *42*(6), 1034-1052.

14. De Schutter, H. (2017a). Global, interlinguistic and intralinguistic linguistic justice (forthcoming).

15. De Schutter, H. (2017b). Two principles of equal language recognition. *Critical Review of International Social and Political Philosophy*, *20*(1), 75-87.

16. Di Bitetti, M. S., & Ferreras, J. A. (2017). Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. *Ambio*, *46*(1), 121-127.

17. Ernst, E., & Kienbacher, T. (1991). Chauvinism. *Nature, 352*, 560.

18. Ferguson, G. (2007). The global spread of English, scientific communication and ESP: questions of equity, access and domain loss. *Ibérica, 13*, 7-38.

19. Guariguata, M. R., Sheil, D., & Murdiyarso, D. (2011). 'Linguistic injustice' is not black and white. *Trends in Ecology and Evolution, 26*(2), 58-59.

20. Hill, A. B. (1965). The environment and disease: Association or causation. *Proceedings of the Royal Society of Medicine*, *58*, 295-300.

21. Holroyd, J. (2012). Responsibility for implicit bias. *Journal of Social Philosophy, 43*(3), 274–306.

22. Hyland, K. (2016a). Academic publishing and the myth of linguistic injustice. *Journal of Second Language Writing*, *31*, 58-69.

23. Hyland, K. (2016b). Language myths and publishing mysteries: A response to Politzer-Ahles et al. *Journal of Second Language Writing*, *34*, 9-11.

24. Jost, J.T., Rudman, L.A., Blair, I.V., Carney, D. R., Dasgupta, N., Glaser, J., & Hardin, C.D. (2009). The existence of implicit bias is beyond reasonable doubt: A refutation of ideological and methodological objections and executive summary of ten studies that no manager should ignore. *Research in Organizational Behavior, 29,* 39-69.

25. Langum, V., & Sullivan, K. P. (2017). Writing academic english as a doctoral student in sweden: Narrative perspectives. *Journal of Second Language Writing*, *35*, 20-25.

26. Lee, C. (2010). Information structure in PA/SN or descriptive/metalinguistic negation: with reference to scalar implicatures. In Dingfang Shu & Ken Turner (eds.), *Contrasting Meanings in Languages of the East and West*, 33-73. New York: Peter Lang.

27. Lee, C. J., Sugimoto, C. R., Zhang, G., & Cronin, B. (2012). Bias in peer review. *Journal of the American Society for Information Science and Technology, 64*(1), 2–17.

28. Li, C. I. (1997). Logical entailment and conversational implication: A discourse-pragmatic account of Taiwanese toh (就) and ciah (才). *Journal of Taiwan Normal University: Humanities & Social Science, 42*, 55-70. doi:10.6210/JNTNULL.1997.42.05

29. Link, A.M. (1998). US and non-US submissions. *Journal of the American Medical Association, 280*(3), 246–247.

30. Lipton, P. (2004). *Inference to the Best Explanation*. 2nd ed. New York: Routledge.

31. MacMahon B, & Pugh T. F. (1970). *Epidemiology: Principles and Methods*. Boston: Little, Brown & Co.

32. Mori, A. S., Qian, S., & Tatsumi, S. (2015). Academic inequality through the lens of community ecology: a meta-analysis. *PeerJ*, *3*, e1457.

33. Politzer-Ahles, S., Holliday, J. J., Girolamo, T., Spychalska, M., & Berkson, K. H. (2016). Is linguistic injustice a myth? A response to Hyland (2016). *Journal of Second Language Writing*, *34*, 3-8.

34. Primack, R. B., Ellwood, E., Miller-Rushing, A. J., Marrs, R., & Mulligan, A. (2009). Do gender, nationality, or academic age affect review decisions? An analysis of submissions to the journal Biological Conservation. *Biological Conservation*, *142*(11), 2415-2418.

35. Radder, H. (2015). How Inclusive Is European Philosophy of Science?. *International Studies in the Philosophy of Science*, *29*(2), 149-165.

36. Resnik, D., & Elmore, S. (2016). Ensuring the quality, fairness, and integrity of journal peer review: A possible role of editors. *Science and Engineering Ethics, 22*(1):169-188.

37. Schaffer, J. (2005). Contrastive causation. *Philosophical Review, 114*, 327-358.

38. Schaffer, J. (2010). Contrastive causation in the law. *Legal Theory, 16*, 259-297.

39. Subtirelu, N. (2016). Denying language privilege in academic publishing. [Web log post]. Retrieved from https://linguisticpulse.com/2016/03/28/denying-language-privilege-in-academic-publishing/.

40. Suhler, C., & Churchland, P. (2009). Control: conscious and otherwise", Trends in cognitive sciences, 13(8), 341–347.

41. Susser, M. W. (1991). What is a cause and how do we know one? A grammar for pragmatic epidemiology. *American Journal of Epidemiology*, *133*, 635-648.

42. Van Parijs, P. (2011). *Linguistic Justice for Europe and for the World*. Oxford University Press on Demand.

43. Van Parijs, P. (2015). Lingua franca and linguistic territoriality. Why they both matter to justice and why justice matters for both. *Critical Review of International Social and Political Philosophy*, *18*(2), 224-240.

44. Weihelm, I., Conklin, S. L., & Hassoun, N. (2017). New data on the representation of women in philosophy journals: 2004-2015. *Philosophical Studies*.

45. Wellmoon, C., & Piper, A. (Forthcoming). Publication, power, and patronage: on inequality and academic publishing. *Critical Inquiry*.

46. Wolters, G. (2015). Globalized Parochialism: Consequences of English as Lingua Franca in Philosophy of Science. *International Studies in the Philosophy of Science*, *29*(2), 189-200.

47. Wright, S. (2015). What is language? A response to Philippe van Parijs. *Critical Review of International Social and Political Philosophy, 18*(2), 113-130.

SCIENCE FROM PERSPECTIVE OF AL-QUR'AN[1]

Mohd Yusof Hj Othman
Institute of Islam Hadhari, Universiti Kebangsaan Malaysia, 43600 Bangi, Malaysia
e-mail: myho@ukm.edu.my

ABSTRACT

According to Toby Huff (1995), from about eighth century till the end of the thirteenth century, the Arabic-Islamic world had the most advanced science to be found anywhere in the world. They established excellent centre of learning called House of Wisdom in Baghdad (786), the world's first university called al-Qarawiyin University in Fez, Morocco in 859, followed by University of al-Azhar in Cairo (972). They produced renowned scientist such as Jabir Ibn Hayyan (722) (the father of chemistry), al-Biruni (973) (in astronomy, mathematics, physics, geography and history), Ibn Sina (980) (in medicine), Ibn Haytham (965) (in optics) and many more. They also introduced terminologies in science which is being used in science until today such as algorithm, algebra, camera, music, chemistry, alkali, alcohol, cornea, and so on. They modified old theories and established new concepts in science such as the concept of light in optic and on how eyes see an object; heliocentric not geocentric concept of solar system as was mentioned by al-Biruni. Al-Qur'an is the holy script of Islam. Muslims in the past and present day regard al-Qur'an as the source of knowledge including the subject of science and technology. This paper discusses the concept of science from perspective of this holy script. Of course al-Qur'an is not the book of science, but al-Qur'an requests Muslims to observe and understand the entire scientific phenomenon around them.

Keywords: The Qur'an, Observation, *Tawhidic* Science, Scientific method, *Quranic* epistemology.

INTRODUCTION

The development of science and technology is in line with human history. Nobody deny that humans need science and technology to develop and prosper in this world. Without science

[1] Paper presented at the 7th Asia-Pacific Conference on Philosophy of Science, held at Department of Philosophy, National Chung Cheng University in Chia-Yi of Taiwan on December 15-16, 2017.

and technology, the world is unlikely to progress as we see today. Human beings, not other creatures, have the potential to develop science and technology, because they are the only ones capable of understanding the nature of this nature in a limited way. Their ability to understand the natural phenomena systematically, rationally and objectively, which later develops the knowledge of science and technology, enabling them to manage and govern the world well.

Science is not a knowledge that can explain everything as claimed by some scholars. It is limited knowledge. At least everyone knows that science cannot explain the aspect of happiness in our life. Science is also unable to explain the concept of human well-being; the emotional and spiritual aspects of a person. It is unfortunate that the over empowering of science has caused this knowledge to dominate all aspects of human life – all knowledge including social sciences and literature needs to be explained through scientific approaches if it wants to be accepted as knowledge universally. Similarly, the knowledge of psychology and humanity must be studied through scientific approach in order to be accepted as knowledge. Most scholars deeply understand that human-related knowledge is subjective and objective in nature, not purely systematic and objective as scientific knowledge. Currently, even religious and cultural knowledge must also be scientifically proven if it is to be accepted universally.

This is the main reason why society tend to ignore the teachings of religion, culture, customs and traditions in their everyday life. Religion and culture are considered as personal matter. Ultimately, it produces scientists that consider science as their ideology which we call it as scientism. It produces the scientists who are blind of religion, culture and customs of their society which undermine the family institution, social well-being and environmental destruction, as well as chaos and disorder in the economic system. This is the consequence of a science that is developed without religion and culture, without the wholistic of humanity that causes them to be blind toward the meaning of life; social well-being, happiness and peaceful co-existence.

As consequence, we produce a system of knowledge that separates science, with the ecosystem of knowledge that enables people to know their religious and culture, and the knowledge that enables people to better understand social responsibility. Some alternative suggestions have been suggested by Gardner (1984) who proposed the concept of Multiple Intelligent; Daniel Goldman (1996) who suggested the Concept of Emotional Intelligence; and the Concept of Spiritual Intelligence by Donah Zohar (2000). But for me, the concept of

human intelligence which they proposed is still beyond the knowledge of revelation as taught in Islam.

There are also people who consider religion and culture unimportant, at least as important as knowledge related to science, technology, engineering, health and economics which they claim, can generate wealth. They also suggest that indicators of progress of a nation should be based on the scientific, technology and economics index, regardless of matters relating to human values.

This paper seeks to examine this title based on the author's observation of today's science education, which also contributes to the overall development of human capital development. This paper also seeks to look the position of science in a wider perspective of the revealed knowledge. We would also like to propose a new concept of science and technology so that it includes cultural, moral and religious values in the science education system – the more divine, environmentally friendly and human-friendly science which we call *Tauhidic* Science.

THE MEANING OF SCIENCE

Before we proceed with our discussion, let as look at the defination of science as we understood today. No single definition can be given to explain what science is. But basically, the method we adopted to understand the nature around us is the fundamental principle of science. The knowledge that we can improve our understanding the nature of nature can be considered as science or in English we call it as 'natural science'. Knowledge related to the technique of using science is also called technology.

To better understand what is science, let's start by looking at terms of 'science' itself. According to Peter Medawar (1984), the word '*science*' is a new term that is not exist in the previous civilization. Therefore to talk about what is meant by science, we will use the meaning given by Western scientists because they are the one who introduce that term.

Traditionally, science as proposed by Western scholars can be defined as (Fowler, 1978) '*Systematic and formulated knowledge'*. The English-Malay Dictionary states that science is (Dewan Bahasa & Pustaka, 1992) '*Systematic study based on observation and* experiment'. To Mortimer J. Adler (1976) '*Science is a search for a rational explanation of natural phenomena. It is continuing activity'*. According to Peter Medawar (1984) the basic

term '*science*' is from the Latin word '*scientia*'; '*scio*' which means 'I know' and '*sce*' which means 'see'. There are 12 words that resemble the term '*science*' as follows,

> *'sienz, ciens, cience, siens, syence, syense, scyence, scyense, scyens, scienc, sciens, scians'*

The word '*scientia*' from Latin can be translated as '*knowledge*', which is the level of 'know' than someone's understand natural phenomena. But not all knowledge can be regarded as science. What is meant by science according to Medawar (1984) is,

> *'The knowledge that the information (as a result of observation on natural phenomena) is generated and developed systematically with a particular methods based on particular premises by its observer to accumulate reliable knowledge, either through experimental or rational arguments'*

To Medawar the truth of the science is the truth based on the goal or objective to be achieved by the work done by scientists through '*asymptote*', which is an inconsistent and not absolute truth, which can still be disputed and criticized, but assumed to be so. Science provides only a direction to which a scientific study can be performed, but has no final goal to be achieved. Thus, the exploration of the knowledge of the natural phenomena studied is constantly changing and not absolute. The history of science has shown to us how science told us that the world is at the end of the horns, but today scientists believe in the 'big bang' theory. Now, people argue whether the big bang theory is still applied. Similarly, the theory of atom is constantly changing when a new theory is introduced later. Dalton's atomic theory which considered atom as smallest particle in an element, then challenged by the theories proposed by Thomson and Rutherford later. Today scientists introduce quantum theory which considers electrons, atoms and other elementary particles as waves. All scientific theories will continue to evolve and change throughout the ages in line with the human ability to understanding nature during its observation. This shows that science is not absolute and the knowledge and theories propose by scientists is constantly changing.

In general, Crump (2002) defines science as,

> *Science is the aggregate of systematized and methodical knowledge concerning nature, developed by speculation, observation and experiment, leading to objective laws governing phenomena and their explanation.*

What is meant by ‘law’ here is a natural rule of thumb that an observed object follows certain rules, which can be studied and can be repeated. For example, studying why an object falls when it is released from a height; why trees need sunlight to grow; why fire burns; why sharp knife cut; why water does not break when chopped, and the study of various other natural phenomena. It turns out, there is a certain rule that needs to be followed and can be understood why the object falls when it is released, why the fire burns, why the sharp knife cuts, and so forth. This natural law is what scientists want to study.

To Shaharir Mohamad Zain (2000), as stated in his book Introduction to Philosophy of Science, the definition of science commonly supported by most scholars can be expressed as,

> *‘Science is the analysis of natural phenomenon systematically, rationally and objectively with the specific devised method to create a reliable knowledge which can be accepted’*

While this definition of science can still be improved and argued, but in general, we can say that pure science is a knowledge that is based on how humans are able to perceive and observe the natural phenomena that behaves according to their nature. This systematic observation result is then argued and analysed in an objective manner, conducted experimentally and make comparison with its assumptions on the theory. This knowledge continues to be nurtured and developed later to add to the treasure of science itself.

THE PREMISES OF SCIENSCE

In order to develop knowledge of science as discussed above, scientists need to make basic assumptions that become the premise of their knowledge. The premise of the knowledge was discussed in detail by Shaharir (1998), indicating its shortcomings in the Islamic perspective, and this is not discussed in this paper. According to Toby E. Huff (1995), pure scientific knowledge requires three basic assumptions.

- First, scientists need to be convinced and believe that properties of nature are regulated and in a certain order. Because it is regulated and organized and in a certain order, this means that the natures are interconnected with one another (coherent), regulated, according to certain rules or laws, and in predictable domains. This is a basic premise in order to develop knowledge of science. Without these assumptions the properties of nature cannot be understood through a scientific approach.

- Secondly, scientific arguments also assume that humans are able to give reasons and causes to the natures he observed. Thus human beings have the ability to understand the properties of the nature by conducting an investigation through rational argument. Yet scientists also believe that certain theories of a nature may be wrong at some point, and at the same time they also believe that they may not be able to know everything about the properties of the nature. But they are convinced that eventually humans can also give their reasons through systematic, rational and objective investigations. Thus, they are also convinced that this scientific investigation has to be done continuously for the sake of purifying their understanding of the nature.
- Thirdly, the philosophy of the natural sciences also assumes that everyone, man or woman, the Eastern or the Western, wherever they are, and with different cultural backgrounds, are permitted to use their intellect that allows them to ask and give reason about a nature they observed. They are free to question anything about the science that they claim. Scientists also believe that after they have scientifically investigated the nature, eventually everyone will give the same conclusions about what they are observing, though they are in different places, with different cultures. With these assumptions, they regard science as a universal knowledge; unlike the knowledge that is related to the culture and religion which localized.

Based on the above definitions and premises of science, we can conclude that to develop science and technology according to the perspective of Western scholar's, we need to pay attention to the following five points,

- This nature has rules and laws that must be followed. Thus a scientist must have the ability to observe the properties of the nature that has certain rules or laws that need to be obeyed.
- Human being has the ability to give reasons and causes why this laws must be obeyed by the nature. The reasons given should be systematic, rational and objective. Objective means 'dealing with outward things or exhibiting facts uncoloured by feelings' (Fowler, 1978). Which means that one should waive all religious, cultural or customary values in making a scientific observation.
- Create or provide a suitable method for understanding the properties of the nature being studied. This method is known as the scientific method. This method must be

accepted and credible enough so that scientific community can accept it. It can be in the form of empirical or in theory.

- Accumulating knowledge about the nature of nature observed. Since scientists believe that scientific study is continuous activity as long as human beings can be systematically, rationally and objectively arguing, then the new knowledge shall prevail from this carefully observation. This will accumulate new knowledge of the same phenomenon previously observed.
- Since scientists are convinced that their study (their understanding of a nature) is not absolute, and can only be explained in detail later by those who are more acquainted with it, they need to express their degree of reliability or accuracy over the properties of nature they describe. In science, the degree of accuracy is expressed in the form of error analysis. There are two things that contribute to the imperfections of the observations which cause error; systematic error caused by the limitations of the tool for careful reading, and random errors caused by the people who observes the nature.

This is the approach used by Western scholars in developing their science knowledge. We can illustrate the Western epistemology of scientific approach as in Figure 1.

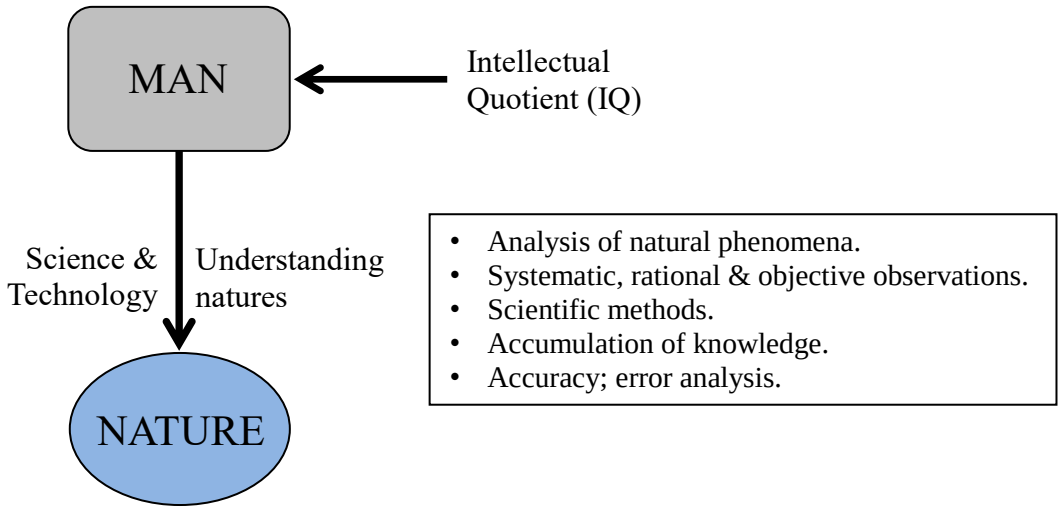


Figure 1: Epistemology of science according to the perspective of Western scholars. Science is intellectual ability or intelligence to understand and discover the nature of nature.

In the Western perspective, science is the knowledge that relates to how human beings able to perceive the properties of worldly nature with their intellectual quotient. With their

understanding of the nature, they are also trying to find the technique to make use the observed nature. This technique is known as technology.

WHAT IS AL-QURAN?

Generally, we can say that the Qur'an is the book containing a set of revelations conveyed to the Messenger of Allah (SWT) to guide all people in their everyday life. According to Subhi Salleh (1978), this book is called the Qur'an as it should be remembered and memorized (*al-Qiraah*) by Muslims. Generally, we can conclude that to a Muslim the Qur'an is as follows,

a. The Qur'an is the book revealed as guidance (*huda*) to all human beings to guide their everyday life. The contents of the Qur'an also clearly state the difference between the right and the wrong ones, as mentioned in verse 185 chapter *al-Baqarah* (2) which means,

   "(Fasting that is enjoined to you) shall be during the month of Ramadhan, in which the Qur'an was revealed as a guidance for the whole of mankind and as self-evident proof for the true guidance, and the criterion (a guidance to distinguish right from wrong)."

   Al-Baqarah (2): 185

   With the guidance that God has revealed, and by stating between true and false, it is hoped that man will make the Qur'an the guideline in handling his activities throughout the day in his life,

a. To Muslims the Qur'an is *kalam* (pen) of God which is a *mukjizat* (miracle) revealed to the Prophet Muhammad and was written, memorized and narrated with *mutawatir* (validity), clear and without doubt. Thus, reading the Qur'an, even without understanding its meaning, is considered worship by all Muslims. This definition is given by Soenarjo et. al. (1412 H) in the Qur'an and its famous translation used as a reference in Malaysia as well as other Malay regions today.

b. Muslims believe that the Quran is also a legitimate, original and authentic book, which has never been challenged by anyone, since it was revealed and bound in the era of Saidina 'Uthman to this day. The legitimacy of its binding was led by Zaid bin Thabit,

together with Abdullah bin Zubair, Sa'id bin 'Ash and Abdurrahman bin Harith bin Hisham (Ishfaq, 2000).

c. The Qur'an is the book that creates the Islamic religious' individual and community. Religion which includes ritual and social aspects. Religions that produce civilized, orderly and systematic human beings in conducting their daily activities. To ensure Muslims continue to awaken, they need to make sure the Qur'an as the source of inspiration, and aspiration in building their future (Dawam Rahardjo, 1996).

d. The Qur'an is the book that must be faithfully and sincerely believed as revealed by God. It is one of the five pillars of faith in Islam. Believing faithfully is the true nature of a Muslim. While questioning facts in the verses of holy Qur'an may jeopardised his *syahadah* (faith). Even though there is a rationale argument which might seem contradict with the verses of the Qur'an; contrary to the thinking of a person or a group of people, such as the story of Prophet Abraham. which was burned by the King of Namrud but did not burn; the story of Jesus a.s. which can bring life to the dead and so on. This is the primary basis for the development of the Islamic faith (Sayid Sabiq, 1991).

SCIENCE FROM ISLAMIC PERSPECTIVE

What is the position of science in the revealed knowledge as contained in the Quran? Muslim scholars never reject the scientific approach as practiced by Western scholars as mentioned above. Even in the Islamic point of view, Muslims are encouraged to observe and study the nature of nature. Verses 190-191 in Surah Ali 'Imran (3) clearly prove this fact, which means,

> *"Surely in the creation of the heavens and the earth, and the alternation of the night and the day there are signs (of Allah's power, wisdom and infinite bounty) for a man of understanding.*
>
> *(That is) those who remember Allah standing or sitting or lying down and reflect on the creation of the heavens and the earth (saying): "Our Lord, Surely, You have not created this in vain. Glory be to You! Save us from the chastisement of the Fire."*
>
> Ali 'Imran: 190-191

Likewise, with verses 27 and 28 in Surah Faathir (35) which means,

> *"Did you not see how Allah sent down water from the sky, whereby We brought forth fruits of different hues? In the mountains there are streaks of various shade of white and red (dark and light), and jet-black rocks,*
>
> *And of men, beasts, and cattle, in like manner have their different colours, too. Verily from among His servants, it is the learned who fear (to go against the command of) Allah. Allah is the Almighty and Most Forgiving."*

Faathir (35): 27&28

In Surah Al-Mulk (67), verse 3, Allah SWT mentions which means,

> *"It is He Who created seven heavens, one above another. You cannot see any fault in the work of the Most Gracious. (If you are in doubt) then look again – can you detect any flaw?"*

Al-Mulk (67), ayat 3

In Surah al-Qamar (54), verse 49, Allah SWT also mentions which means,

> *"Indeed, We have created all things according to the measure (which has been decreed)."*

Al-Qamar (54): 49

The above verses, God clearly asks Muslims to observe and study the events and phenomena of creation of heaven and earth. They are asked to give reason and cause rationally about the phenomenon of the natural occurrence that is in balanced, harmony and according to the rate that Allah SWT decides to occur. Seeing the phenomenon of nature in heaven includes the field of astronomy as it is known today. While the verses in Surah Faathir, Allah SWT asks Muslims to look and think about the occurrence of phenomena of water coming down from the sky, which then grows a variety of plants, and producing varieties of fruits. Likewise, with mountains, structures of stone and minerals that are useful to human beings. Not enough by nature, Allah SWT also urges Muslims to study about human beings, various wild animals and livestock that humans can learn and benefit from.

In the above verse, Allah SWT mentions those who observe the properties of nature (*creation of the heavens and the earth, and the alternation of the night and the day*), there are *ayatillah* (signs of Allah) for a man of understanding, followed by the words '*those who*

remember Allah while standing or sitting or lying down' saying "*O our Lord, You have not created this in vain. Glory be to Thee, then guard us from the punishment of hell*". That means, in Islam, knowledge in relation with the understanding of the nature of nature, which is known today as science, is very important to be learned, so that he can be a good caliph of Allah in this world. But that knowledge of science must not make us forget our responsibility to God that creates the nature of nature. In other words, Islam urges its followers to master science and technology, but mastering of science must not make us forgets our responsibility to Allah SWT.

Those who remember Allah SWT wherever and whenever they are, including standing, or sitting or lying down or anywhere, at the same time can understand the nature of nature created by Allah SWT are called *ulul al-bab* or sensible and wise person. In verses in surah *Fathiir* above, Allah SWT named this group of people as *al-Ulama*. These al-Ulama (the learned) are those who can understand the nature of nature – the phenomenon of rain descending from the sky, which fosters various species of plants, which can understand the nature of mountains and its rocks, who understand the atmosphere, and who understand the diversity of human beings, wild animals and livestock. In other words, they are the natural scientists and social scientists as we understand today, but at the same time they are the most feared people to Allah SWT; who fear His will, fearing His warnings, obeying His commands, who did not go beyond the limits, and who do not feel that they are mighty because of their ability to understanding the nature of nature and human nature. They know that Allah SWT is more valiant and mighty; and who immediately apologizes and seek forgiving from Allah SWT, if they are inexperienced in understanding the nature of nature.

In other words, Muslim scientists aware that this nature is inherent, there is a certain nature that can be understood and predicted, but the nature of this nature is not absolute, and is not determined by nature (whether plants, animals, matter or human beings), but the absolute properties of nature are determined by Allah SWT. That's why Muslim scientists or *Tauhidic* scientists believe that the heat of the burning fire burns, but the burning of objects that are exposed to the hot flame does not lie in the fire, but lies in the consent of Allah SWT to burn. Therefore, a Tauhidic scientist believes in the story of the Prophet Ibrahim AS as mentioned in the Qur'an, which is not burned when cast into a flaming fire because God Almighty does not allow the fire to burn Prophet Ibrahim AS. Similarly, the story that the Qur'an states about Prophet Ismail AS who was slaughtered with a sharp sword by his father

Prophet Ibrahim AS, but did not cause him to die; The story of the Prophet Musa AS who crossed the Red Sea just by throwing his stick, and other stories that were not acceptable by rational thought, were all believed by the Tauhidic scientists who believed in the Qur'an without any doubt.

Tauhidic scientists also believe that they have the ability to comprehend the nature of nature systematically, rationally and objectively with limited knowledge because they have been entrusted by Allah SWT to become the manager and administrator of this world – they are the caliph of Allah SWT. They may not be able to perform their duties properly as caliph if Allah SWT does not empower them with the potential to understand the nature of this world. At the same time, they (as the caliph of Allah SWT) need to sharpen their talents and skills in understanding the natures entrusted to them. Thus, to them, the obligation to develop science is considered as a duty of *fardu kifayah*.

For Tauhidic scientists, knowledge is not limited to what can be observed by human intellect alone. The knowledge that can be responded to by human reason is called *acquired knowledge*, while the knowledge given by Allah SWT to a person through reading the Qur'an is referred to as *revealed knowledge*. Both of these knowledge is indispensable in the life of Tauhidic scientist. Acquired knowledge is the knowledge gained by a person based on the power of reasoning, while revealed knowledge is the knowledge given by Allah SWT based on guidance from Allah SWT when he seeks knowledge by reading the Qur'an. The strength of the revealed knowledge is based on the extent to which one's faith develops his Tauhidic aspect – a belief system on Oneness of Allah SWT based on the ritual and worshiping practices of a person.

Thus, for the Tauhidic scientists, there is hierarchy in knowledge. The highest hierarchy of knowledge is knowledge that can introduce himself to his God. This knowledge not only allows one to know his God, but also makes all his practices done because of the God he knows. The hierarchy of knowledge in Islam are discussed in detail by Imam al-Ghazali (1990) in his book *Mau'zatul Mukminin* (the Guidance for Mukminin) and ibn-Khaldun (1993) in his most famous book, *Mukaddimah*.

In conclusion, the epistemology of Tauhidic science that is based on the divine revelation is illustrated in Figure 2.

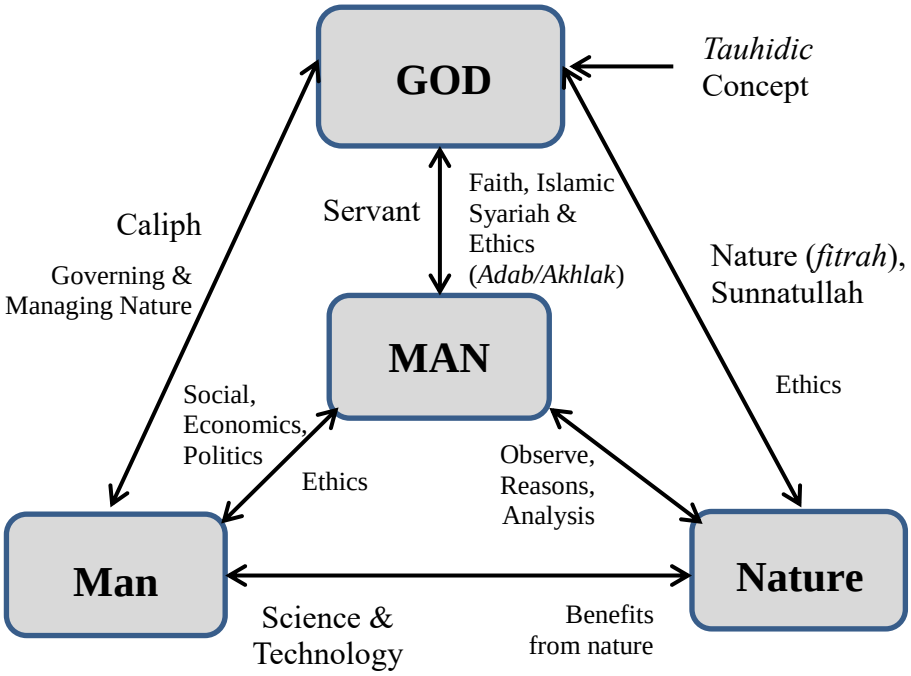


Figure 2: Epistemology of Tauhidic science that illustrates the relationship between human beings with Allah SWT, between man with other human beings, and between man with nature in its environment that develops science and technology.

The notable difference between science from the Western perspective and Tauhidic science is that Western science rejects any form of religious argument, although their scientists understand that their knowledge is limited, not absolute and secular, while Tauhidic's science is a science that is of particular concern to the divine, even the spirit and their final goal of developing science is simply because of Allah SWT. All innovations, creativity and inventions produced by Tauhidic scientists are manifestations of his own responsibility to his God. Thus, the destruction, uncivilized and destructive creativity of the environment and the well-being of life is strictly prohibited.

CONCLUSION

The world badly needs the development of science and technology. The development of science and technology will generate economic growth and hopefully, will bring prosperity to a nation. But unfortunately, the development of science and technology that ignores the development of culture and religion makes people blind in the direction in which the development will take place. Prosperity and luxury are merely a mirage in the arid desert, not

a real peace that comforts and reconciles one's heart and soul. As a result, we see in front of us now, how high-literate society is unable to address environmental issues and the problems of global warming; problems related to speculation and manipulation of currency in order to generate wealth; and issues related to moral, religious and cultural-based values among society. Even the white collar crime that is in front of us, which also undermines the world economic system today as stated by Francis Fukuyama in his writings The Fall of America Inc., in Newsweek October 13, 2008, teaches us how poor the system is based on scientific achievement but neglecting responsibility to God. Therefore, we will be 'blinded' if we persistence build science and technology without consideration on the need of culture and religion.

On the other hand, religious development also requires the development of science and technology. What does it mean to us as a caliph of Allah SWT who is entrusted to manage and administer this world, but fails to understand the nature of nature well? How can this nature be managed and administered well if we are ignorant of the nature of nature? How can we make clothes to cover our body as our religion obligations if we fail to make proper clothing? How can we fulfil all the practices contained in the pillars of Islam if we do not master science and technology? That's why religion will be lame without science and technology.

ACKNOWLEDGEMENT

The author would like to record his appreciation to the Universiti Kebangsaan Malaysia which allows this paper to be written under the funding of GPP-2011-011 and STEM-2014-002. The author also wishes to extend his gratitude to the organizational committee of the 7th Asia-Pacific Conference on Philosophy of Science for inviting and sponsoring the author to actively participate in the conference.

REFERENCES

Abdullah Yusuf Ali, 1989. *The Meaning of The Holy Qur'an*. Maryland, Amana Corporation. All translation of verses from al-Qur'an are referred to this reference.

Adler M.J., 1976. *Great Ideas from the Great Books*. N. York. Washington Square Press.

Crump T., 2002. *A Brief History of Science: As Seen Through The Development of Scientific Instrument*. London, Robinson.

Dewan Bahasa dan Pustaka, 1992. *Kamus Inggeris-Melayu Dewan*. Kuala Lumpur. Dewan Bahasa dan Pustaka.

Fowler F.G. and Fowler H.W, 1978. *The Oxford English Dictionary.* UK, Oxford University Press.

Fukuyama F., 2008. *The Fall of America Inc.*, dalam Newsweek keluaran 13 Oktober.

Gardner H., 1984. *Frames of Mind: Theories of Multiple Intelligences*. London, Heinemann.

Goleman D., 1996. *Emotional Intelligence: Why It Can Matter More Than IQ*. London, Bloomsbury.

Huff, T.E., 1995. *Islam, Science and Fundamentalism*. J. of Arabic, Islam and Middle Eastern Studies, Vol 2(2). pp. 1-27.

Ibn Khaldun, 1993. *Mukadimah*, Kuala Lumpur. Dewan Bahasa dan Pustaka.

Imam Ghazali, 1990. *Bimbingan Mu'minin*. Singapura, Pustaka Nasional Pte. Ltd.

Medawar P., 1984. *The Limits of Science*, Oxford University Press, UK.

Mohd. Yusof Hj. Othman, 1998. Isu dalam Ilmu dan Pemikiran. Kajang, Aras Mega Sdn. Bhd.

Shaharir Mohamad Zain, 2000. *Pengenalan Sejarah dan Falsafah Sains*. Bangi, Penerbit Universiti Kebangsaan Malaysia. Hlm 25.

Shaharir Mohamad Zain, 1998. *Kritikan Awal Kepada Premis Ilmu Sains Tabii*. Kesturi 1(1).

Zohar D. and Matshall I., 2000. *Spiritual Intelligence – The Ultimate Intelligence*. London, Bloomsbury.

Woo-Chong, K.1993. *Kekayaan Ada Di Mana-mana*. Shah Alam, Times Subang.

MALAY VALUES IN SCIENTIFIC INQUIRY[1]

Abdul Latif Samian
Institute of Civilizational Islam
Universiti Kebangsaan Malaysia
abdlatif@ukm.edu.my

Abstract

From the Malay perspective, as a general teory of ethics and values, God is *The Good*, i.e., *Yang Maha Baik*. Whatever that is good entails from The Good. There are a plethora of hierarchical goodness, either in the tangible or intangible forms and the dominant Malay worldview is founded on the Unity of God by way of the teachings of Islam. In this paper, the author examines the values of scientific inquiry espoused by scientists and scholars (particularly the Malay thinker Hamka) in the Malay world and civilization, taking into account both the esoteric and exoteric values of the Malays and the position of Islam in their worldview.

Keywords: The Good, Ethics, Values, Truth

[1] Paper presented at The Asia-Pacific Conference on Philosophy of Science (APCPOS) organized by National Chung-Cheng University, Taiwan, 15-16th December 2017. The author is grateful for the grant received from the organizer.

1. INTRODUCTION

Generally speaking, Islamic values are fundamentally non-other than those values that benefits humanity because Muslims believe that the most valuable people are those that are most valued by others since anything that cause harm to others is considered unIslamic. Values such as striving for excellence, mutual respect, respect for elders, accommodativeness (inclusion), seeking knowledge, punctuality, trustworthiness, justice, and sincerity are praiseworthy whereas betrayal, slander, envy, and avarice are blameworthy (Hamka, 2007).

Islam advocates a well-balanced change, development and progress based on the saying of the Prophet that "The best in all things is their mean". The ultimate objective in all endeavors is the mean since both extremes are blameworthy precisely because submitting to the extremes will cause imbalanced. It is by way of striving for the total educational experience that the mean is achieved, bearing in mind that what is 'the mean' is contextually defined, i.e., relative to the existing fundamentals of the person.

From an Islamic perspective, an ethical act is a Godly act. An act is virtuous if it is done with a noble intention and praiseworthy consequence. Therefore the purity of intention which is further based on faith is a significant matter in Islam. In fact those who are perceived to be good must be construed as having a superior moral character.

Thus from the theological point of view it is not possible to have an unethically good muslim. In so far as professionalism is concerned, the prophet says that "God loves those who do their best when they perform any given duty." This is the 'right' or 'Godly' attitude so to speak. Thus we have the ethical concepts of goodness (khayr), righteousness (birr) and striving to achieve distinction (itqan). The prophet says that "success comes with patience, relief with affliction and ease with hardship." Adhering to noble values is a time honored approach to ensure the right scientific decisions in problem solving are taken. As the saying goes, if one does not know what harbor he seeks, any wind is the right wind. However, values alone are inconsequential. It has to be translated into acts in order for scientists to realize the lofty objectives of their scientific endeavor.

2. MALAY TRADITIONAL VALUES

Unlike Malays in other countries, all Malays in Malaysia are Muslims. The Abbasids coin dated 8AD was discovered at Lembah Bujang (Bujang Valley) in northern Malaysia. The Terengganu Islamic Inscription (13AD) is another evidence apart from the discovery of tombstones. The significance of religion can be gauged from the traditional Malay children *pantun* '*Halia ini tanam-tanaman, ke barat juga akan tumbuhnya, Dunia ini pinjam-pinjaman, Akhirat juga akan sungguhnya*'

(Garlic is not a tree, yet to the West it will grow, the World is but temporary, the Hereafter is eternal indeed). Suffice is to say that Islam is part and parcel of the Malay *weltanschauung*.

Malays are pragmatic without adhering completely to pragmatism as its foundation of ethics. The value of the proverb, *masuk kandang kambing mengembik, masuk kandang lembu menguak* (be a goat with the goats, be a cow with the cows) has a limit. Absolute pragmatism-in Rome do as the Romans do- is not an integral behavioral principle of the Malays. Otherwise, there will not be the aphorism *adat bersendikan syarak, syarak bersendikan kitabullah* (religion is the cornerstone of tradition, the Holy book is the corner stone of religion).

Generally, Malay accept their fate and destiny with open hearts. As Muslims they believe that sometimes God works in a mysterious way. There must be a *hikmah* (a divine intervention) if things do not turn out as planned. They do not espoused the *jabbariyyah* doctrine (resigning completely to fate) neither do they subscribe fully to the *qadariyyah* position (man is the measure of all things). Otherwise there will not be the proverb *alang-alang menyeluk pekasam biar sampai ke pangkal lengan , alang-alang berdakwat biar hitam* (if you must dig for the fruit sauce, dig all the way to the bottom of the vase, if you must write, write really well) or the aphorism *kalau tidak*

dipecahkan ruyung masakan dapat sagunya (if you do not break the bark of the sago tree, how could you get the sago). Ditto with the proverb *sedikit-sedikit lama-lama menjadi bukit, sehari selembar benang lama-lama menjadi kain, sehari tulis sesuku lama-lama menjadi buku* (bit by bit the hill is made, thread by thread the cloth is sown, word by word the book is written). Malays writ large subscribe to perennial value of the middle path of these two theological positions-*manusia punya asa tapi Tuhan punya kuasa* (man proposes, God disposes).

3. SCIENTIFIC PROBLEMS AS RELIGIOUS PROBLEMS

First and foremost, Malay scientists and scholars such as Sheikh Abdul Latif al-Khatib al-Minangkabawi (1855-1915) in his book *'Alam al-Hussab fi ilm al-Hisab* and Sheikh Nuruddin Muhammad Jailani Ibn Hasanji Ibn Muhammad Hamid ar-Raniri (d.1658 (1068H)) in his *Bustan as-Salatin* never uses the word 'science' in the sense that the word is understood today; that knowledge which is 'exact', objective, veritable, deductive and systematic, especially the latter's view concerning *Tibb an-Nabawi*. The closest term that they ever used is the Arabic word '^{c}ilm', which also means knowledge. Al-^{c}ilm in the language of the Quran and *sunnah* (traditions of the Prophet) implies knowledge which makes man conscious of God, of His attributes, of the eternals, of the next world and of the return to Him and Him alone.

Taking into account the comprehensiveness of their scientific views, science according to them is a problem solving activity comprising of aspects of intention and actions prior to arriving at the solutions.

Scientific Problems→Intention→Values→Experimentation (Trial and Error) →Scientific Solutions (New Scientific and Technological Knowledge)→New Problems

Fig. 1. Process of Scientific Creativity (Abdul Latif, 2009)

Scientists seek solutions to scientific problems. A scientific problem to them is a problem circumscribed by the Holy Quran and as-Sunnah which is enjoined by God. It is a problem that arises and needs to be solved in order for a Muslim to improve his '*taqwa*'. It is also a problem posed by an Islamic society, arising out of their efforts to practice Islam as correct, and as accurate as possible, in order to please God. The orientation of the problem determines its 'scientificity'. A scientist, to al-Khatib and ar-Raniri, does not solve a scientific problem simply for the sake of solving problem. He does not solve a problem because the problem ought to be solved since it is technologically possible to do so. His motive of solving problems is dominated by this consciousness of seeking God's pleasure, "that which yields Him satisfaction".

The interesting thing is that the evaluation of the problems tackled is not given *post hoc* or *ad hoc.* It is not the case that al-Khatib solved scientific problems before thinking of its necessity, its worthiness for the ummah, *viz.*, its legitimacy from the Quranic and *Sunnah* point of view. Thus,

> Tatkala ilmu hisab tinggi darjatnya kerana tiada terkahaya daripadanya tiap-tiap manusia pada musafirnya dan tiada pula pada ketika mukimnya kerana segala manusia berkehendak kepada makan dan minum dan ibadat…

Clearly, to al-Khatib, there is a 'sacred' orientation in scientific problems. Al-Khatib shows that there is a "revealed perspective" on scientific problems which the scientist should take into account and in this case the Divine Name of God, '*al-Hasib'* (The Reckoner).

Moreover, from the perspective of the Quran and *Sunnah*, nature and history (the days of God (*ayyam Allah*)), can enlighten man in knowing more about himself and his Creator. Says the Holy Prophet: "He who knoweth his self knowest his Lord". The Holy Quran views the alternation of night and day, the lengthening of shadows, the variation in human colour and language, the vicissitudes of nations, as signs of God that warrant examination in our quest of knowing Him (al-Qur'an, 10:6). The science of mathematics to al-Khatib, for an example, has its origin from Prophet Idris (the Biblical Enoch).

There is an element of transcendence in seeking scientific solution to the end that problem solving is an act of contemplation. Al-Khatib and ar-Raniri, more often that not, is always conscious of God while solving problems. He strives to be among those who "… remember Allah, standing, sitting and reclining and consider the creation of the heavens and the earth, (and say): O Lord; Thou created not this in vain". (al-Qur'an 3:191). While it is not the case that all problems of religion are scientific problems, all scientific problems are religious problems (Abdul Latif, 1999).

4. NATURE OF SCIENTIFIC AND TECHNOLOGICAL KNOWLEDGE

From the practical point of view, both al-Khatib and ar-Raniri believe that science is governed by religion because science is extremely useful in solving problems of humanity. Mathematics for example, enhance prayer, almsgiving, distribution of wealth and so forth, the many virtuous esoteric and exoteric acts in Islam. In this sense, mathematics, as any other scientific disciplines, is circumscribed by religion. There is an organic relationship between mathematics and that of Islam.

Eventhough al-Khatib concedes to the importance of the intellect in acquiring mathematical knowledge, he does not submit to the philosophical position that mathematical knowledge is localized. Mathematical knowledge is not personal, so to speak. Mathematicians have their own academic comunities and al-Khatib believes that

this social aspect also contributes in the development of mathematical knowledge and values. In other words, if one "proceeds systematically from first principles," the principles of Iman dan the principles of Islam, one can arrive at truth.

Ditto for other disciplines of science and technology in the Malay World, those contemporary included. The aforementioned first principles of Iman and Islam subscribed by al-Khatib and ar-Raniri also applys in the case of bioethics which should be construed as a subset of Islamic knowledge, i.e circumscribed by religion, whereby contribution of modern biotechnology to the betterment of the agricultural sector and human health is undeniable. Thus problems of ethics of modern biotechnology must consist of ethical issues related to God, ethics among human beings and ethics related to living things.

Ethics among human being must include issues of health, food safety and monopoly. From an Islamic perspective, health is viewed as one of the greatest blessings that God has bestowed on mankind "Eat of the good things We have provided for your sustenance…" (al-Qur'an, 20:81). In so far as food safety and technology is concerned, scientists, industries and government regulators must ensure that modern biotechnology and its products is not in anyway harmful "..and make not your own hands contribute to destruction…" (al-Qur'an, 2:195). Food production should not be a monopoly to the extent that the interests of biotechnology companies should not

exceed the general well-being of the people; trade must be based on mutual goodwill "…let there amongst you traffic and trade by mutual good will…" (al-Qur'an 4:29).

5. VIRTUES OF SAFE-GUARDING THE SOUL

Since the pursuit of scientific knowledge is a virtuous act , it is crucial to safeguard the soul in so far as the ethics of science are concerned. Al-Khatib, al-Raniri & others in the Malay World (for examples, Syeikh Daud Abdullah Fathani and Syeikh Muhammad Tahir Jalaluddin) believed that the joy of the spirit in the virtuous act of discovering the true nature of things is better than the pleasures of the flesh since the former is more lasting. Since the scientist does vacillate between the two levels of awareness and forgetfulness of the Good, al-Khatib for example, gives some advice for the mathematician so that the latter can always be in the blessed state.

Al-Khatib maintains that as a seeker of a sacred knowledge, the mathematician should live according to a set of virtues revealed by God through His Prophets (peace be upon them). According to him, the mathematician should be actively involved in solving problems for the society because man cannot live by himself. The code of conduct described by al-Khatib, ar-Raniri, al-Fatani rests on the scientists consciousness of The Divine. They realize this most important axis, the awarenessof

God as the most important aspect that binds and characterizes the scientist's quest of scientific knowledge. The scientist should be mindful of God as much as he can. His private and public life should be in accordance to the famous saying of the Holy Prophet : "that you should worship God as though you saw Him…" (*an ta^{c}buda Allaha ka 'annaka tarahu* ...). In other words, the scientist should always be in a state of gratitude to his Lord.

In light of their views on virtuous conduct for the scientists, we should never interpret that their problem solving activity equals to the utilitarian normative ethical doctrine. Utilitarians maintain that if a scientist is faced with a number of scientific problems related to the society, he should prefer solving the problem that can promote the greatest happiness of the greatest number irrespective of guidance from the scripture. Choices are judged by their consequences and the amount of pleasure derived from those consequences. Clearly his code of ethics cannot be called utilitarian because choices are never analysed entirely through actions and consequences. Rather, motives and underlying *intention* are crucial in his problem solving approach. As we have shown earlier, al-Khatib for example believes that problems are religiously defined. From the external aspect, problems are solved for the betterment of the society but to al-Khatib, the welfare of the society is never the endpoint. The endpoint, the ultimate

cause, the foremost reason problems are solved by the mathematician is so that both he and the virtuous society will enjoy continuous Divine Blessing. There is an equally important esoteric aspect to it.

Al-Khatib and al-Fatani view mathematics as a very powerful tool of studying nature. However to say that they were instrumentalists as the word is understood today would not do justice to their philosophies of science. Instrumentalists believe that in the case of mathematics, the latter is nothing more than a tool in our noble quest of knowledge where as al-Khatib believes that mathematics has an important role in man's understanding of the relationship between nature, science, religion and in order for him to become a virtuous man. Nature can be scientifically analyzed through mathematics and religion plays a critical role in some of the processes. In more specific terms, mathematics as practiced by al-Khatib are circumscribed by religion wherein the mathematician is immersed above all, from observing God's handiwork in deciphering nature with the consequence of knowing more about his mode of existence and as a matter of fact, about Existence Itself.

6. CONCLUSION

In view of their philosophies of science, there must be a primary link that connects ethical problems and the spiritual realm. Guided by this basic belief that eveything is

rooted in the Divine, scientists' contemplation of scientific problems are facilitated by studying the Scripture; through which they can know and internalize the qualitative aspects of God in solving those problems (Hamka:2007). Ethical problems, in turn, are related to God in a manner corresponding to their mode of existence. The Divine is the beginning and the end of all scientific problems there is. From the aspect of Divine Unity, theology is central to their frame work and thus it functions as the dominating factor. More important than that, it is a consequence of their deep rooted belief and knowledge in the ever encompassing, ever knowing God; the Absolute Good. The Good is the Supreme Reality, The One (al-Qur'an 112:1-4) who is at once transcendant and immanent, yet "nothing is like Him" (al-Qur'an 42:11). Indeed, we are from the The Lord (al-Rabb) but not part of The Lord (Abdul Latif:2012).

Ethics is concerned with practice, with human decision and conduct in solving problems. Scientists have to decide between goals of action, objects worth pursuing or need to be avoided. That acts done for the sake of The Good is the only act that is good for its own sake can be proved from the fact that no two events are necessarily consequential yet the goal for any act is always to experience goodness. Granted that our existence is contingent upon the existence of God, what more of moral events? If the purpose of moral deliberation is to find a reasonable ground of obligation, what is more reasonable than to act for the sake of the The Most Reasonable? Likewise, if the

objective of the inquiry of moral value is to evaluate desirable acts, what is more desirable than that which is more desireable to the Absolutely Desirable?

The non-humans, part of nature or the phenomena which are the objects of scientific inquiries, those at the end of the chain of being-automobile, painting, medicine, money, biotechnological products *et cetera*, are nonetheless variety of goods by virtue of extension. Their values are derivative or more specifically, dependent on the self and its various manifestations guided by the principles of faith and principles of religion.

In conclusion, in order to avoid misunderstanding, we must reiterate that the ethical framework espoused is not a reductionist kind. It is also a version of axiological pluralism, with the unifying presence of The Divine as the common denominator, which in essence is The Good. In this framework, The Absolute Good remains and is the Most Universal, the Ultimate Source of all goodness.

BIBLIOGRAPHY

Abdul Latif Samian. 1999. *Falsafah Matematik*. Kuala Lumpur: Dewan bahasa dan Pustaka.

Abdul Latif Samian. 2000. Falsafah Matematik Ahmad bin Abdul Latif al-Khatib. *Jurnal Akademi Sains Islam Malaysia*. Jilid 10 (1&2); hal.52-67

Abdul Latif Samian.2002. *Falsafah Moral dalam Etika Melayu*. Pemikir. (28)(73-90).

Abdul Latif Samian, *et. al*. 2009. Islamic ethics and modern biotechnology. *International Journal of the*

Malay World and Civilisation.(27)(2) Pp.285-29

Abdul Latif Samian. 2009. 'Pemikiran Saintifik', dlm Che Husna Azahari (ed.) *Sains dan Teknologi di Alam Melayu.* UKM: Institut Alam dan Tamadun Melayu.

Abdul Latif Samian.2010. Sains Ketauhidan Dalam Melestarikan Tamadun dlm. Rokiah@Rozita Ahmad et.al,(ed) *Prosiding Bengkel Pengajaran Sains Tauhidik*. Bangi: Universiti Kebangsaan Malaysia.

Abdul Latif Samian. 2012. *Memetakan Metamatematik*. Bangi: UKM Press.

Abdul Rahim Masaridi & Mohd. Alif Redzuan. 2006. Perhubungan Alam Tumbuhan, Mistik dan Ayat al-Quran dalam Tradisi Perubatan Tradisional. *Kertas kerja Seminar Sains, Agama dan Budaya di Alam Melayu.* Dewan Bahasa dan Pustaka dan Universiti Malaya.

Ahmad al-Khatib. 1895. '*Alam al-Hussab Fi'ilm al-Hisab*. Mesir; Kaherah.

Ahmad Daudy.2006. *Sheikh Nuruddin ar-Raniri: Sejarah hidup, karya dan pemikiran*. Banda Aceh:Pusat Penelitian dan Pengkajian Kebudayaan Islam.

Hamka.2007. *Tafsir Al-Azhar*. Singapura: Pustaka Nasional Pte Ltd.

Nur aldin al-Raniri. *Bustan al-salatin*. 1966. Ed. Teuku Iskandar Kuala Lumpur; Dewan Bahasa dan Pustaka.

JABIR IBN HAYYAN: THE ISLAMIC PHILOSOPHY OF THE FATHER OF CHEMISTRY

Ibrahim N. Hassan*, Mohd Yusof Hj Othman, Abdul Latif Samian

Institute of Islam Hadhari, Universiti Kebangsaan Malaysia, 43560 Bangi, Selangor, Malaysia.

*ibnhum@ukm.edu.my

ABSTRACT

Nearly 3000 cursive about chemistry, as well as several other sciences, was found belonging to the father of chemistry, Jabir ibn Hayyan. The foremost Muslim alchemist was born c. 721, Ṭūs, Khurasan and died c. 815, Al' Kūfah, Iraq. Jabir was Ja'far as-Sadiq's most noticeable student and a colleague of Imam Abu Hanifa, the founders of the Sunni Hanafi School of fiqh (Islamic jurisprudence). In addition to chemical and laboratory equipment and apparatuses, Jabir has developed a lot of chemical compounds, as well as medicines, aiming to help his people who suffer from diseases. The Jabirian corpus is renowned for its contributions to alchemy. It perfectly expresses the recognition of the importance of experimentation, "The first essential in chemistry is that thou shouldest perform practical work and conduct experiments, for he who performs not practical work nor makes experiments will never attain to the least degree of mastery.Therefore, in this paper, we will try to look at Jabir ibn Hayyan from Islamic point of view, attempting to discover his philosophy as a Muslim Chemist and how he developed Chemistry based on his viewpoint as a Muslim.

INTRODUCTION

Arab Muslims knew chemistry since the first century AH / seventh century, has led them to engage in this concept books chemistry early, and this explains that the first book was transferred to the Arabic language was a book in chemistry. Islamic conquests has played an important role in opening eyes on forms of literature in the science of chemistry, including books on the gold

industry, and the types of various chemical processes, so that was the outcome of Muslims work in chemistry field - in the end – rather than physics field.

Muslim Contribution to Chemistry

Before addressing the subject of Muslim chemistry, however, one crucial matter needs to be raised. It concerns the use of the word Alchemy instead of chemistry. This is another instance of historical corruption fooling so many who have no perception of the depths some scholarship can descend to in order to convey distorted images of aspects of history, such as that of Islamic science. Alchemy, indeed, is a corrupt translation of the Arabic word Chemia (chemistry,) proceeded by the article Al (which means: the), and which the Arabs always use (like the French and others for that matter) in front of their subject such as Al-Tib (medicine) al-Riyadiyat (mathematics) etc... If this was applied to other subjects, it would become al-medicine; al-mathematics, al-geography and so on... Only Baron Carra de Vaux had had the presence of mind to pointing to this, however briefly. Somehow al-Chemy should be translated literally The Chemistry and not Alchemy in English; and La Chimie and not l'alchimie in French. The fact that only Westerners translated or dealt with the subject, followed by rather very respectful or shy Muslim scholars means that this corrupt word of al-chemy has remained, and has become the norm.

The reason why alchemy is used instead of chemistry might have another motive behind it. Chemistry means a modern science; alchemy means the amateur, the occult, the second or third rate. Alchemy belongs to the Muslims; chemistry, of course, does not; instead is the realm of the good. This notion conveyed by some Western scholars, that alchemy ended with the Muslims and chemistry began with the Westerners has no historical ground. The reason is simple: all sciences began in some part of the world, most likely China or the Ancient Middle East, or India, at level: 1, the most basic, and then graduated to levels 2, 3, 4, and higher, through the centuries, until they reached us at the level they are, and will evolve in different places in the future. This is the story of every science, and of every sign of our modern world. Thus, it was not that we had alchemy at one point, and then, with the Europeans it became chemistry. This is a crass notion like much else coming from scholars holding such a view. Chemistry began under one form, associated with

occult and similar practices, and then evolved, gradually becoming more refined through the centuries until it took our modern forms and rules. Many elements concourse to support this point; here they follow.

Muslims Revolutionized Chemistry

First and foremost many of the products or discoveries made by the Muslims have become part of our modern chemical world; in fact were revolutions in the advance of the science. Mathe summarizes the legacy of Muslim chemists, which include the discovery of alcohol, nitric and sulphuric acids, silver nitrate and potassium, the determination of the weight of many bodies, the mastery of techniques of sublimation, crystallization and distillation. Muslim chemistry also took many industrial uses including: tinctures and their applications in tanning and textiles; distillation of plants, of flowers, the making of perfumes and therapeutic pharmacy. More specifically, some such advances that have revolutionized our world are expertly raised by Multhauf (1919-2004). Thus in the De aluminibus, composed in Muslim Spain, (whose author Multhauf does not recognize) but could be Al-Majriti, are described experiments to obtain the chloride of mercury, corrosive sublimate ($HgCl_2$), process and outcome which mark the beginning of synthetic chemistry. Multhauf notes indeed that the chloride of mercury obtained did not just become part of the chemist's repertoire but also inspired the discovery of other synthetic substances. Corrosive sublimate is capable of chlorinating other materials, and this, Multhauf, again, notes, marks the beginning of mineral acids. In the field of industrial chemistry and heavy chemicals, Multhauf notes again that one of the greatest advances of the medieval times was the manufacture of alum from `aluminous' rocks, through artificial weathering of alunite, which he describes. And in the same context the Muslims managed to perform the crystallization of `ammonia alum' (ammonium aluminum sulphate). Multhauf, however, falls in the same trap as many of his colleagues, asserting in his conclusion that it was European Renaissance which gave chemistry a secure and significant place in science, and that with the Muslims all that was, was `alchemy;' and Multhauf states this in full contradiction of what he had just described, and so expertly, and he had himself classified under modern chemistry.

Fair Historians of Chemistry

A scholar who from the initial point gave Islamic chemistry its due, and hardly failed to call it so, was Holmyard (Holmyard, E. J. 1929). Holmyard, indeed, has the right qualifications to discuss Islamic chemistry, and more than any other scholar, with the exception of Ruska, and also Levey. Holmyard is indeed both a chemist with great renown, and also an Arabist in training, rightly qualified to look at the science from the expert angles, unlike others, who are either Arabists, and so understand little in chemistry, or are experts in chemistry and understand nothing in Arabic. Holmyard notes that the rise and progress of Islamic chemistry is given very little space, and whatever information exists is erroneous and misleading, a fact due partly to Kopp's unfavorable opinion of Islamic chemistry, and the hasty conclusions drawn by Berthelot from his superficial studies of Islamic material. And neither Kopp, nor Berthelot were Arabists, which, as Holmyard notes, makes their conclusions on Muslim chemistry unable to stand the test of criticism as more information is available. Of course, today's scholars can always ignore evidence that has come out since Kopp and Berthelot, and still stick with their misinformation, errors, or distorted statements, and blame such on either one of them. This tactic is in fact very common amongst scholars writing in any field of history, who shape and reshape events at will and have all the necessary sources and references to justify their writing. Some `scholars' even go as far as blaming the material in the library of their university, stating in their preface or conclusion that any shortcoming in their work was the result of their access to such limited material.

To return to Holmyard, in his Makers of Chemistry, tracing the evolution of the science from the very early times until our century, and even if not having at his disposal the vast amount of information many of today's scholars have, he produced an excellent and encompassing, thorough work. It includes none of the usual gaps of centuries one finds with other historians; nor does it include the discrepancies caused by 'sudden', 'enlightened' `miraculous' breakthroughs out of nothing.

Transmission of Chemistry to Europe

Of course Muslim chemistry, like other sciences was heavily translated into Latin, and also into local languages, which explains its spread to Europe (more on this in the chapter on the transfer of Muslim science to Europe). Many of the manuscripts translated have anonymous authors. Of the known ones, Robert of Chester, a twelfth century scholar, translated Liber de compositione alchemise. At about the same time, Hugh of Santalla made the earliest Latin translation of lawh azzabarjad (the Emerald table). Alfred of Sareshel translated the part of Ibn Sinna's Kitab al-Shiffa (the Book of Healing) that deals with chemistry. It is, however, as per usual, the Italian, Gerard of Cremona, who made the more valuable translations of Al-Razi's study and classification of salts and alums (sulphates) and the related operations the De aluminibus et salibus, whose Arabic original is preserved. The many versions of this work had a decisive influence on subsequent operations in the West, more generally on mineralogy; as did others in the formation of the foundations of such science. In fairly recent times, Holmyard, Kraus, and above all Ruska, have devoted considerable focus to Muslim chemistry, much of which, unfortunately, is not accessible to non-German speakers, who thus will be deprived from forming a truest picture of Islamic chemistry.

After such an expose, however brief, should we still consider Muslim chemistry as an occult or magical practice called alchemia? Are not many aspects of such science exactly what we have in our modern chemistry? And if this is not enough, here is what Muslims thought of the occult alchemia. Both Ibn Sina and Ibn Khaldoun attacked the experimentalists who sought to turn ordinary metals into precious ones, gold in particular. Ibn Sina, for instance, in The Book of Minerals, denounces the artisans who dye metals in order to give them the outside resemblance of silver and gold. He asserts that fabrication of silver and gold from other metals is `practically impossible and unsustainable from a scientific and philosophical point of view.' Ibn Khaldoun, for his part, denounces the frauds who apply on top of silver jewelry a thin layer of gold, and make other manipulations of metals. To Ibn Khaldoun, the Divine wisdom wanted gold and silver to be rare metals to guarantee profits and wealth. Their disproportionate growth would make transactions useless and would run contrary to such wisdom.

JABIR IBN HAYYAN

Nearly 3000 cursive about chemistry, as well as several other sciences, was found belonging to the father of chemistry, Jabir ibn Hayyan (12). Jabir was born and educated in Tus, and he later traveled to Kufa south of Iraq. He is generally known as the father of chemistry, and has contributed a lot in the field of chemistry. He introduced experimental investigation into alchemy, which rapidly changed its character into modern chemistry. His contribution of fundamental importance to chemistry includes perfection of scientific techniques such as crystallization, distillation, calcinations, sublimation and evaporation and development of several instruments for the same. The fact of early development of chemistry as a distinct branch of science by the Arabs, instead of the earlier vague ideas, is well-established and the very name chemistry is derived from the Arabic word al-Kimya, which was studied and developed extensively by the Muslim scientists (7&8).

The seeds of the modern classification of elements into metals and non-metals could be seen in his chemical nomenclature (9). He proposed three categories:

• "Spirits" which vaporise on heating, like arsenic (realgar, orpiment), camphor, mercury, sulfur, sal ammoniac, and ammonium chloride.

• "Metals", like gold, silver, lead, tin, copper, iron, and khar-sini (Chinese iron).

• Non-malleable substances, which can be converted into powders, such as stones.

The origins of the idea of chemical equivalents might be traced back to Jabir, in whose time it was recognized that "a certain quantity of acid is necessary in order to neutralize a given amount of base". (10).

Despite the research carried out by Jabir Ibn Hayyan was in the field of chemistry, he was a prominent chemist and alchemist, physician, pharmacist, physicist, philosopher, geographer, engineer, astrologer, and astronomer (22). Yet, his intention was to solve the problems of mankind, which is, In fact, the noble mission of Islam and noble and duty of Muslims. Hence, it can

obviously be understood that it was a manifestation of the implementation of his responsibilities as a Muslim.

Conclusion

Jabir Ibn Haiyan, the father of chemistry, has contributed a lot in the field of chemistry. He introduced experimental investigation into alchemy, which rapidly changed its character into modern chemistry. His contribution of fundamental importance to chemistry includes perfection of scientific techniques such as crystallization, distillation, calcinations, sublimation and evaporation and development of several instruments for the same. The fact of early development of chemistry as a distinct branch of science by the Arabs, instead of the earlier vague ideas, is well-established and the very name chemistry is derived from the Arabic word al-Kimya, which was studied and developed extensively by the Muslim scientists. Muslim scholars have developed and transferred Chemistry, as well as other sciences, from Greece civilization to us; However, Crusades were the reason behind the demise of Muslims contribution in science. It is, thus, time to give Muslim chemistry its due place in history. For that to happen, the concentrated effort of Arabic speaking, able scholars, with some honesty, ought to get on with the task of writing truest accounts of Islamic chemistry in history, do for this science what Rashed, Djebbar and Yuskevitch did for Islamic mathematics, or what al-Hasan and Hill did for Islamic engineering, and what King, Saliba, Kennedy and Samso seek to do for Islamic astronomy, bringing Islamic chemistry out of the slumber others have dug in for it.

References

Al-Majriti, M. https://en.wikipedia.org/wiki/Maslama_al-Majriti

D. R. Hill. 1993. *Islamic Science and Engineering*. Edinburgh: Edinburgh University Press.

E. J. Holmyard. 1931. *Makers of Chemistry*. Oxford: Claredon Press.

Hayyan, J. I. https://en.wikipedia.org/wiki/Jabir_ibn_Hayyan

Holmyard, E. J. 1929. *The Great Chemists*. London. Methuen & Co. Ltd; 3rd Edition.

Holmyard, E. J. 1931. *Makers of chemistry*. The Clarendon press.

Holmyard, E. J. 1961. *Chemistry in Islam in Toward Modern Science*, edited by R. M. Palter edition. New York: Noonday Press, , vol. 1, pp. 160-70.

Levey, M. 1973. *Early Arabic Pharmacology*. Leiden: E. J. Brill.

Lindberg, D. C. 2007. *Islamic Science*. *The Beginnings of Western Science: The European Scientific Tradition in Philosophical, Religious, and Institutional Context, Prehistory to A.D. 1450*. Chicago: U of Chicago. pp. 163–92.

Mathé, J. 1980. *The Civilisation of Islam*, translated by David Macrae. New York: Crescent Books.

Meyerhof, M. 1931. Science and Medicine", in *The Legacy of Islam*, edited by Sir T. Arnold and A. Guillaume, Oxford: Oxford University Press, pp. 311-55.

Multhauf, R. P. 1993. *The Origins of Chemistry*. London: Gordon and Breach Science Publishers.

Principe, L. M. 2011. Alchemy Restored. *Isis* **102** (2): 305–12.

Sabra, A. I. 1996. *Situating Arabic Science: Locality versus Essence*. *Isis* **87** (4): 654–70.

KNOWLEDGE AND BIG DATA

Introduction

> If it exists, the Master Algorithm can derive all knowledge in the world – past, present and future – from data. Pedro Domingos *The Master Algorithm*, Penguin Books, 2015.

The meaning of the above sentence is not entirely clear. What exactly does "derive knowledge from data" mean? How is this to be understood and what does "all knowledge" refer to here? Does it refer to everything that can be learned, to all forms of cognition or only to knowledge proper – that is to true and justified beliefs, or again to only what we call science? Further, what is or what qualifies as an algorithm in this context? As we will see later on, the answer to that question is not clear; and that leads to uncertainty as to what it means to say that a system or algorithm "learns".

Leaving – for now – those difficulties aside, one thing is clear; this is a very ambitious, some would say outlandish, claim. Actually, there are two claims here. One claim is that "all knowledge" can be "derived" from data. Something which we may be tempted to describe as "radical empiricism", but whether or not that is the case depends on how we understand "data". The second claim is that there is a unique, or unitary, effective procedure – an algorithm – that could "derive" all knowledge from data in each and every field of knowledge. The important point is that this procedure is understood to be unique, or unitary, meaning that there is one and only procedure that applies across the board to all forms of knowledge. This implies that the division of science (or knowledge) into different domains, while it is not necessarily an illusion, is nonetheless a reflection of our failure until now to discover the "master algorithm". The effective procedure that constitutes the universal key, provided sufficient data. This, at first sight, may be viewed as "reductionism with a vengeance", but I will argue that this claim to universal competence is actually quite different. Because reductionism concerns inter-theoretic relations, while the importance of theories is precisely what is being questioned here.

We would be wrong to think that Pedro Domingos is some sensationalist journalist or a representative of a high-tech company trying to sale his ware to gullible clients. In fact. he is professor of computer science at the University of Washington and a highly respected specialist

of machine learning. Therefore, his claims should be taken seriously as identifying a project that is viewed among computer scientists as a legitimate scientific enterprise. The search for a mechanical procedure that can derive from any given data set all the knowledge that can be derived from it. Given the underlying assumption that *all knowledge* can be derived from data, the immediate conclusion is that the "master algorithm" should be able to derive from data all the knowledge there is, past, present and future.

The claim is that once the master algorithm is found, once we have in hand the universal key that can unlock any set of data whatsoever, the search for knowledge will be over. First because this effective procedure will be able to learn whatever can be learned from any future set of data, just as it can learn whatever can be learned from any set of data concerning the present or the past. Second, because it is the algorithm, rather than you or me that will "learn", that is to say, that will derive knowledge from data. This begins to answer our original question: what does 'derive knowledge from data' mean? It means for an algorithm to "learn". An algorithm derives knowledge from data when it can learn from the data.

Big Data

Big data is a rather ill-defined expression which can refer to different types of large scale data sets. Recently Big data, with a capital B, has been used to refers to extremely large data sets which contain trillions or even more data points. These data sets are characterized by what has been called the 3Vs: high volume, high velocity and high variety.

[1] High volume refers to the very large size of the data sets. High velocity means that big data sets are constantly changing, because new data is permanently being added at a very rapid rate. High variety corresponds to the range and diversity of data types and sources.

All three Vs relate to another characteristic of Big data which is viewed as fundamental by both its critics and advocates. Unlike ordinary statistics – either governmental, commercial or scientific – big data is not collected with a particular objective in mind, but arises as a by-product found in business and administrative systems, social networks, and the internet of things.[2] While a national census or business statistics are compiled in view of previously chosen goals and in response to specific questions, big data collects the results of what is described as "low

[1] I will adopt that convention and write "Big data" with a capital B to refer to extremely large data set characterized by the 3Vs and "big data" to refer to any large data set.

[2] See, *Big Data: Potential Challenge and Statistical Implications* C.L. Hammer, D.C. Kostroch, G. Quiros and STA internal group. IMF Staff Discussion Note, September 2017.

intentionality acts" -- for example, clicking on an add or viewing a web site even for a faction of a second[3]-- or arises as a result of normal administrative functioning – for example, the medical records of a hospital or the NHS. These data sets are permanently being upgraded because the data is collected in "real time" with the net traffic, instead of having to wait, say, five years for the next census; the variety of sources of the data is not pre-constrained by hypotheses which allows new domains to be explored and the data is cheap because it is a by-product of other activities. It is thus possible to compile extremely large data sets.[4]

The absence of directionality in the collection of its data is central to Big data's claim to "by-pass causality". Because the collection of data is not pre-formatted by specific hypotheses, surprising and unexpected correlations, it is argued, are discovered when sufficiently large data bases are explored by the appropriate algorithm, while ordinary statistics compile data in view of causal hypotheses concerning the interrelations between different variables. This claim about non-directionality provides additional insight into what "deriving knowledge from data" means. First deriving knowledge *from data* supposes that the data collected does not contain any prior hypothesis and that it is from the data itself that knowledge is derived without the help or guidance of imbedded hypotheses. Second deriving *knowledge* from data implies that knowledge *is* or corresponds to the correlations which appear.

However, critics are quick to point out that the "accidental"[5] nature of the data collection, the fact that it is a by-product of another activity, and absence of hypotheses guiding the collection do not rule out bias. Though the data is not compiled with any particular scientific hypotheses concerning the data in mind, it does not follow that it is compiled without any purpose, nor that it is without bias.

For if Big data is a by-product of business and administrative systems, of social networks and of the net traffic in general it is not an entirely accidental by-product. To a large extent this data concerns – and originally essentially concerned – the buying habits of internet users, their

[3] It is low intentionality data sources because it does not always involve a conscious intention and therefore these sources do not distinguish between clicking on an ad out of curiosity and sliding the mouse over it by accident. For big data, low intentionality it is not "by accident" but meaningful.

[4] The main reason why the data is cheap, and readily compiled is purely technological. The internet can automatically keep traces of all our low intentionality act that result in web traffic (and also some which don't but that demand special and intrusive software) and sending information to someone on the internet is not sending to someone something that you had and that you do not have anymore, now that you sent it, but simply making a copy of it.

[5] In the sense that the data is a by-product of another activity than collecting it.

reactions to advertisement, what they read online, what kind of sites they visit, etc. The data was compiled and used for commercial purposes, for marketing, especially advertisement. Data concerning patients' intake of various drugs and their reactions to these drugs came to be recorded as a by-product of hospitals and health services managements giving rise to bid data sets in medicine also. Again, even if this data arises as a by-product of normal administrative functioning, not all the data that so arises is of interest. So, the indirection of the collection and the variety of sources do not automatically rule out bias. To put it otherwise, absence of bias is not a by-product of the fact that Big data is a by-product.

In both these fields, commerce and medicine, the promise of Big data is, what may be called, narrowing the domain of inference. Statistics are classically understood to be governed by the laws of large numbers. In consequence statistics apply to groups or to class of people who share some common characteristics, but they are silent about individual reality. Big data, precisely because of the very large amount of data it assembles, pretends to be able to overcome this limitation by creating profiles or models, that can predict individual behavior. As Domingos writes: "As useful as averages are, we can do even better; indeed, the whole point of big data is to avoid thinking at such coarse level. Our clusters can be very specialized sets of people or even different aspects of the same person."[6]

In medicine, that would mean, for example, the ability to predict the reaction of a given person suffering from a uniquely defined type of cancer, to this or that particular drug. The objective is to be able to calibrate the prescription of drugs to the needs of individual patients in view of their unique genetic code and of the specific characteristics of their ailment. In commerce, ultimately it would mean elaborating on the basis of a person's past choices and "low intensity acts" an individual model of that person, that knows, or can "learn" his or her tastes and desires, and can predict, for example, where that person would love to go on holiday, the type of music he like, the books she is interested in reading (and buying), and so on. [7]

As the word "cluster" suggests Big data does not take us out of the world where the laws of large numbers apply. It is rather that Big data sets because of the three Vs allow us to represent

---

[6] Pablo Domingos, *The Master Algorithm* Penguin Books, 2015, p. 207.

[7] As P. Domingos puts it: "The company I am envisaging would do several things in return for a subscription fee. It would anonymize your online interactions, routing them through its servers and aggregating them with its other users'. It would store all the data from you in one place … It would learn a complete model of you and your world and continually update it. And it would use the model on your behalf, always doing exactly what you would do, to the best of the model's ability." *The Master Algorithm*, op. cit., p. 273.

an individual as a cluster of data points and to make hypotheses about his or her future behavior on the basis of past behavior. The knowledge derived from the data is nothing but these hypotheses, that is to say, nothing other than the probability distribution of the agent's future behavior.

In the case of individually calibrated drug prescription it may be objected that the knowledge will not simply be derived from data, but also, for example, from hypotheses concerning how patients who have this or that gene react to a drug having this or that molecule. This may be the case, but the big data specialist will respond that in the last analysis such hypotheses simply boil down to a lot of data, and that if it does not now, it will tomorrow when enough data is in. At that point, given the high velocity of big data we will be able to follow *in real time* the evolution of that correlation, and we will have no need for any hypothesis or theory concerning the relations between genes and molecules.

The implicit idea – in this albeit imagined answer – is that hypotheses, theories, knowledge as we know it, all are just "short hand" for data, or perhaps a better (Platonic) image is that they are "shadows" of data sets projected on the wall of the cave of data scarcity. Once we have enough data and the ability to treat it rapidly, theories and hypotheses will be superfluous. To put it in a language closer to that of analytic philosophy of science,Big data pretends to resolve the problem of the underdetermination of theory by evidence. Because when you have sufficient evidence you do not need a theory and because the problem of underdetermination is simply the problem of the lack of data.[8] Such is, I take it, one of, if not *the* central philosophical claim of big data.

Note though, that under one plausible description, at least in the medical case mentioned above, it may be claimed that not one iota of knowledge has been (directly) added to our understanding or explanation of cancer by the (future) prowess of big data. What individually calibrated drug prescription does, for better or worse, is to replace the "art" of the physician who after years of practice, knowing his patients and their disease has an, albeit obscure but often excellent, intuition as to what is the best most appropriate treatment for them. This is the

[8] The entry on the "Underdetermination of Scientific Theories" in the *Stanford Encyclopedia of Philosophy* begins with a very simple example: "if all I know is that you spent $10 on apples and oranges and that apples cost $1 while oranges cost $2, then I know that you did not buy six oranges, but I do not know whether you bought one orange and eight apples, two oranges and six apples, and so on." to which the data scientist will simply respond, that uncertainty only exists because you do not have all the data, if you have enough data the underdetermination disappears and you do not need a theory. The entry is by Kyle Stanford and may be found at https://plato.stanford.edu/entries/scientific-underdetermination/

competence which big data claims to be able to derive from data and to sharpen compared to the individual's intuition. Is this knowledge? Well clearly it is a form of knowledge, and to the extent that it proves to be successful it certainly is a remarkable achievement also, but it is a very specific form of knowledge.

It is more akin to "knowing how" than to "knowing what" and it could allow, us not only to tutor the "knowing how" of experienced physicians, but also to discover and systematize the "knowing that", the knowledge about diseases and patients which underlies it. Yet, discovering that it seems will require hypotheses and theories that are not in the data itself. That claim, however, is precisely what will be disputed by the data scientists and advocates of the master algorithm. Once you have the data of enough cases of individually calibrated drug prescription for this and that disease you will not need a theory to explain them, all the knowledge will be there in the data itself.[9]

Machine learning

Big data is extensively used in machine learning, among other things to bring neural networks to learn – either by themselves from the very beginning or after a period of human training – to recognize faces, or a picture of a cat, or that of a horse, or to drive a car, or to translate from one language to another, and so on. In other words, Big data provides the source, the raw material on which artificial systems that learn by-themselves are trained. However, the relation between Big data and machine learning runs deeper.

In fact, Big data require machine learning in two senses. First Big data can only be collected or compiled using artificial intelligent systems. Second without the help of machines to group the data, to mine it for information and to present it into forms that are manageable for humans, we would be unable to access not only Big data as such, but also data sets that are much smaller, but already way to large for humans. For example, already in 1992, the NASA cosmic background explorer satellite produced 160 million measurements per 26-week period.[10] Imagine, a mere 160 million measurements as opposed to the trillion of data points characteristic of big data, ridiculous! Yet, as Paul Humphreys writes,

[9] This could be viewed as the final stage of the dispute between "understanding" and "explanation", beyond the point where it is argued that explanation without understanding characterizes knowledge in the natural sciences, knowledge without explanation characterizes data sciences.

[10] In case you are interested that comes out to about 10 measurements per second, assuming that it is taking measurements continuously.

> For such data sets, the image of a human sitting in front of an instrument and conscientiously recording observations in propositional forms before comparing them with theoretical predictions is completely unrealistic. Technological enhancements of our native cognitive abilities are required to process this information and have become a routine part of scientific life. [11]

Artificial cognitive systems -- computers of course, but also gene sequencers, automatic telescopes, magnetic resonance imaging, satellites that take ten measurements per seconds, or instruments that record fluid data flow 1,000 times per second -- constitute the backbone of contemporary science and have been central to the progress of our knowledge for, at least, the last forty years.[12] Without machines to record it Big data would not exists, and even if it did, without artificial cognitive systems (often they are the same as the recording machines) to organize, to analyze and present us this data in an amenable form, we would not understand it.

These artificial cognitive systems have brought about an important transformation of the way in which science is now practiced and understood: human epistemic abilities have ceased to be the ultimate arbiter of scientific knowledge,[13] for at least two reasons. First, because we cannot reproduce the million of lines of calculation that a machine does to arrive at a numerical solution to a problem which does not have any analytic solution. We are left we a result which we can neither check, nor understand, mere data. The second reason is that many complex models, for example those of climate change, are so complex and include so many variables that no one can have an adequate representation of the model. Therefore, nobody knows exactly how they work and how they obtain the results they obtain. For example, we calibrate models of climate change by using them to make "predictions" concerning past climate change. Then we tweak this or that parameter until we obtain satisfactory results, but we do not always know why modifying this or that parameter brings about better or worse results.

Many of the tools which are indispensable to todays science and technology are thus epistemically opaque. As an analogy, one could say that such artificial cognitive systems are to contemporary science as modular systems are to general intelligence in Jerry Fodor's conception of the mind. They are "informationally encapsulated" we have only access to their results.[14] That analogy of course is not exactly correct. These tools are not complete black boxes, because we

---

[11] Paul Humphreys, *Extending Ourselves. Computational Science, Empiricism and Scientific Method*, Oxford University Press, 2004, p. 7-8.
[12] Ibid.
[13] Ibid., p. 53
[14] J. Fodor, *The Modularity of Mind*, MIT Press, 1983.

have made them using theories that we devised, understand, and tested. Nonetheless we cannot not entirely grasp how they work and how they arrive at the specific result at which they arrive. In any case we cannot reproduce that result manually, that is to say, without them. We depend on them and are not the final arbiter of what constitutes sound knowledge.

The claims of Big data and the search for a master algorithm should be viewed, I believe, in the context of this transformation of the practice science and changes in the conception of knowledge that it implies.

To learn

"If evolution can learn us," writes Pablo Domingos, "it can conceivably also learn everything that can be learned, provided we implement it on a powerful enough computer."[15] What does it mean to say that evolution can "learn" us? What does to "learn" someone mean? This rather curious grammatical use of "to learn" in revealing. Domingos considers that evolution is the ultimate example of how much a simple algorithm can achieve.[16] However, contrary to what he seems to assume, the algorithm here is not evolution, which rather corresponds to the results of the effective procedure, but natural selection which is the procedure that brings about evolution and through which the "data set" of all living creature is permanently being upgraded. Therefore, it would be more appropriate to say that natural selection can learn us. However, to say that natural selection can learn us, is simply to say that it can (or could)[17] produce us.

Similarly, a neural network is deemed to have learned when it can produce a certain type of result. For example, when it can "recognize" with a high level of success pictures that contain a horse. "Recognize" means to be able to separate pictures which contain horses from others in which no horse is present. Many of us would immediately respond, yes, but that does not mean that the system knows what a horse is or that it has a concept of a horse. This last assertion is disputable, at least in some cases. However, that I think is not really the issue.

---

[15] *The Master Algorithm*, p. 28-29.

[16] Ibid; see also chapter 5 p. 121 – 142.

[17] There is a real issue here that I cannot address. Given our present understanding of evolution, if natural selection could produce me or you, it is not clear that it can produce me or you again. Not only because we are genetically unique, but also because we consider that if we could replay the tape of evolution not only the ending would be different but also the whole story. See S. Gould *Wonderful Life*. It follows that to "learn" something, someone, here corresponds to a unique unrepeatable performance. However, the clearest sign that you know something, either knowing how or knowing that, is that you can do it again.

The data scientists, if she were consistent, and somewhat philosophically inclined, should respond, of course that this learning machine does not “know”, what it does is much better than knowledge. It can recognize patterns and discover regularities that knowledge cannot, because knowledge is constrained by theories, by causal hypotheses, by the need to understand, by representations without which it is not knowledge. For the last four hundred years, science, our best form of knowledge, has been mainly used to obtain results that we want. Machines that learn promise to be able to do that – to obtain results that we want – better than knowledge can.

Paul Dumouchel
Ritsumeikan University

What the unsupervised learning could deliver us (or, what not)

Insok Ko (INHA University, Korea; insok@inha.ac.kr)

The unsupervised learning, as a mode of machine learning, shall make a machine so clever that it recognizes, for instance, (the pictures of) cats among other sort of things. It is remarkable that there is no need to feed the machine explicit information about the category-specific properties of the cat in order to promote the machine to such level of cat-discerning competence. Though it is not known yet, how far this kind of competence would reach, it gives us certain hope for an *objective* classification, i.e. for one that is free from prejudices or cultural biases that infects us, human observers. Utilizing its power we might also be able to overcome the problem of theoryladenness of observation, whereby the Baconian ideal of observation shall be realized. In this talk I will present an evaluation of this prospect, investigating the process and structure of unsupervised learning.

1. What happens, when a machine learns a la mode 'unsupervised'?

In June 2012, Quoc V. Le, Andrew Ng and Jeff Dean et al. from Google published their accomplishment in the paper "Building High-level Features Using Large Scale Unsupervised Learning". Their AI program has learned to recognize the cats in the pictures without being taught how to do it. WIRED reported:

> "The "brain" simulation was exposed to 10 million randomly selected YouTube video thumbnails over the course of three days and, after being presented with a list of 20,000 different items, it began to recognize pictures of cats using a "deep learning" algorithm. This was despite being fed no information on distinguishing features that might help identify one."
> [1]

What is an unsupervised learning? Machine learning is called unsupervised, if "the agent learns patterns in the input even though no explicit feedback is supplied".[2] In a different formulation, a machine learning is classified to be unsupervised if no labeled data but only unlabeled data are used as inputs. The most common task of the unsupervised learning is

[1] Clark (2012).
[2] Russell&Norvig (2009), p.705.

clustering.[3]

> "We consider the problem of building high-level, class-specific feature detectors from only unlabeled data. For example, is it possible to learn a face detector using only unlabeled images? […] Contrary to what appears to be a widely-held intuition, our experimental results reveal that it is possible to train a face detector without having to label images as containing a face or not."[4]

Their accomplishment was that they succeeded (or, claimed to succeed) in *building a high-level detector* [of the category "cat" for example] *by feeding unlabeled data*, without teaching anything about cat. It was a remarkable event that the system built a face, front view, of cat that it found from the huge bulk of 10 million unlabeled thumbnails of YouTube videos. This sure is an impressive accomplishment, but we will not exaggerate its implication. Jeff Dean, a co-author, said in an interview in *New York Times*, "[The AI system] basically invented the concept of a cat."[5]

But, to be fair, his evaluation would be reformulated like this: "*The machine found a category of the images that come up with high frequency and are similar to each other, i.e. vary from each other within a narrow range of variation. Further, it extracted a representative image of this category.*" Thus reinterpreted, what the AI system of Google has performed does not amount to "inventing a concept", even less than "inventing a concept of cat".

Philosophical debate in last few decades made it clear that it is more than a tough challenge to render a machine intelligence what we call meaning of a concept, even if the machine can process symbols of the concept as well as we human beings do it. I agree with the pessimists on this evaluation. It is not strict impossibility, but shall be very hard to realize. To be brief with the reason, the meaning of those concepts we use in the real world is too complicated, with ever changing, also increasing dimensions of connotation.

Two examples (both from Korea). 1) Korean students call the (nationwide) entrance exam 'water exam', if they think it definitely easier than the usual exam of other years. I think that the concept of water got a new (applied) dimension to its meaning with this usage. 2) Everybody knows what a candle is. But the meaning of the term 'candle' got much richer with politically and emotionally loaded aspects that are added during last one year in Korea. These examples show that a concept with its meaning is a historical entity that emerges, grows and changes with time and in cultural-social context.

---

[3] Ibid.

[4] Le et al. (2012), Abstract. Underlined by Ko.

[5] Cited in: URL=http://www.dailytech.com/Googles+Unsupervised+SelfLearning+Neural+Network+Searches+For+Cat+Pics/article25025.htm.

Even when a machine discerns with 99.99% probability, thus practical certainty, the pictures that include A (image), it does not mean that it is justified to say that it has conceptual understanding of A. On the one hand, it can accomplish such task without having the concept of A in proper sense. On the other hand, it is not guaranteed that a cognitive system understanding the concept of A distinguishes without failure the image of the objects that belong to the category A. Let me show these (negative) relations by examples.

Even if you have ordinary concept of cats, it is still possible that you cannot accurately determine whether a beast you see in the video clip is a cat. It would be inappropriate to evaluate someone who looks in the picture at a rare breed of cat with unusual appearance, far from a typical cat and says "I am not sure. It may be a weird-looking weasel.", that she lacks the concept of cat or has a wrong cat-concept.

On the other hand, someone who identifies 100% of the autumn scenes in the video data possibly does not know that autumn is a season in which many people may feel lonely, or that it is season of harvest in the agricultural tradition, or that it is related to the tilt of the axis of rotation of the earth. It would be fair, if we say somebody has no proper concepts of autumn, if he does not know that it is one of the four seasons we have on the earth and it comes after summer.

2. Boundary of utility of the learning AI systems (AI systems are not learners that will grow to be independent cognitive agents like us. It shall not.)

Human beings are learners, as many other animals. They grow to be an independent cognitive agent that perceives its environment, processes the collected information, making judgments and decisions, and performs action thereby. In contrast, Every AI system is an artefact, how high its level of sophistication may be. It is a thing made by human agents for a specific purpose, conceived to perform certain surrogate function for human being in the techno-society that we are talking about these days.

Considering this context in which the AI systems are coming into being at all, learning is an essential process for AI to realize its *raison d'être*. The reason is as following:

1) They are expected to work as the human beings (would) do whom they shall surrogate in the situations they will have the role; and thus
2) They should be equipped with such (however restricted) cognitive propensity that coincides, or at least harmonizes with that of the human agents.

These are not purely theoretical but practical requirement having social and ethical implications.

3. k-means clustering: an example of unsupervised learning

Let us look into the k-means clustering, one of the basic types of unsupervised learning. It proceeds toward an optimized set of clusters into which the data are arranged. It is different from classification in ordinary sense in which there are fixed set of classes into which the data shall be arranged, in that the classes are made by the clustering process.

The procedure of k-means clustering starts with an assumption that there is a certain number of the clusters. In other words, the number of the clusters should be known, or simply assumed, in order that the whole clustering procedure would start at all.[6]

Let us look into the procedure of k-means clustering, known to be one of the most basic and also simplest type of unsupervised learning. Before we investigate the procedure, let us be aware that every clustering commences with the data that is represented by a set of points in an n-dimensional (property) space. [Q: What is/are data?] Each item of the data which is to be classified, or distributed into clusters, is represented by a point or a restricted n-dimensional region in this space.

> 1) Let the number of the clusters be given as K. Choose K distinct centers ("cluster centroids") for the whole distribution of the data points, one for each cluster. A centroid is a point in the n-dimensional space. (Those initially presumed cluster centroids may be collected arbitrarily.)
> 2) Let each point of the data belong to a certain cluster, so that the distance from this point to the assumed center of the cluster is minimized. Now we have K clusters with concrete data points!
> 3) Take the average point for each cluster. Let it be the new (modified/corrected) center of the cluster.
> 4) Measure anew the distance from each data to the new centers. Let each of the data belong to the cluster whose center lies nearest to it. (Re-clustering!)
> 5) Repeat the steps 3) and 4), until there is no more change of the clusters. Now we have K clusters stabilized.[7]

One should notice that the k-means clustering may only proceed with a presupposed number of clusters to be made. It means that either one should know the right number of the clusters, or one has to be satisfied with the more-or-less pretty outcome (clustering) of the trials. But it would be unfair to call this flaw of the method. It just constitutes an intrinsic feature of it. I will mention some more essential limitations of this sort of clustering methods

---

[6] Q: Can this number be modified as an effect of learning process? (→recursive determination of k?)
[7] Kim (2016), Chapter 6.

in section 5.

It might be interesting to look at the following cases of two hypothetical sets of data to be processed into clusters [Figure 1]. The k-means clustering with K=2 will produce the clustering (green/blue) like the left ones. In contrast, DBSCAN[8], being another clustering method and basic type of unsupervised learning, will deliver different clustering like on the right hand side. Do these imaginary cases suggest that the k-means clustering is deficient while DBSCAN's performance is a step better? No. We cannot make a judgment like that, until we have a golden standard by which the clustering is evaluated.

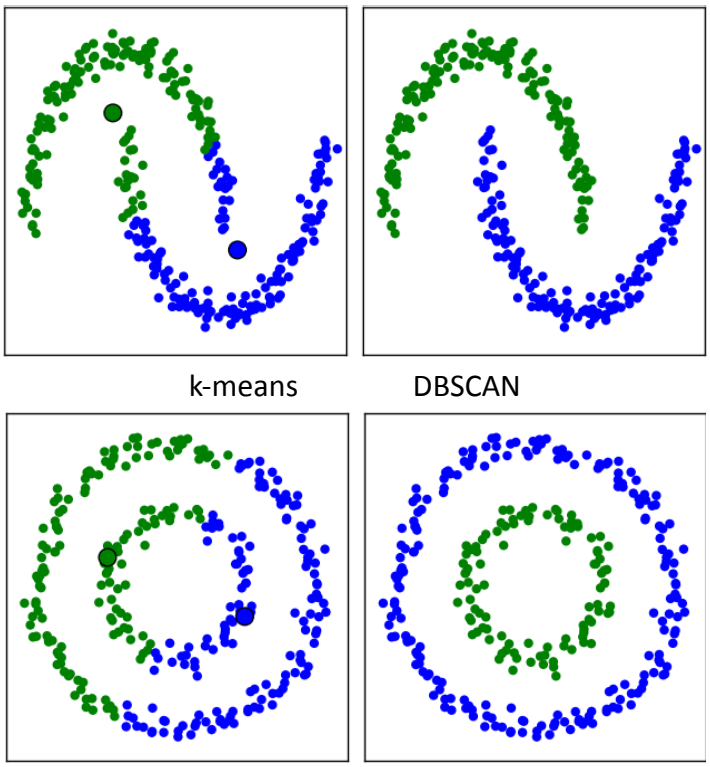


Figure 1. k-means vs. DBSCAN[9]

4. Value of the (unsupervised) machine learning

Nonetheless, the power of unsupervised learning ought not to be underestimated. The learning process using only the graphic images as data enabled the system to discern a certain specific kind of object. The system summarized the result of learning with a sort of average image of that object. This implies certain practical potential in such process of machine learning.

Cat is a familiar object for us. For a human agent, to recognize cat in various kinds of

[8] Density Based Spatial Clustering of Application with Noise.

[9] URL=http://commons.apache.org/proper/commons-math/userguide/ml.html.

pictures is no special achievement that deserves mentioning. But suppose that the machine finds a certain recurring pattern in the pictures of tissue of human organ malfunctioning in specific way, while no medical doctor or laboratory scientist has found it yet. It would be a special achievement.

Even if it is not that the artificial system discovered a new disease or new conception of the disease, it probably makes a decisive step toward the solution of the problem with that disease. In other words, it *helps us human agents* understanding better and dealing more effectively with the disease, while it is not correct to say that the artificial system itself has the concept of that disease.

This gives an example of the relation between machine intelligence and human intelligence. The former shall supplement but not substitute the role of the latter. The supplementary function of AI with learning mechanism will be especially manifest, where either of the following conditions is met:

> (1) The degree of agreement is high (on the matters at hand, including the way relevant concepts are used) AND there is a fairly well defined and fixed set of variables to be considered; or
> (2) The set of variables to be considered is well defined and practically fixed, BUT the degree of agreement is not sufficiently high.

For the situations of the category (1), supervised learning seems to be efficient. We shall be able to teach the machine, by some wisely selected set of examples, the way we see the things and let it mimic us. Those mimicking machines could apparently substitute human counterparts. But in such cases they are not genuinely replacing human agents, but only function as a form of extension of the relevant human agency, thus not independent from the latter.

I evaluate that the situation in medicine, like oncology, belong to the category (2). In this light-grey [or halfway transparent] area, the unsupervised learning and semi-supervised learning will help the professionals of the domain compensating their restricted and sometimes mutually incompatible view.

5. Some more fundamental limitations

The unsupervised learning performed by AI systems will not render a machine genuine conception, but it would enable *us* to look into the nature more accurately and efficiently. It would also help us diminish disagreements among ourselves in the inquiry of the nature. This can be seen as steps toward objectivity in our investigations of reality.

Now I will flip once more and conclude the talk by saying that there are more basic

elements restraining the hope that the unsupervised learning will get us nearer to "objective" clustering and objective classification. This hope says that we would approach realizing the idea of "cutting the nature at its joints" with it. But it is naïve idea.

First, machine learning takes place by way of some sorts of computation, whether in algorithmic way (Turing machine) or by a multi-layered network (deep learning). This process makes it necessary to choose the variables and to decide the distance metric of the space. Similarity relation is weighed and compared in the space made of those variables. Evaluation of similarity will vary with the metric of the space and the scale of the dimensions.[10] There is no nature-given set of variables and metric. It is us, the investigating minds, and not the super-powerful AI systems that make basic and essential decisions for the computation, maybe often unconsciously. Without intervention of such discriminating decisions, there is just "blooming, buzzing confusion" and no 'things' or their 'properties'.[11]

Some would reply that this problem is (at least partially) to overcome by considering ALL the (possible) variables. We have giga, tera, and exa bite computers with ever growing computation power. But this kind of approach misses the aspect of economy, essential in every context of reality. We do not afford consideration of all the possibly relevant variables, even if there *were* a set of such variables.

Second, it seems that joints in the nature are rare, if the term is understood to mean categories engraved in the nature itself. Those rare categories are biological species and chemical elements. All the other classes except those rare cases are man-made. It is simply unreasonable, if one expects that the machine intelligence working independently of human intervention will find such classes from the course of nature.

Reference

Clark, L. (2012), "Google's Artificial Brain Learns to Find Cat Videos", *WIRED*, URL=https://www.wired.com/2012/06/google-x-neural-network/.

James, W. (1890), *Principles of Psychology*, Henry Holt.

Kelleher, J. D., B. McNamee, A. D'Arcy (2015), *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*, MIT Press.

Kim, E. J. (2016), *Introduction to AI, Machine Learning and Deep Learning*, Wikibooks (in Korean).

Ko, I. (2017), "We human beings and AI in the Age of AI", Paper read in the KIAS Symposium on AI in the Era of BigData (in Korean).

[10] For instance, whether the scale of the axis is *w* or *log w* makes difference to such an effect.

[11] James (1890), p.462. He adequately formulates: "The noticing of any part whatever of our object is an act of discrimination."(Ibid.)

Le, Q. V., M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean, A. Y. Ng (2012), "Building high-level features using large scale unsupervised learning", *Proceedings of the 29th International Conference on Machine Learning*, Edinburgh, UK.

Russell, S. & P. Norvig (2009), *Artificial Intelligence: A Modern Approach* (3rd ed.), Pearson.

URL=http://commons.apache.org/proper/commons-math/userguide/ml.html

A taxonomy of experiments or modes of intervention?

Hsiao-Fan Yeh, PhD, Department of Philosophy, National Chung Cheng University, Taiwan.
Ruey-Lin Chen, Professor, Department of Philosophy, National Chung Cheng University, Taiwan.

Abstract

Carl F. Craver and Lindley Darden (2013) build up a taxonomy of experiments in biology. They distinguish loosely three categories of experiments and reclassify every category into several subkinds. We think that this taxonomy of experiments is so complicated as to raise some problems. We think that it is better to consider modes of intervention used in a single experiment or a series of experiments. Our goal in this paper is to argue that the three points: (1) Intervention is used as essential means to realize the testing and discovering functions of experimentation. (2) Two kinds of interventional directions (vertical or inter-level and horizontal or inter-stage) and two kinds of interventional effects (excitatory and inhibitory) are identified. (3) A series of related experiments may be realized testing and discovering functions by using multiple modes of interventions at one time. To illustrate our arguments, we take the synthesis of β-galactosidase in *E. coli* as a case.

1. Introduction

This paper explores the role of intervention and its different modes in experiments in the life sciences. Our goal in this paper is to argue that the three points: (1) Intervention is used as essential means to realize the testing and discovering functions of experimentation. (2) Two kinds of interventional directions (vertical or inter-level and horizontal or inter-stage) and two kinds of interventional effects (excitatory and inhibitory) are identified. (3) A series of related experiments may be realized testing and discovering functions by using multiple modes of interventions at one time. Our study is motivated and inspired by Craver and Darden's taxonomy of experiments in their 2013 work and Waters' view of experimentation as an investigative tool in his 2008 article. We begin by introducing the new mechanistic philosophy and Waters' view of investigation briefly.

Mechanistic explanations are often used for given phenomena in the life sciences. A influentially philosophical analysis of mechanistic explanations has been moved by some new mechanistic philosophers, especially, Lindley Darden and Carl F. Craver. Their research focuses on how scientists discover various mechanisms in the life sciences by using different reasoning strategies (Craver 2007; Craver and Darden 2013; Darden 2006; Machamer, Darden and Craver 2000). They also emphasize that
experimentation plays an important role in discovery of mechanisms. They argue that discovery of mechanisms occurs piecemeal via iterative refinements, which may be guided by interventional experiments. Scientists may manipulate some part of the mechanism and then observe what changes occur in a termination condition. The observed changes offer a useful guide for discerning the parts that are relevant to the behavior of a mechanism as a whole from the parts that are not. The role of interventions in experiments is a good tool to make a how-possible hypothesis become a how-actual description of a mechanism. Nevertheless, in most accounts from Craver and Darden, experimentation is largely be treated as an instrument for testing hypotheses about mechanisms.

C. Kenneth Waters questions the adequateness of the explanation-centered approach, including the mechanism-centered approach. He argues that scientists are more earnest in investigating new phenomena by means of explanatory theories instead of pursuing reductive explanations in biological practice. In his view, scientists use investigative practices by means of combining investigating new phenomena and explanatory reasoning to advance biological development. Philosophers would get wrong if they misplace the peripheral things in the center. (Waters 2008)

According to Waters, the investigation of new phenomena is a kernel action of scientific practice in the life science. However, it is important to distinguish the significant phenomena from the puzzling ones, because the outcome of investigating new phenomena may produce scientific discoveries or may not. We call those significant phenomena experimental discoveries, by following Chen's term (2013).

Chen argues that Mendel made an *experimental discovery*, which is independent of any existing theories at that time, via establishing data model from breeding experiments (Chen 2013). According Chen, experimental discovery comprises three steps: organizing raw data into significant phenomena, producing the need and motivation to discover mechanisms, and constraining the direction for construction of theoretical hypotheses. When scientists establish an adequate and correct data model representing a significant phenomenon, they will be motivated to search for underlying mechanisms or propose new hypotheses. (A detailed development, see Yeh and Chen 2017)

As we have seen in the literature discussed above, we find that different modes of interventions have not been clarified. Interventions in experiments are frequently used to test or discover something. Are there different modes of interventions that are effective in experimentation?

We begin in Section 2 with a brief characterization of Craver and Darden's of taxonomy of experimental types. Section 3 detects some defects of the taxonomy, and thus raises a need for a new approach. Section 4 identifies and discerns different modes of interventions in the experiments in the sciences of life. In the fifth section, we illustrate the arguments by the case study of the synthesis of β-galactosidase in *E. coli*.

2. The catergorization of experiments

Carl F. Craver and Lindley Darden jointly develop a mechanism-based and discovery-oriented new methodology of biological sciences and propose a highly systematic work on the discovery of mechanisms in life sciences, i.e., *In Search of Mechanisms* (Craver and Darden 2013). In this book, they explore the problems of how these experiments lead how-possibly mechanism schemas into how-actually mechanisms and provide a sophisticated account of how experiments work to discover mechanisms (see chapter 8). It not only exhibits the actual way by which scientists discover mechanisms but also recommends experimental strategies for those who want to investigate strange phenomena.

Craver and Darden first distinguish loosely three categories of experiments: those for testing causal relevance, those (interlevel experiments) for testing

componential relevance, and those (complex experiments) for asking specific mechanistic questions. The second category is reclassified into three subkinds based on the directions of intervention and detection, and based on the types of interventions and results. They are interference experiments that are bottom-up and inhibitory; stimulation experiments that are bottom-up and excitatory; and activation experiments that are top-down and excitatory. The third category is in turn classified into four subkinds: by-what-activity experiments, by-what-entity experiments, series of experiment with multiple interventions, and preparing the experimental system.

Craver and Darden emphasize that their goal "is not to offer a systematic taxonomy of experimental types but rather to call attention to the ways…, to answer specific questions about how a mechanism works." (Craver and Darden 2013:119) However, the above contents still give us a strong impression that they are making a taxonomic system of experiments. This impression is strengthened not only by using the term "kind" in the context but also by classifying kinds into subkinds. For example, they say that interlevel experiments have "the three most common kinds" (p.126) and consider that "some alternative kinds experiments" that do not fit the intervene-and detects structure (p.129). One problem we pose to Craver and Darden is that the taxonomy of experiments is a bit complicated and puzzling. The picture that provided by them seems to be fragmented rather than integrated.

The other problem is that Carver and Darden seem to pay more attention to the testing function of experiments in the process of mechanism discovery and to keep silence on the discovering function of experiments. But we think that the experiments that are used to discover novel phenomena are an important starting-point for discovering mechanisms. Gregor Mendel's hybridization experiment with peas discovered the segregation and the independent assortment of hereditary units and led to the discovery of Mendelian mechanism of heredity. Frederick Griffith's experiment with *Pneumococcus* discovered the transformation of bacteria cells and led to a series of discoveries of molecular mechanisms of heredity (Chen 2013). Molecular biologists, M. Hammarlund, E. Jorgensen, and M. Bastianis, learned the crucial guidelines from the unexpected phenomena in the experiment and lead to the discovery of the function of β-spectrin protein in neuron (Waters 2008). Those novel phenomena urged scientist to search for the underlying mechanisms. Craver and Darden pay less attention to the "discovering" function of experiments. If we deal with the complexity of the problem and supply the discovering part, then we may offer a more integrated and complete categorization of the roles of experiments that contribute to the discovery of mechanisms.

3. Intervention and experiment

As mentioned above, Craver and Darden reclassify the interlevel experiments into three subkinds: experiments are bottom-up with inhibitory intervention; bottom-up with excitatory intervention; top-down with excitatory intervention. Here we find an obvious asymmetry. One would wonder whether there are experiments with top-down and inhibitory interventions. In such a pattern of intervention, one intervenes the start conditions to inactivate or inhibit the explanandum phenomenon, and one detects the behavior of putative components of its mechanism to see if they change as a consequence. We think that vaccination is the very pattern. In a vaccination experiment, scientists inactivate the lethality of some kind of pathogenic bacteria or kill them, inject the avirulent or dead bacteria into subjects, and see if target organs of subjects no longer manifest relevant symptoms.

By rights, all "subkinds" of interlevel experiments share the intervene-and-detect structure, so there should be four patterns of interventions. Bottom-up and top-down are two *intervention directions* and they are not mutual exclusive, because one may exert different directions of intervention into the same mechanism. Excitation and inhibition are two *intervention effects* and they are not mutual exclusive either, because an intervention may produce both excitatory effect on one component and inhibitory effects on another component in the same mechanism. Furthermore, we think that this intervention framework can be adequately applied to the category of experiments for testing causal relevance.

Consider Julius Axelord's experiments Craver and Darden use to exemplify the category of series of experiments with multiple interventions. In performing the series of experiments, Axelrod and his colleagues intervened by injecting norepinephrine to increase rats' blood pressure, killing cats' nerves, injecting labeled norepinephrine, stimulating the live sympathetic nerves to release labeled norepinephrine from neurons, at different stages. These interventions are operated in both *top-down* direction on killing the nerves and *bottom-up* direction on stimulating the live nerves. And, in a fourth intervention, the scientists treated the nerves with cocaine and prevented the labeled transmitter from being re-sequestered. The effect appeared because cocaine could block the retake of neurotransmitters and enhance the effect of endogenous neurotransmitters. It produces both an *inhibitory* effect on the reuptake of neurotransmitters and an *excitatory* effect on endogenous neurotransmitters between two neuron synapses. This case shows that the framework of intervention directions may be combined with the framework of intervention effect to form a *scaffolding* of a more united framework.

Before we get into that, there is another thing to consider. Since that kind of pathogenic bacteria in the vaccination experiment is confirmed as the *etiological*

cause of the relevant symptoms and disease, the interlevel experiments for testing componential relevance can be also interpreted as experiments for testing causal relevance if we take an integrated view of causality and mechanisms. From this view, we see that there are at least five causal aspects of mechanisms (Chen, 2017:38-39):

(1) In the aspect of a mechanism consider as a whole, we understand a mechanism to constitute the complete cause of a phenomenon.
(2) In the aspect a mechanism piece comprising a part of the whole, we understand the piece to be a partial cause of a phenomenon.
(3) In the aspect of a stage in a mechanistic process, we understand each stage to be a part of causal chain, a causal net, or a causal mechanism.
(4) In the aspect of activity, we understand each activity occurring in a mechanism as a micro-cause, i.e., a cause of a micro-change.
(5) In the aspect of disturbance, we understand a disturbing factor or an activity exercised by a disturbing factoring to be a cause of the abnormal output or the malfunction of the disturbed mechanism.

For the purpose of this paper, we would like to especially focus on (3). Because, in biological practices, the scientists often need to design and operate the interventional experiments involved embryology and genetics. In order to understand how cell with identical genomes may develop differentiated and transmit particular characters to the offspring, they need to produce various mutations as interventional tools and breed them. That's why we need to take the temporal and etiological factor into our account. We attempt to extend the interventional aspect of the causal relevance from different levels to different stages. Thus, Craver and Darden's distinction between "experiments for testing causal relevance" and "interval experiments for test componential relevance" might not be necessary.

Craver and Darden's great contribution to the taxonomy of experiments has offer a basis for us to discuss different patterns of intervention. We are also quite agreeable with their conclusion: "In particular it tells us that different kinds of questions about a mechanism are answered with different kinds of experimental strategies. It tells us the conservation is often protracted, involving multiple interventions and series of experiments." (Craver and Darden 2013:142) On the basis that they have paved, we want to develop a more united and complete framework to account for a plurality of intervention in the life sciences.

4. A variety modes of interventions

In this section, we propose a new framework for analyzing a variety of interventions in the discovery of new mechanism. In this framework, we do not categorize different kinds of experiments, nor distinguish among abstraction types. Instead, we want to analyze different modes of intervention that can be used to realize the testing and discovering functions of experimentation. First of all, it will be helpful for us to clarify the two functions of experimentation: testing and discovering.

Under the testing function, we may categorize different kinds of experimental testing based on different targets. For examples, we have experiments for testing a causal hypothesis, experiments for testing a mechanism model (schema), and experiments for testing a putative part (an entity or an activity) or a putative stage of a mechanism model. However, we may have experiments or series of experiments for testing a causal hypothesis related to the mechanism model, the whole of the mechanism model, all of putative parts or stages of a mechanism model. Therefore, this categorization is really a taxonomy of testing targets rather than a taxonomy of experiments. Different targets may be the common goal of one and the same experiment.

We may similarly categorize different kinds of experimental discoveries based on different discovered objects. For example, we may have experiments that discover a significant phenomenon, experiments that discover a new entity, experiments that discovery a kind of new activity, and experiments that discover a new mechanism. Again, we may have experiments or series of experiments that can discover all objects as the previous sentences say. Therefore, this categorization is not a taxonomy of experiments, either. It is a taxonomy of experimental discoveries. A single experiment or a single set of experiments may discover different objects.

Considering the experimental testing and discovering together, we find that even a single experiment or a single set of experiments may perform both testing and discovering functions, as we clearly see Craver and Darden's fruitful discussion about the goal of discovering an underlying mechanism via experimentally testing the causal relevance or componential relevance of an entity or an activities to the target mechanism.

Interventions are used as essential means to realize the two typical functions. As we have argued in section 3, we may discern two kinds of interventions based on different interventional directions: vertical and horizontal. The intervention in the vertical direction occurs between two levels, say neuroscientists may put the rat in a maze and record the electrical activity of neurons in the rat hippocampus or molecular biologists may intervene one nucleotide sequence that codes some genetic information in organisms and observe the changes in the behavior of a mechanism as a whole. It can be also divided into the "up-down" and the "bottom-up" directions. So we call it

“inter-level” intervention, too. The intervention in the horizontal directioin occurs in different stages in a mechanistic process, say, biologists may engineer a part of a mechanism at an earlier stage and investigate changes at a later stage. Particularly the regulatory mechanisms have different working entities serially operating at different times in an extended process. So we call it “inter-stage” intervention, too. The inter-level intervention can be used to test and discover any putative part of some mechanism model while the inter-stage intervention can be used to test and discover any putative stage of some mechanistic process.

To determine the feature of an entity or activity involved in the mechanism, scientists attempt to make some difference at the inputs and see whether such an intervention brings about some corresponding change at the outputs. There are two kinds of consequences: excitatory effect and inhibitory effect. We call the consequences those are tending to activate, excite, stimuli the original states are “excitatory effects”. Alternatively, those are tending to eliminate, disable, shut-down, inhibit the original states are “inhibitory effects”. In principle, we may have the four modes of intervention: an inter-level and excitatory, an inter-level and inhibitory, an inter-stage excitatory, and an inter-stage inhibitory intervention. In the first case one intervenes on the start level to enhance or activate and observe the reactions on another level, say injecting radiolabeled norepinephrine or stimulating the live nerves. In the second case one intervenes on the start level to weaken or inhibit and detect the reactions on another level, say removing neurons or making a mutation on some sequence of genes. In the third case one intervenes to trigger or activate at the upstream stage and see the changes at the downstream stage, say adding inducers into mechanistic environments. In the fourth case one intervenes to shut down or inactivate or shut down a part in the upstream stage and assess the changes at the downstream stage, say knock out functional protein from a system.

All experimental functions, interventional directions and effects are not mutual exclusive in experiments. That is, they all may occur in one and the same experiment. A series of experimental interventions may perform two kinds of experimental functions, use two directions of intervention, and acquire two kinds of interventional effects. Because one can exert an intervention into some mechanism and produce excitatory effects on one component and inhibitory effects on another in the same mechanism, depending on whether the feature of entity is essentially excitatory or inhibitory. Interventions may produce novel or unexpected phenomena so as to test a mechanism model and discover the whole mechanism. All these modes of intervention in experimentation make important discoveries in biology.

5. The interventional experiment in discovering of the synthesis of β-galactosidas

(1) Experimental background

Consider the famous PaJaMa bacterial mating experiment[1]. In order to investigate a puzzling phenomenon of enzyme induction, Arthur Pardee, Francois Jacob, and Jacque Monod performed a series of interventional experiments. Monod characterized the phenomenon by various inducing experiments first (Monod 1947). He found that when the bacteria were grown in a single carbohydrate as carbon and energy source (say, glucose), the amount of the growth of bacteria was strictly proportional to the concentration of carbohydrate (see Fig. 4, Monod 1947:249). Alternatively, when the bacteria were grown in *two* carbohydrates (say, glucose and lactose), the bacterial growth-curves exhibited two successive complete growth cycles and separated by a period of lag (Fig. 5, Monod 1947:249). He called that the phenomenon of "diauxie" (Monod 1947:249).

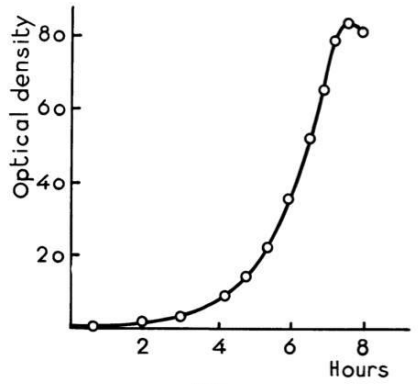


FIG. 4
GROWTH OF *B. subtilis* IN SYNTHETIC MEDIUM WITH SACCHAROSE + D-MANNOSE AS CARBON SOURCE; NORMAL GROWTH CURVE (82).

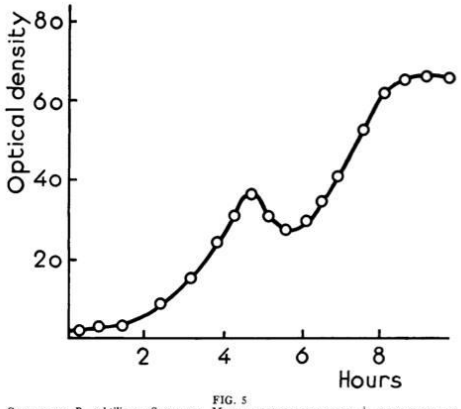


FIG. 5
GROWTH OF *B. subtilis* IN SYNTHETIC MEDIUM WITH D-FRUCTOSE + L-ARABINOSE AS CARBON SOURCE, "DIAUXIC" CURVE (82).

Now we know that there is a gene regulatory mechanism underlying the phenomenon. When glucose is consumed, lactose being as an inducer induces the gen to produce a particular enzyme, i.e., β-galactosidase, that breaking down lactose for energy. In the 1930s and 40s, biologists suggested the "enzyme adaptation" model for this phenomenon. The enzyme adaptation model may be simply described as that when bacteria were given two kinds of carbohydrates (so-called "substrates"), they will synthesize one kind of enzymes to digest glucose first, which is the long-known

[1] Schaffner notes that the "PaJaMa" is a combination of the first two letters of each of the scientists' names: Pardee, Jacob, and Monod. "PaJaMa" is more appropriate than "PaJaMo" because that is mating experiments. But nobody knows who exactly coined the abbreviation (Schaffner 1974:361, footnote 42).

"glucose-effect" (Monod 1956:21), whereby the presence of glucose (and other carbohydrate) generally inhibits enzyme synthesis. When glucose is empty, bacteria will develop the ability to synthesize another enzymes to digest lactose. Due to bacteria need moments to *adapt* another carbohydrate then synthesis a specific enzyme, they called the enzymes "adaptive enzymes". Yet just as Kenneth Schaffner said:

> "…the term 'adaptation' was felt to be somewhat confusing, …and was also thought by some to have teleological implications, and in 1953 the phenomenon was rechristened 'enzyme induction'." (Schanffner 1974:351)

In the enzyme induction model, the enzyme-forming system consists of the fundamental production mechanisms of *constitutive* (continuous produced) enzymes and *induced* (production can be turned on and off) enzymes. Monod's group previously discovered that the z gene is responsible for producing the induced enzymes, i.e., β-galactosidase, and the y gene is responsible for producing β-galactosidase permease. At the same, they also discovered that the z gene is closely linked with the y gene and the i gene (which represents inducibility or the inducible gene) has a specific effect on both of them (Hogness and Monod 1955; Cohen and Monod 1957).

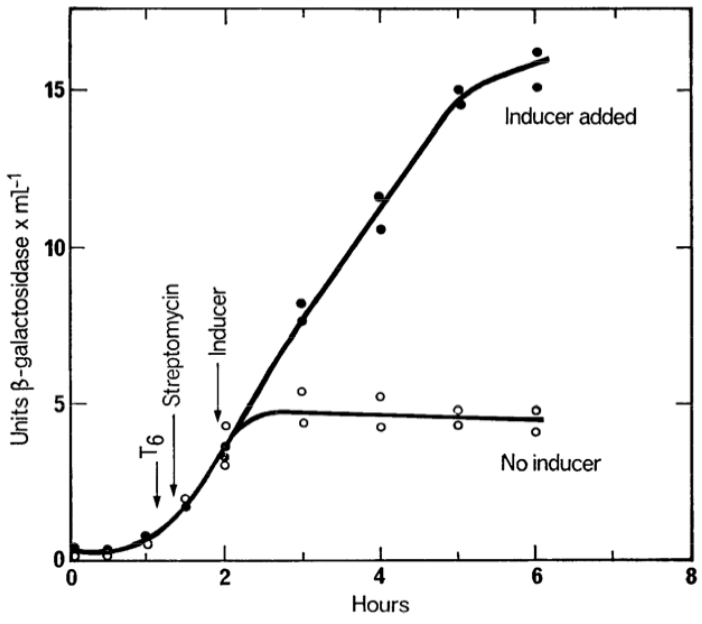
The PaJaMa research group prepared various important interventional tools. First, there were specific *mutant* strains of bacteria that had changed from being inducible to being constitutive. These mutants were used for comparison of constitutive and inducible enzymes and the cause of their production. They used i+ for the gene determining the inducible feature (bacteria require external inducers to induce the enzymes), i- for the gene determining the constitutive feature (i- gene itself produces internal or endogenous inducers to continuously induce the enzymes; hereafter "the internal model"), z+ for the normal (or wild type) gene producing β-galactosidase, and z- for the mutant (nonfunctional) gene. Secondly, the research group used *gratuitous inducer* as the interventional tool that is not hydrolyzed by the β-galactosidase nor metabolized by the bacteria)[2]. This showed the research groups were able to decide the synthesis of the enzymes to be initiated or stopped in terms of putting gratuitous inducers in or move them out of the experimental systems at will. Thirdly, the scientists conducted *interrupted mating experiments* in which ones were able to determine the degrees of the chromosomal transfer from male bacteria into female bacteria at will as well. To put it somewhat metaphorically, these

[2] Methyl-β-D-thiogalactoside (MTG) was used in Hogness, Cohn, and Monod 1955; Isopropyl-thio-β-D-galactoside (IPTG) was used in PaJaMa 1959.

interventional tools function as switchgears for turning machines (enzyme synthesis) on and off and as adjust valves for modulating its operation (bacterial conjugation) in the related experiments. At last, the researchers put *radioactive isotope* of sulfur (35S) into β-galactosidase as labeling markers for tracking the complex interaction among different genotypes, different enzymes, and different carbohydrates.

(2) The PaJaMa experiment: the internal inducer model or the repressor model?

Pardee, Jacob and Monod performed a series of experiments that conjugating inducible and constitutive bacterial strains for discovering the synthesis mechanism. They feed male's z-i- genes into female's z+i+ genes in the absence of inducers and then feed male's z+i+ genes into the female's z-i- genes in the absence and in the presence of inducers. According to the internal inducer model, when male's i- gene enters female's cytoplasm to meet z+i+ genes, the zygote (i-/z+) should form the presumed internal inducers then induce the normal z+ gene to begin synthesizing the enzymes even in the absence of external inducer. But the result was that *no* enzyme was synthesized. In the second experiment, when male's z+ gene enters the female's cytoplasm to meet z-i- genes, the zygote (z+/i-) should produce internal inducers then will begin to synthesis and continues without stopping even in the absence of external inducer. But the result was that the synthesis began at a normal rate about few minutes later, then *stopped* after about two hours. Yet they had the interesting discovery was that the synthesis would be restarted if external inducers were added (Fig. 4, PaJaMa 1959:173).



According to the internal inducer model, in the first experiment, the behavior of zygotes (z-i-/z+i+) should change from inducible to constitutive because z- injects first and meet i+, but the fact (no enzyme is synthesized) "means that the constitutive (i-) allele from male is never expressed. This suggests that the dominant allele is the inducible (i+)." (PaJaMa 1959:174) To test this hypothesis, they conducted the second experiment. The behavior of zygotes (z+i+/z-i-) should be constitutive because z+ injects first and meets i-, but the fact (enzyme synthesis produced but stopped after about two hours) showed that the internal inducer model was wrong. At this stage, the research has experimental tested the existing model was disproved by facts and experimental discovered that i+ gene's inducibility is dominant over i- gene's constitutiveity.

The main thing here is that when male's z+ and i+ both enter into female's cytoplasm at the same time, whether z+/i- constitutiveity conversion to i+ inducibility? There are two opposite effects in the zygotes. The scientists found that the presumed i- gene's constitutiveity effect just sustained in a short while. From the observation they concluded that "the constituitive (i-) allele in inactive, while the i+ is dominant, provoking the synthesis of a substance responsible specifically for the inducible behavior of the galactosidase enzyme-forming-center." (PaJaMa 1959:175)

In Monod's Nobel lecture, he had some important comments concerning the substitute "the repressor model" for the internal inducer model. The i gene determines the synthesis, not of an inducer, but a "repressor" which *blocks* the synthesis.

> "Of course I learned, like any schoolboy, that two negatives are equivalent to a positive statement, …, [we] detected this logical possibility that we called the ' theory of double bluff,'… I had always hope that the regulation of 'constitutive' and inducible systems would be explained one day by a similar mechanism. Why not suppose, …, that induction could be effected by an anti-repressor rather than by repression by an anti-inducer? " (Monod [1965]1977:479)

We can therefore say that the research group has experimentally discovered the repressor model that the enzyme synthesis was regulated by the interaction between i gene and inducer (say, lactose) in a negative feedback manner. The i+ gene produces a repressor, which inhibits enzyme synthesis by blocking the upstream region in the gene. When lactose exists in the cell, it binds to the repressor, which then allows synthesis to begin. By contrast, the mutant i- gene does not produce a repressor, so synthesis occurs constitutively in the presence of lactose. The repressor model seemed to be supported by the two hours delay in the second experiment. It can predict and

explain the phenomenon that the synthesis began about few minutes after z+ enters the i- cytoplasmic environment (mutant i- gene does not produce a repressor for inhibiting the synthesis at this stage, say constitutive) and that the synthesis restarted if an external inducer was added (functional i+ enters the cytoplasmic environment, thus the phenotype changed constitutive to inducible).

(3) Modes of interventions

In the series of experiments, the scientists use multiple interventions as essential means to realize the two functions of experiments: testing the internal inducer model as inappropriate and discovering the repressor model. They carry out inter-stage mode of interventions (interrupted mating experiments) several times and bring about the inhibitory effects (terminating the synthesis by injection i+ gene into female's cytoplasm; turning gene functional into non-functional; turning constitutive into inducible) and excitatory effects (initiating the synthesis by adding external inducers; turning inducible into constitutive). Finally, the scientists discover that the subtle regulatory mechanism underlying the phenomena of "diauxie".

6. Concluding remarks

Experiments always occupy a central status in scientific practices. Francis Bacon metaphorically expressed that nature has to be intervened so as to reveal her secret. Although Craver and Darden believe that that metaphor is no longer so appealing, they still write: "…the idea that we learn from nature by manipulating it and detecting the consequences of these interventions remains fundamentally correct." (Craver and Darden 2013:141). Inspired by Craver and Darden's illuminating insights, we propose an advanced framework of biological experiments that considers interventional functions, modes and effects. There is much yet to be discussed in our framework, for example, the labeling intervention, the negative control intervention and other kinds. They are left to us to continue the work via more investigations.

References

Chen, Ruey-Lin (2013). Experimental Discovery, Data Models, and Mechanisms in Biology: An Example from Mendel's Work. In Chao, Hsiang-Ke, Szu-Ting Chen, and Roberta L. Millstein (eds.). *Mechanism and Causality in Biology and Economics*. Dordrecht: Springer Press.

Chen, Ruey-Lin (2017). Mechanisms, Capacities, and Nomological Machines:

Integrating Cartwright's account of nomological machines and Machamer, Darden and Craver's account of mechanisms. In Chao, Hsiang-Ke, Szu-Ting Chen, and Julian Reiss (eds.). *Philosophy of Science in Practice: Nancy Cartwright and the Nature of Scientific Reasoning*. New York: Springer Press.

Cohen, G. N. and Jacques Monod (1957). "Bacterial Permeases." In Andre Lwoff and Agnes Ullmann (eds.) *Selected Papers in Molecular Biology by Jacques Monod*. Academic Press Inc.

Craver, Carl F. (2001). "Discovering Mechanisms in Neurobiology: The case of Spatial Memory." In Machamer, Peter, R. Grush, and P. McLaughlin (eds.) *Theory and Method in the Neuroscience*. Pittsburgh: University of Pittsburgh Press. Reprinted in Darden 2006, ch. 2.

Craver, Carl F. (2002). Interlevel Experiments, Multilevel Mechanisms in the Neuroscience of Memory. *Philosophy of Science* (Supplement) 69:S83-S97.

Craver, Carl F. (2005). "Introduction: Mechanisms Then and Now." In Craver, Carl F. and Lindley Darden (eds.) *Studies in History and Philosophy of Biological and Biomedical Science*. Special Issue, "Mechanisms in Biology" 36:233-244.

Craver, Carl F. (2007). *Explaining the Brain: Mechanisms and the Mosaic Unity of Neuroscience*. Oxford: Clarendon Press.

Craver, Carl F. (2008). Axelrod, Julius In Noretta Koertge (ed.) *New Dictionary of Scientific Biography*. Detroit, MI: Charles Scribner's Sons/Thomson Gale.

Craver, Carl F. and Lindley Darden (2013). *In Search of Mechanisms: Discoveries across the Life Sciences*. Chicago: The University of Chicago Press.

Darden, Lindley (1991). *Theory Change in Science: Strategies from Mendelian Genetics*. Oxford: Oxford University Press.

Darden, Lindley (2006). *Reasoning in Biological Discoveries: Essay on Mechanisms, Interfield Relations, and Anomaly Resolution*. Cambridge: Cambridge University Press.

Darden, Lindley and Carl.F. Craver (2002). "Strategies in the Interfield Discovery of the Mechanism of Protein Synthesis." *Studies in History and Philosophy of Biological and Biomedical Sciences* 33:1-28. Corrected and reprinted in Darden 2006, ch. 3.

Hogness, David S., Melvin Cohn and Jacques Monod (1955). "Studies on the Induced Synthesis of β-galactosidase in Escherichia Coli: The Kinetics and Mechanims of Sulfur Incorporation." In Andre Lwoff and Agnes Ullmann (eds.) *Selected Papers in Molecular Biology by Jacques Monod*. Academic Press Inc.

Judson, Horace F. (1996). *The Eighth Day of Creation: The Makers of the Revolution in Biology*. Expanded Edition. Cold Spring Harbor., NY: Cold Spring Harbor Laboratory Press.Monod, Jacques (1947). "The Phenomenon of Enzymatic and

its Bearing on Problem of Genetics and Cellular Differentiation." In Andre Lwoff and Agnes Ullmann (eds.) *Selected Papers in Molecular Biology by Jacques Monod*. Academic Press Inc.

Monod, Jacques (1956). "Remarks on the Mechanism of Enzyme Induction." In Andre Lwoff and Agnes Ullmann (eds.) *Selected Papers in Molecular Biology by Jacques Monod*. Academic Press Inc.

Monod, Jacques ([1965] 1977). "From Enzymatic Adaptation to Allosteric Transition." Reprinted in *Nobel Lectures in Molecular Biology: 1933-1975*, pp. 259-82. New York: Elsevier.

Pardee, Arthur B., François Jacob and Jacques Monod (1959). "The Genetic Control and Cytoplasmic Expression of 'Inducibility' in the Synthesis of β-galactosidase by E. Coli." In Andre Lwoff and Agnes Ullmann (eds.) *Selected Papers in Molecular Biology by Jacques Monod*. Academic Press Inc.

Pardee, Arthur (1979). "The PaJaMa Experiment." In Lwoff, Andre and Agnes Ullmann (eds.) *Origins of Molecular Biology: A Tribute to Jacque Monod*, pp. 109-116. New York: Academic Press.

Schaffner, Kenneth (1974). Logical of Discovery and Justification in Regulatory Genetics. *Studies in History and Philosophy of Science* 4:349-385.

Waters, C. Kenneth (2008). Beyond Theoretical Reduction and Layer-Cake Antireduction: How DNA Retooled Genetics and Transformed Biological Practice. In Michael Ruse (ed.) *The Oxford Handbook of Philosophy of Biology*. Oxford: Oxford University Press.

Woodward, James (2003). *Making Thing Happen: A Theory of Causal Explanation*. Oxford: Oxford University Press.

Yeh, Hsaio-Fan and Ruey-Lin Chen (2017). An Experiment-Based Methodology for Classical Genetics and Molecular Biology. *Annals of the Japan Association for Philosophy of Science* 26:39-60.

Staying Regular?

Alan Hájek

ALI G: *So what is the chances that me will eventually die?*
C. EVERETT KOOP: *That you will die? – 100%. I can guarantee that 100%: you will die.*
ALI G: *You is being a bit of a pessimist…*

–Ali G, interviewing the Surgeon General, C. Everett Koop

Autobiographical back story

- Over my philosophical career I've been interested in various topics, but certain topics have especially gripped me…

Introduction

- I'll discuss the fluctuating fortunes of *regularity:*

If X is possible, then the probability of X is positive.

$\Diamond X \rightarrow P(X) > 0.$

Introduction

- I'll give many reasons to care about regularity.
- So it's important to formulate it carefully.
- I'll offer what I take to be its most plausible version: a constraint that bridges *doxastic* modality and *doxastic* (subjective) probability.
- But even that will fail.

Introduction

- There will be two different ways to violate regularity
  - zero probabilities
  - no probabilities at all (probability gaps).
- Both ways create trouble for pillars of Bayesian orthodoxy:
  - the ratio formula for conditional probability
  - conditionalization, characterized with that formula
  - the multiplication formula for independence
  - expected utility theory

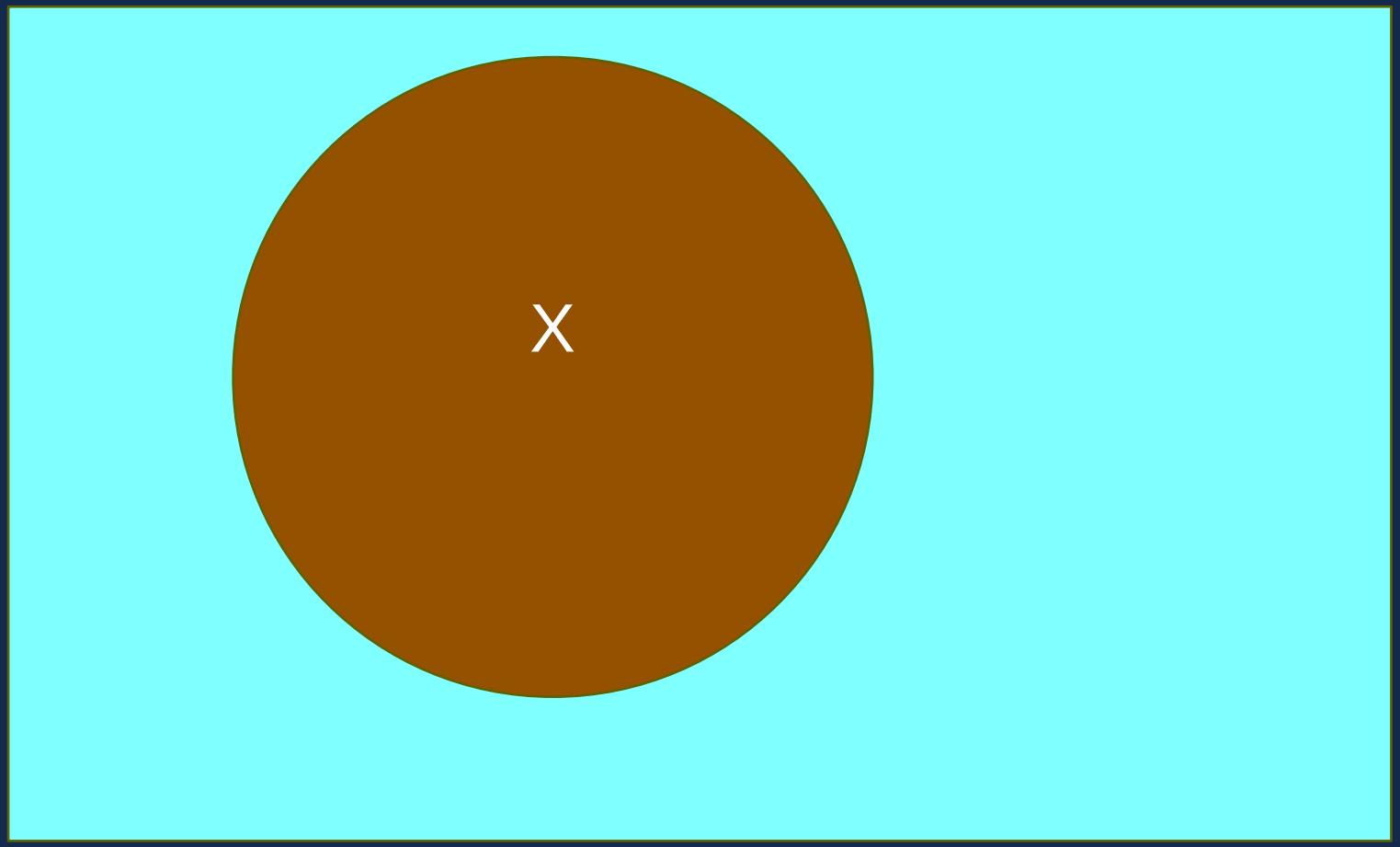
Introduction

- The failure of this seemingly innocuous constraint has ramifications that strike at the heart of probability theory and formal epistemology.

Regularity

If X is possible, then the probability of X is positive.

- Muddy Venn diagram: no bald spots.



Regularity

- An unmnemonic name, but a commonsensical idea.
- “If it can happen, then it has a chance of happening”…

Advocates of regularity

- Regularity has been suggested or advocated by Jeffreys, Jeffrey, Carnap, Shimony, Kemeny, Edwards, Lindman, Savage, Stalnaker, Lewis, Skyrms, Appiah, Jackson, Hofweber, …

Reasons to care about regularity

- Why care about regularity? …

Reasons to care about regularity

- Regularity promises a reduction of modality to probability: X is possible iff X has positive probability.

Reasons to care about regularity

- Regularity promises a bridge between probability and truth:

  If X has probability 0, then X is impossible, hence (actually) false.

  If X has probability 1, then X is necessary,

  hence (actually) true.

- If regularity fails, even this is a bridge too far!

Reasons to care about regularity

- Regularity may provide a bridge between traditional epistemology and Bayesian epistemology.

Reasons to care about regularity

- Regularity promises to illuminate *rationality*.
- It would provide a much-needed additional constraint on rational credence that goes beyond coherence.

Reasons to care about regularity

- Centrepieces of synchronic Bayesian epistemology face problems when regularity fails.

Reasons to care about regularity

- The centrepiece of *diachronic* Bayesian epistemology – conditionalisation – faces problems without a version of regularity; yet it also conflicts with regularity.

Reasons to care about regularity

- Bayesian decision theory faces problems if regularity fails.
- So failures of regularity pose some of the most important problems for probability theory as a representation of uncertainty.

Reasons to care about regularity

- These failures motivate other representations of uncertainty – Popper functions, ranking functions, NAP, comparative probabilities…

Reasons to care about regularity

- So the stakes are high!

Formulating regularity

If X is possible, then the probability of X is positive.

- This is just a schema.
- There are many senses of ‘possible’ in the antecedent…
- There are also many senses of ‘probability’ in the consequent…

Formulating regularity

- Pair them up, and we get many, many regularity conditions.
- Some are interesting, and some are not; some are plausible, and some are not.
- Focus on a pairing that is definitely interesting, and somewhat plausible, at least initially.

Formulating regularity

- In the consequent, let's restrict our attention to *rational subjective* probabilities.
- If X is possible, $C(X) > 0$.
- In the antecedent? ...

Formulating regularity

- *Doxastic* possibility seems to be a promising candidate for pairing with subjective probability.
- Doxastic regularity:

  If X is doxastically possible then C(X) > 0.

Formulating regularity

- We might think of a doxastic possibility for an agent as:
  - something that is compatible with what she *believes*;
  - or something that she is not certain is false;
  - or perhaps some other understanding …
  - I will speak of a doxastically *live* possibility—for short, a *live* possibility.

Formulating regularity

- So from now on I will understand regularity as:

  if X is a live possibility then C(X) > 0

- All the better that this can be understood in multiple ways. For I believe that on any reasonable understanding of 'live possibility', it is false.

Formulating regularity

- If doxastic regularity is violated, then offhand two different attitudes are conflated: to genuine impossibilities, and to some improbable possibilities.

Formulating regularity

- And yet doxastic regularity appears to be untenable.

Two ways to be irregular

- There are two ways in which an agent's probability function could fail to be regular:

  1) It assigns zero to some live possibility.

  2) It fails to assign *anything* to a live possibility.

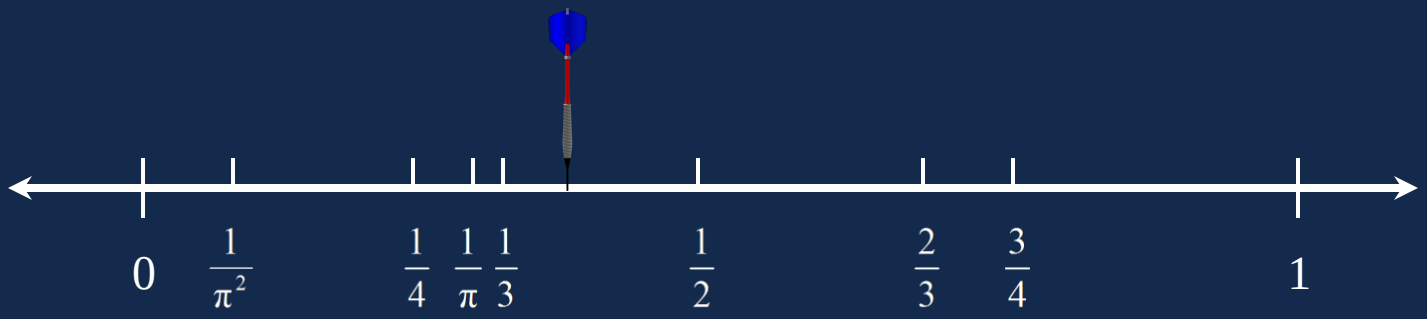
Two ways to be irregular

- Those who regard regularity as a norm of rationality must insist that all instances of 1) *and* all instances of 2) are violations of rationality.
- I will argue that there are rational instances of both 1) and 2).

Dart example

Throw a dart at random at the [0, 1] interval of the reals …

Dart example



Dart example

- Various non-empty subsets get assigned probability 0:
  - All the singletons
  - Indeed, all the finite subsets
  - Indeed, all the countable subsets
  - Even various uncountable subsets (e.g. Cantor's 'ternary set')

Dart example

- Examples like this pose a threat to regularity as a norm of rationality.
- Any landing point in [0, 1] is a live possibility for our ideal agent.

Arguments against regularity

- In order for a probability function *P* to be regular, there has to be a certain harmony between the *cardinalities* of *P*'s sample space and its range.
- If the sample space is too large relative to *P*, regularity will be violated.

Arguments against regularity

Kolmogorov's axiomatization requires *P* to be *real*-valued. This means that any uncountable probability space is automatically irregular. (Hájek 2003).

Arguments against regularity

- It is curious that this axiomatization is restrictive on the *range* of all probability functions: the real numbers in [0,1], and not a richer set;
- yet it is almost completely permissive about their *domains:* Ω (the sample space) can be any set you like, however large, and *F* (the set of subsets that get assigned probabilities) can be any field on Ω, however large.

Arguments against regularity

- *We can apparently make the set of contents of an agent's thoughts as big as we like.*
- *But we limit the attitudes that she can bear to those contents*—the attitudes can only achieve a certain fineness of grain.
- Put a rich set of contents together with a relatively impoverished set of attitudes, and you violate regularity.

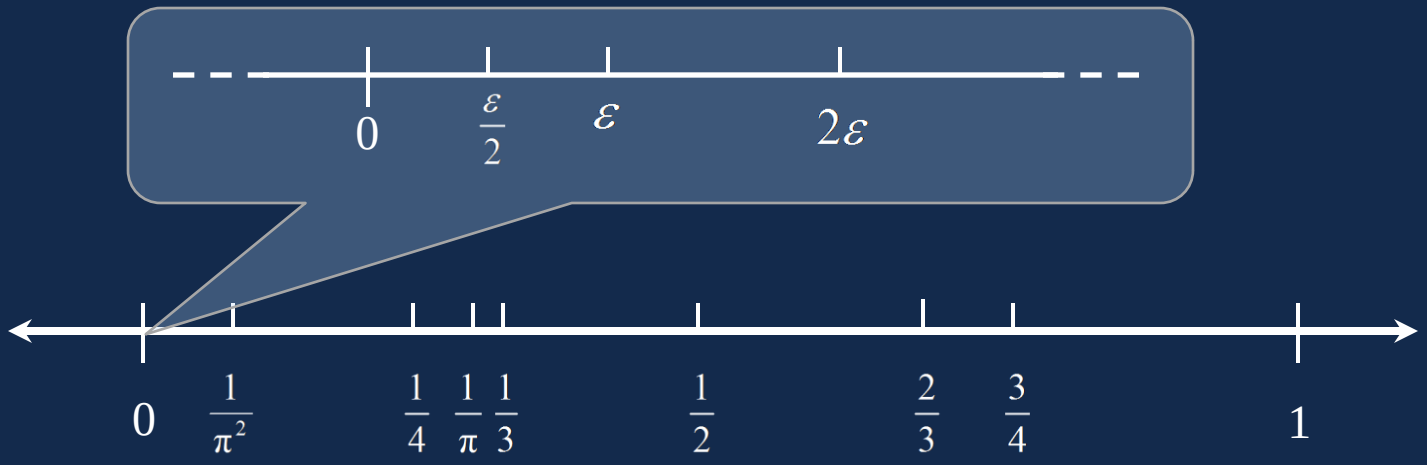
Infinitesimals to the rescue?

The friend of regularity replies: if you're going to have a rich *domain* of the probability function, you'd better have a rich *range.*

Lewis:

"You may protest that there are too many alternative possible worlds to permit regularity. But that is so only if we suppose, as I do not, that the values of the function *C* are restricted to the standard reals. Many propositions must have infinitesimal *C*-values ... (See Bernstein and Wattenberg (1969).)"

Infinitesimals to the rescue?



Infinitesimals to the rescue?

- Bernstein and Wattenberg's article does not substantiate Lewis' strong claim that there are too many possible worlds to permit regularity *only if C*'s values are restricted to the reals.
- Bernstein and Wattenberg show that using infinitesimals, one can give a regular probability assignment to the landing points of our fair dart throw.

Infinitesimals to the rescue?

- But that's a very specific case, with a specific cardinality!
- Lewis himself thinks that the cardinality of the set of possible worlds is greater than that (at least beth-2).
- We need a similar result that holds if the set of possibilities has higher cardinality than that of the real interval [0, 1].
- Indeed, the set of doxastic possibilities may well be *a proper class! …*

Arguments against regularity, even allowing infinitesimals

- I conjectured that a version of the cardinality problem would always arise.
- Pruss proved it: if the cardinality of Ω is greater than that of the range of *P*, then regularity fails.

Arguments against regularity, even allowing infinitesimals

I envisage a kind of arms race:

- We scotched regularity for real-valued probability functions with sufficiently large domains (uncountable).
- The friends of regularity fought back, enriching their ranges: making them hyperreal-valued.
- The enemy of regularity counters by enriching the domain.
- And so it goes.
- By Pruss's result, the enemy can always win (for anything that looks like Kolmogorov's probability theory).

Arguments against regularity, even allowing infinitesimals

- Could we tailor the range of the probability function to the domain, for each particular application? (Like the general of a defense force …)
- The trouble is that in a Kolmogorov-style axiomatization the commitment to the range of *P* comes *first*…
- On the tailoring approach, a probability function is a mapping from *F* to …—well, to *what*?
- What will the additivity axiom look like?
- In any case, this 'wait and see' approach is quite a departure from Kolmogorov.

Arguments against regularity, even allowing infinitesimals

- On a Kolmogorov-style approach, there will always be an Ω that will have non-empty subsets assigned probability 0.

Doxastically possible credence gaps

- I will argue that you can rationally have credence gaps.

Examples of doxastically possible credence gaps

- Non-measurable sets

Dart example

- Certain subsets of Ω—so-called *non-measurable* sets—get no probability assignments whatsoever.

Examples of doxastically possible credence gaps

- Chance gaps
- The Principal Principle says (roughly!!):

  your credence in *X*, *conditional* on it having chance *x*, should be *x*:

  $$C(X \mid \text{chance}(X) = x) = x.$$

Examples of doxastically possible credence gaps

- A relative of the Principal Principle? Roughly:

  your credence in *X*, *conditional* on it being a chance gap, should be gappy:

  $C(X \mid \text{chance}(X) \text{ is } \textit{undefined}) \text{ is } \textit{undefined}$.

- All I need is that rationality *sometimes permits* your credence to be gappy for a hypothesized chance gap.

Examples of doxastically possible credence gaps

- There are arguably various examples of chance gaps:
  - Chance statements themselves
  - Cases of indeterminism without chances: Earman's space invaders, Norton's dome (Eagle)

Examples of doxastically possible credence gaps

- One's own free choices
- Kyburg, Gilboa, Spohn, Levi, Briggs, Liu and Price contend that when I am making a choice, I must regard it as free. In doing so, I *cannot* assign probabilities to my acting in one way rather than another (even though onlookers may be able to do so).
- "Deliberation crowds out prediction"—or better, it crowds out probability.

Examples of doxastically possible credence gaps

- To be sure, these cases of probability gaps are controversial.
- But these authors are committed to there being credence gaps, and thus violations of regularity.
- All I need is that it is *permissible* for them to be credence gaps.

Ramifications of irregularity for Bayesian epistemology and decision theory

- I have argued for two kinds of counterexamples to regularity: rational assignments of zero credences, and rational credence gaps, for doxastic possibilities.
- I now want to explore some of the unwelcome consequences these failures of regularity have for traditional Bayesian epistemology and decision theory.

Problems for the conditional probability ratio formula

- The ratio analysis of conditional probability:

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}$$

... provided P(B) > 0

Problems for the conditional probability ratio formula

- What is the probability that the dart lands on ½, *given* that it lands on ½?
- 1.
- But the ratio formula cannot deliver that result, because P(dart lands on ½) = 0.

Problems for the conditional probability ratio formula

- Gaps create similar problems.
- Take your favorite probability gap, G.
- The probability of G, given G, is 1.
- But the ratio formula cannot deliver that result, because

  $P(G)$ is undefined.

Problems for the conditional probability ratio formula

- We need a more sophisticated account of conditional probability.
- I advocate taking conditional probability as primitive (in the style of Popper and Rényi).

Problems for conditionalization

- The zero-probability problem for the conditional probability formula quickly becomes a problem for the updating rule of *conditionalization*, which is defined in terms of it.
- Suppose the agent learns evidence *E.*

$$P_{new}(X) = P_{old}(X \mid E) \text{ (provided } P_{old}(E) > 0)$$

Problems for conditionalization

- Suppose you *learn* that the dart lands on ½. What should be your *new* probability that the dart lands on ½?
- 1.
- But

$$P_{old}(\text{dart lands on } ½ \mid \text{dart lands on } ½)$$

is undefined, so conditionalization (so defined) cannot give you this advice.

Problems for conditionalization

- Gaps create similar problems.
- Suppose you *learn* that G. What should be your *new* probability for G?
- 1.
- But

  $P_{old}(G \mid G)$

  is undefined, so conditionalization cannot give you this advice.

Problems for conditionalization

- We need a more sophisticated account of conditionalization.
- Primitive conditional probabilities to the rescue!

Problems for independence

- We want to capture the idea of *A* being probabilistically uninformative about *B.*
- *A* and *B* are said to be *independent* just in case

$$P(A \cap B) = P(A)\,P(B).$$

Problems for independence

- According to this account of probabilistic independence, anything with probability 0 is independent of *itself*:

  If $P(X) = 0$, then $P(X \cap X) = 0 = P(X)P(X)$.

- But identity is the ultimate case of (probabilistic) *de*pendence.

Problems for independence

- Suppose you are wondering whether the dart landed on ½. *Nothing* could be more informative than your learning: the dart landed on ½.
- But according to this account of independence, the dart landing on ½ is independent of the dart landing on ½!

Problems for independence

- Gaps create similar problems.
- Suppose you are wondering whether G. *Nothing* could be more informative than your learning: G.
- But there is no verdict from this account of independence.

Problems for independence

- We need a more sophisticated account of independence – e.g. using primitive conditional probabilities.
- Branden Fitelson and I have been working on this!

Problems for expected utility theory

- Arguably the two most important foundations of decision theory are the notion of *expected utility*, and *dominance reasoning.*

Problems for expected utility theory

- And yet probability 0 propositions apparently show that expected utility theory and dominance reasoning can give conflicting verdicts.

Problems for expected utility theory

- Suppose that two options yield the same utility except on a proposition of probability 0; but if that proposition is true, option 1 is far superior to option 2.

Problems for expected utility theory

- You can choose between these two options:
  - Option 1: If the dart lands on 1/2, you get a million dollars; otherwise you get nothing.
  - Option 2: You get nothing.

Problems for expected utility theory

- Expected utility theory apparently says that these options are equally good: they both have an expected utility of 0.
- But dominance reasoning says that option 1 is strictly better than option 2. Which is it to be?
- I say that option 1 is better.
- I think that this is a counterexample to expected utility theory as it is usually interpreted.
- Both evidential and causal.
- (To be sure, there are replies …)

Problems for expected utility theory

- Gaps create similar problems.
- You can choose between these two options:
  - Option 1: If G, you get a million dollars; otherwise you get nothing.
  - Option 2: You get nothing.

Problems for expected utility theory

- Expected utility theory goes silent.
- I say that option 1 is better.
- We need a more sophisticated decision theory.

Conclusion

- Irregularity makes things go bad for the orthodox Bayesian; that is a reason to insist on regularity.
- The trouble is that regularity appears to be untenable.
- I focused on doxastic regularity, but other interesting regularities will meet similar downfalls.
- I think, then, that irregularity is a reason for the orthodox Bayesian to become unorthodox.

Conclusion

- I have advocated replacing the orthodox theory of conditional probability, conditionalization, and independence with alternatives based on Popper/Rényi functions.
- Expected utility theory appears to be similarly in need of revision.

Conclusion

- And then there are some possibilities that *really should* be assigned zero probability …

Thanks especially to Rachael Briggs, David Chalmers, John Cusbert, Kenny Easwaran, Branden Fitelson, Renée Hájek, Thomas Hofweber, Leon Leontyev, Hanti Lin, Aidan Lyon, John Maier, Daniel Nolan, Alexander Pruss, Wolfgang Schwarz, Mike Smithson, Weng Hong Tang, Peter Vranas, Clas Weber, and Sylvia Wenmackers for very helpful comments that led to improvements; to audiences at Stirling, the ANU, the AAP, UBC, Alberta, Rutgers, NYU, Berkeley, MIT, Miami, Princeton, Cornell, Northwestern, the Lofotens Epistemology conference, the Formal Epistemology Workshop (Munich), the Epistemic Rationality Conference (Barcelona); to Carl Brusse and Elle Benjamin for help with the slides; and to Tilly.

Big data, logic of scientific discovery, and abduction

Young E. Rhee
Kangwon National University
Korea

Purpose

- To examine the very odd view that big data is enough and we don't need theories and models in doing science.

- Especially, to examine the main implications of the view to philosophy of science: causation- correlation, theory generation, logic of scientific discovery.

Contents

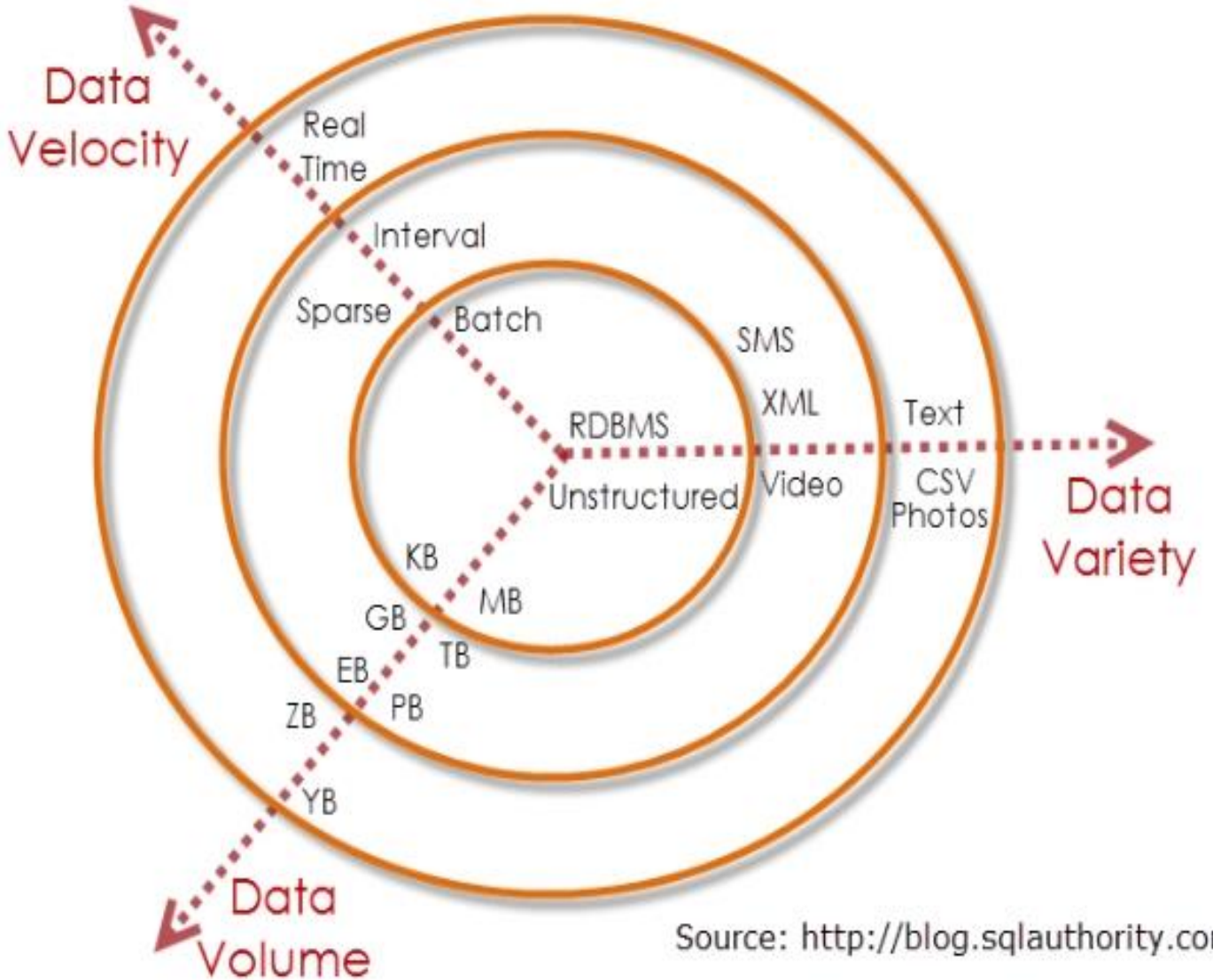
1. Characteristics of big data
2. The very odd view
3. Causation-correlation
4. No need for theories in doing science
5. Big data with the logic of scientific discovery
6. Prospects

1. Characteristics of big data

- 3Vs

a. Volume: great volume
b. Velocity: rapid collecting, generating, processing
c. Variety: various modalities such as audio, video, image, …

3Vs of Big Data



Source: http://blog.sqlauthority.com

- “Big data is not a ‘thing’ but instead a dynamic/activity that across many IT borders.” (M. Chen, S. Mao, and Y. Liu, 2014)

- “Big Data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis.” (J. Gantz and D. Reinsel, 2011)

2. The very odd view

"There is now a better way. Petabytes allow us to say: Correlation is enough.' . . .We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot . . . Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all. There's no reason to cling to our old ways. It's time to ask: What can science learn from Google?"

(C. Anderson, 2008)

Main implications of the view

- The data deluge makes the scientific method obsolete.

a. No causation: correlation is enough and it supersedes causation.
b. No theory: Big data is enough and we don’t need scientific theories or models.
c. Big data as the logic of scientific discovery.

3. Causation-correlation

- Hume's view on causation (1738-40)

- When we say of two types of object or event that "X causes Y" (e.g., fire causes smoke), we mean that (a) Xs are "constantly conjoined" with Ys, (b) Ys follow Xs and not vice versa, and (c) there is a "necessary connection" between Xs and Ys such that whenever an X occurs, a Y must follow.

- Unlike the ideas of contiguity and succession, the idea of necessary connection is subjective.

Where is causation in scientific laws?

- Newton's Second Law: F= ma
- Planck's law: $B_\nu(\nu, T) = \frac{2h\nu^3}{c^2} \frac{1}{e^{\frac{h\nu}{k_\mathrm{B}T}} - 1}$
- It depends upon the notion of hidden variables in scientific realism that the apparent randomness of a system depends not upon collapsing wave functions but rather due to unseen or unmeasurable variables.

What matters

- Not between causation and correlation
- But between reliable correlation and spurious correlation

- Then, can big data find the reliable correlations in nature?

Maybe, but impossible in principle

- Maybe, because of the 4Vs(Volume, Velocity, Variety, values) of big data.

- Impossible, because the chance of finding it will be zero.

$$\text{Probability} = \frac{\text{however big}}{\infty} = 0$$

From correlation to causation

- Combinational explosion found in neural system.

- Building block approach(J. Searle, 2004)

a. Find neural correlates of consciousness
b. Test to see if the correlation is causal
c. Get a theory

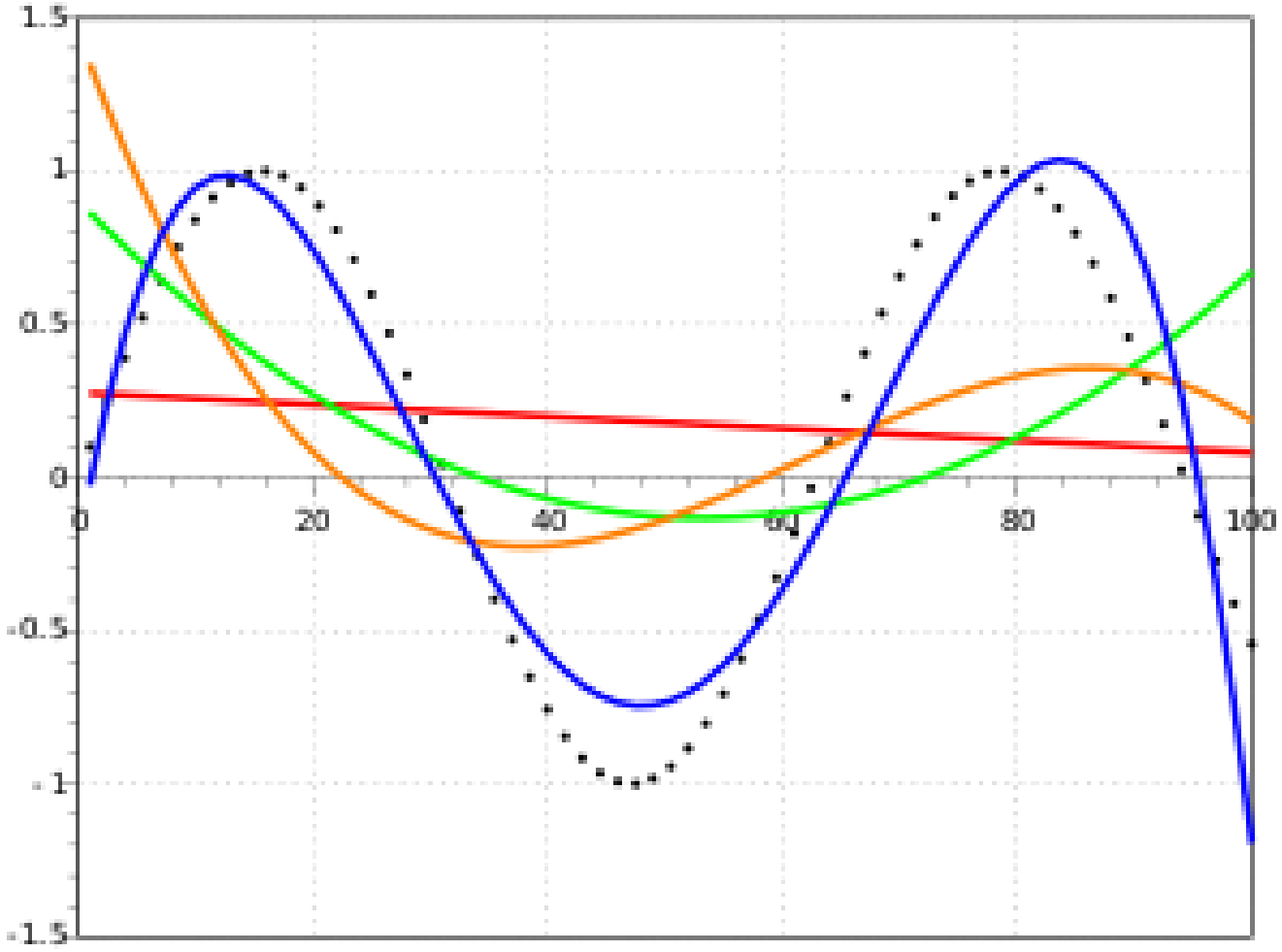
4. No need for theories in doing science

- Newton: *Hypotheses non fingo* (1713)

"I have not as yet been able to discover the reason for these properties of gravity from phenomena, and I do not feign hypotheses. For whatever is not deduced from the phenomena must be called a hypothesis; and hypotheses, whether metaphysical or physical, or based on occult qualities, or mechanical, have no place in experimental philosophy. In this philosophy particular propositions are inferred from the phenomena, and afterwards rendered general by induction"

Limits of big data

a. Big data but few patterns: “We have shifted from the problem of what to save to the problem of what to erase.” (l. Floridi, 2012)
b. Low degree of reliability
c. Possibility of spurious correlation
d. Curve-fitting problem



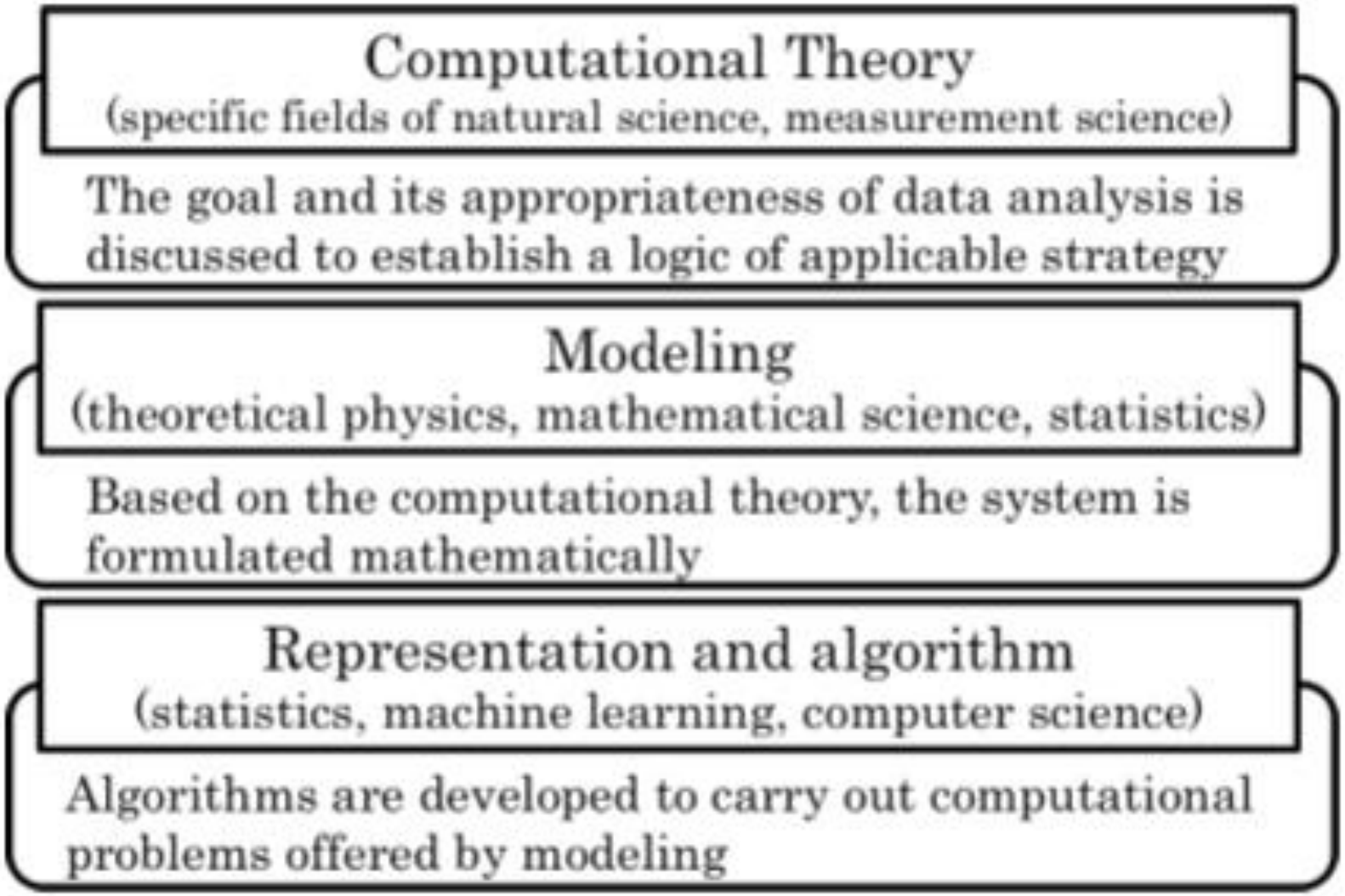
Polynomial curves fitting points generated with a sine function.
Red line is a first degree polynomial, green line is second degree, orange line is third degree and blue line is fourth degree. From Wikipedia.

Data-driven science

- The great empiricists of the 17th century believed that if we used our senses to collect as much data as possible, we would ultimately understand our world.

- Johannes Kepler(1627): Discovery of the laws of planetary motion in Rudolphine Tables based on his analysis of Tycho Brahe's observational data.

- Three levels of data-driven science(Y. Igarashi et al, 2015)



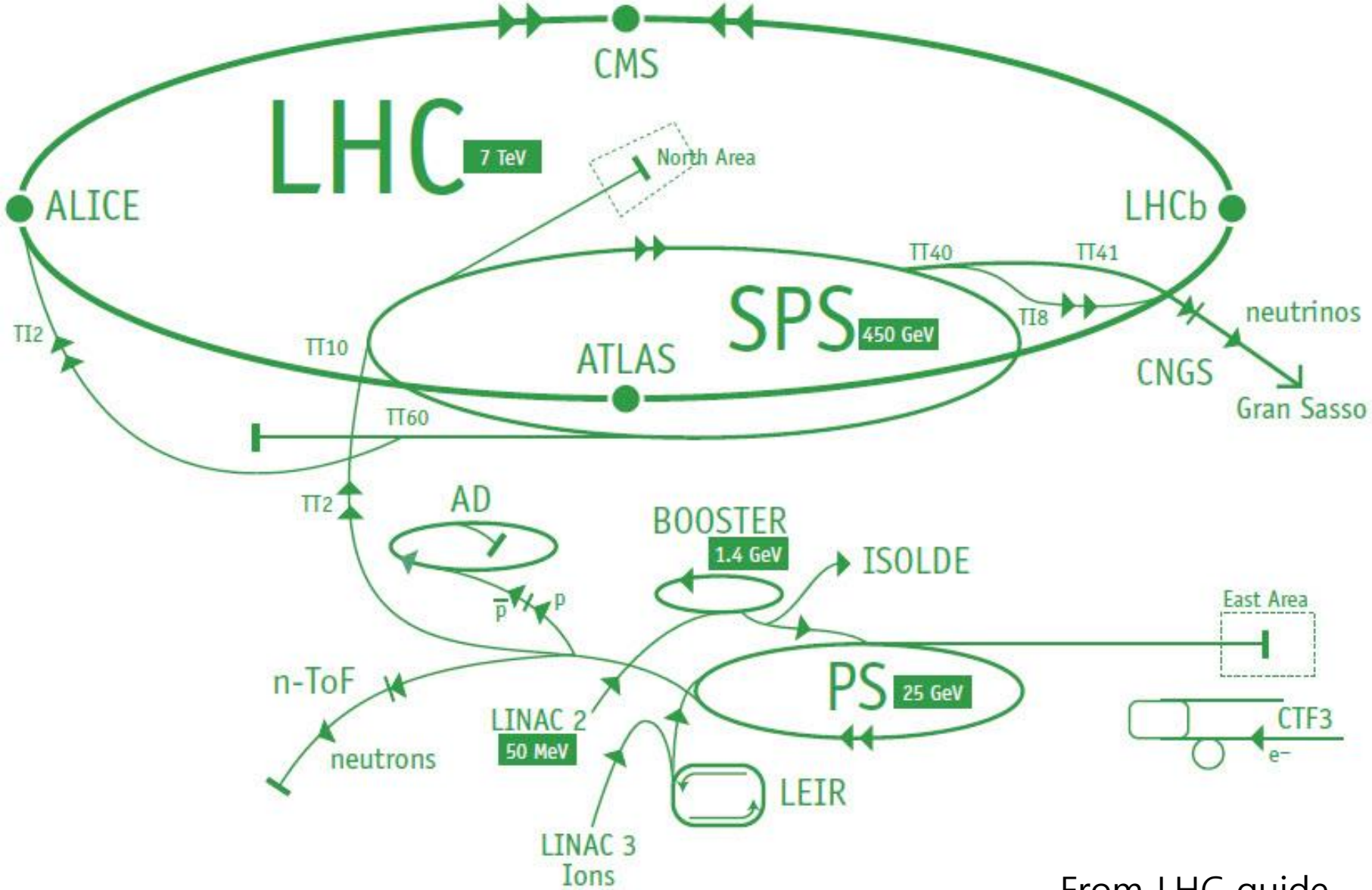
Henri Poincaré

“It is often said that experiments should be made without preconceived ideas. That is impossible. Not only would it make every experiment fruitless, but even if we wished to do so, it could not be done. (1902)

Discovery of Higgs boson

- The discovery of the Higgs boson was not data-driven. The Large Hadron Collider(LHC, 27km) at CERN experiments were mostly driven by theoretical predictions. (F. Mazzocchi, 2015)

- The LHC, big data, distributed computing, and sophisticated data analysis all played a crucial role in the discovery of the Higgs boson.



From LHC guide

Physical Review Letters 114 (15 May 2015)

- Title : “Combined Measurement of the Higgs Boson Mass in pp Collisions at s √ =7 and 8 TeV with the ATLAS and CMS Experiments”

- No. of authors: 5154
- Research contents took only 7 pages among 33 pages.

How come without theories or models?

- Without theories or models,

a. Impossible to design experiments
b. Impossible to choose data to be collected
c. Impossible to classify data collected
d. Enormous chunk of data

So far,

- It turned out

a. It is not clear how to discriminate between reliable correlation and spurious correlation, admitting that there is no causation in nature.

b. It is impossible to design experiments, to choose data to be collected, and to classify date collected.

Dangerous dichotomy

Data-driven approach	Theory-driven approach
Inductivism	Deductivism
Empiricism	Hypothetico-deductivism

- Big data does not belong either of the two, but a combination of them.

5. Big data with the logic of scientific discovery

- K. Popper, C. G. Hempel

a. Scientific activity is fundamentally composed of two distinguishable aspects, discovery and justification of a hypothesis.
b. The justification process obeys the application of logical rules, so justification is a rational activity.
c. Discovery is of a non-rational character.

K. Popper

"… there is no such thing as a logical method of having new ideas, or a logical reconstruction of this process. My view may be expressed by saying that every discovery contains 'an irrational element', or 'a creative intuition', in Bergson's sense"

(1959, p. 8)

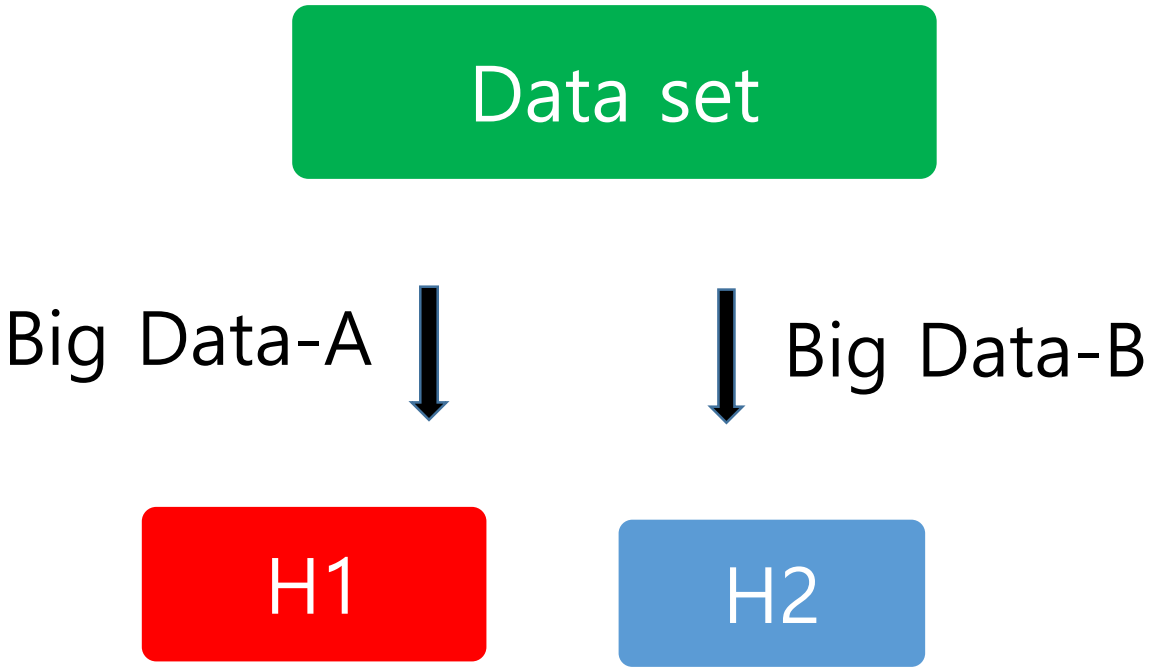
C. S. Peirce

- Abduction as a logic of science (1903)

  The surprising fact, *C*, is observed.
  But if *A* were true, *C* would be a matter of course.
  Hence, there is reason to suspect that *A* is true.

- Big data with abduction can generate hypotheses such that it can overcome the problem of poor patterns.
- How can we construct the system? Deep neural networks such as AlphaGo?

Underdetermination of theories by data



When H1 and H2 are competing, how big data can decide them?

6. Prospects

- There will be many such gig data that can discover quantitative empirical laws, such as BACON or GLAUBER(P. Langley et al, 1987)
- But it would be very difficult to develop such big data equipped with inference engines that generate hypotheses such as abduction.
- Science to save just phenomena dominate in the future.

References

- Anderson, C. 2008. The end of theory: The data deluge makes the scientific method obsolete. *Wired* 23. http://www.wired.com/science/discoveries/magazine/16-07/pb_theory(Accsed 17 November 2017).
- Chen M, Mao, S., and Liu. Y. 2014. Big Data: A survey. *Mobile Networks and Applications* 19(2): 171–209.
- Floridi, L. 2012. Big data and their epistemological challenge. *Philosophy & Technology* 25: 435-437.
- Gantz, J. and Reinsel, D. 2011. Extracting value from chaos. *IDC iview* 1142: 9–10.

- Hume,. D. 1739-40/1978. *A treatise of human nature*. Oxford University Press
- Igarashi, Y. Nagata, K. Kuwatani, T. Omori, T. Nakanishi-Ohno, Y. and Okada M. 2016. Three levels of data-driven science. *Journal of Physics: Conference Series* 699: 2-13
- Kitchin, R. 2013. Big data and human geography: Opportunities, challenges and risks. *Dialogues in Human Geography* 3(3): 262–267.
- Langley, P. Simon, H. A. Bradshaw, G. L. Zytkow, J. M. 1987. Scientific discovery: Computational explorations of the creative process. MIT Press.

- Mayer-Schönberger, V. and Cukier, K. 2013. *Big data: A revolution that will transform how we live, work and think*. Houghton Mifflin Harcourt.
- Mazzocchi, F. 2015. Could big data be the end of theory in science? *EMBO reports* 16(10): 1250-1255.
- Newton, I. 1713. *Philosophiae naturalis principia mathematica* 2nd edition.
- Peirce, C. S. 1903/1998. Harvard Lectures on Pragmatism. *The essential Peirce*, Vol. 2. Indiana University Press.
- Poincaré, H. 1902/1913. *Science and hypothesis*. Science Press.
- Popper, K. 1935/59. *The logic of scientific discovery*. Routledge.
- Searle, J. 2004. *Mind* : *A brief introduction*. Oxford University Press.

Individuating Genes as Types or Individuals

Ruey-Lin Chen
pyrlc@ccu.edu.tw
Department of Philosophy
National Chung Cheng University, Taiwan

Abstract

In this paper, I argue that there are at least two kinds of individuation of genes. The transgenic technique can individuate "a gene" as an individual while the technique of gene mapping in classical genetics can only individuate "a gene" as a type or a kind. The two kinds of individuation involve different techniques, different objects that are individuated, and different references of the term "a gene". Thus, I also discuss this semantic phenomenon in using "gene" and the problem about the relation between kinds and individuals in the individuation of genes.

1. Introduction

This paper discusses individuation of genes from the perspective of scientific practice. It thus involves both the issue of biological individuality and the issue of approaches from theoretic constructions or experimental practices. In this paper, I argue that the transgenic technique can individuate "a gene" as an individual while the technique of gene mapping in classical genetics can only individuate "a gene" as a type or a kind. Herein we find double extensions of the term "a gene". Therefore, I also discuss this semantic phenomenon in using "gene".

Biological individuality has become a central issue in the philosophical discussion of biology, in which organismal entities such as a human, a dog, a banyan tree, a mushroom, a bacterial cell, etc. grasp most philosophers' attention (Wilson and Baker 2013; Guay and Pradeu 2016; Ligard and Nyhart 2017).. However, there are many diverse cases of supra-organismal entities such as colonies, groups, populations, species, etc. (Bouchard and Huneman 2013) and sub-organismal entities such as stem cells, genes, gene-networks, genomes, etc. (Dupré and O'Malley 2007, 2009; Dupré 2012; Fagan 2016) – all of which plausibly qualify as living individuals. Among a variety of biological individuals, the issue about the individuality of the gene is the other focus which has gotten most philosophers' concerns ever since the 1980s.

To date, "what is a gene?" and other related questions have been asked many times by philosophers, historians, and scientists of biology (Kitcher 1982, 1992; Falk 1986, 2010; Carlson 1991; Maienchein 1992; Portin 1993; Waters 1994, 2007, 2018; Beurton, Falk, and Rheinberger 2000; Snyder and Gerstein 2003; Stotz and Griffiths 2004, 2013; Pearson 2006; Reydon 2009; Baetu 2012; Rheinberger 2015). Those questions were frequently embedded in the discussion about the definition of "gene" and the gene concept. Philosophers with one another have disputed and continued to dispute on whether or not there is a single or united definition or concept of gene. Synthetizing the

influential literature, let me identify the following four positions: skepticism (Kitcher 1992), dualism (Moss 2003), pluralism (Griffths and Stotz 2013), and pragmatism (Waters 2007, 2018).
[1]

In my view, the discussion about "what is a gene" in the literature has two distinctive features: (1) "A gene" refers to a type of gene (i.e., a genotype) rather than an individual gene.[2] (2) Although philosophers inquire into the gene concept and the conceptual change of gene by examining scientific practices, they seldom consider the role of the transgenic technique developed in biotechnology may play in the philosophical discussion. This paper explores experimental individuation of genes along the alternative direction (i.e., the transgenic technology) and considers the possibility that a gene is individuated as an individual.

The question of what a gene is explicitly presupposes the problem of the gene individuality; and identifying a gene presupposes individuating the gene. According to the literature of analytic metaphysics, "individuation" is traditionally understood in a metaphysical and an epistemological sense.[3] Beuno, Chen, and Fagan (2018) add a practical sense to the term and interpret "individuation" by connecting the three senses from the process perspective and the scientific practice perspective. They characterize "individuation" and "individuals" as "*an individual emerges from a process of individuation in the metaphysical sense. Epistemic and practical individuation, then, are processes that aim to uncover stages of that metaphysical process*." (Beuno, Chen, and Fagan 2018, in production) The approach to individuation of genes adopted in this paper follows their characterization, especially focusing on the process of epistemic and practical individuation.

In the case of the classical gene, Gregor Mendel assumed that hereditary factors of features are unitary and corpuscular things – scientists called them genes later. The classical geneticists built up a theory of genes and performed breeding experiments to identify a specific gene in a specific chromosome. In this sense, one may well say that the classical geneticists individuate genes via experiments. Philosophers usually call the gene in the sense of classical genetics Mendelian gene.

From the view of molecular biology, the boundary and size of a located Mendelian gene cannot be delineated exactly and clearly. Moreover, as many philosophers have strongly argued, a Mendelian gene does not exactly correspond to a molecular gene. Therefore, the technique of genetic maps does not genuinely individuate a (Mendelian) gene (as a particular). However, the definition of the molecular gene concept has its own troubles, as we have seen in the literature mentioned above.

In this paper, I address the two questions: (Q1) In what sense, one can reasonably say that the classical geneticists have individuated a gene? The answer is that they individuate a gene as a type. (Q2) Are there experiments that can individuate a gene as a particular (i.e., an individual)? The answer is the experiments by using the transgenic technique. The answers to the two questions indicate two different objects of

---

[1] Gene pluralism and gene pragmatism may be viewed as the same position, because they both are established by considering scientific practices.

[2] Rosenberg (2006) raises the issue of "gene individuality thesis" which is parallel to the "species individuality thesis". Also, see Reydon's objection (2009).

[3] A summary discussion, see Beuno, Chen, and Fagan (2018).

individuation: individuation of a type and individuation of an individual. One may wonder whether or not "the individuation of a type" is an inconsistent phrase. In order to answer this question, I will discuss in what sense we individuate a type. My discussion thus involves the relationship between kind and individual in the context of experimentation.

2. Chromosomal location of a gene

The geneticists could locate and label some specific genes at some specific chromosomes. As figure 1 shows, the chromosomal location of the CFTR (Cystic fibrosis transmembrane conductance regulator) gene is <7, q, 3, 1 .2>, which is taken as the locus of the CFTR gene at the chromosome in human cells.[4]

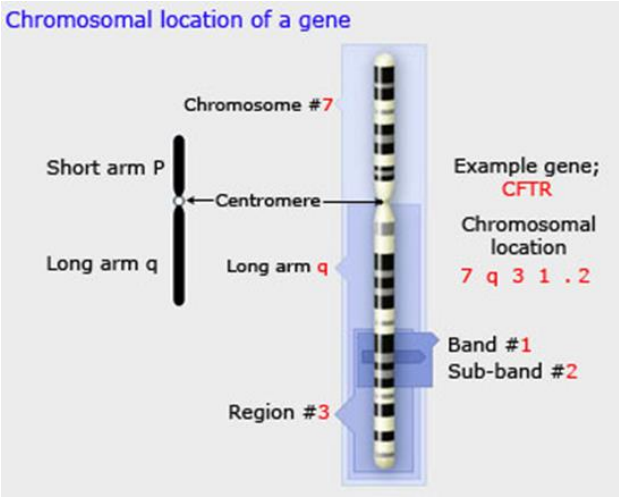


Fig. 1 Chromosome Seven in human cells

In the symbolic tuple <7, q, 3, 1 .2>, "7" represents the seventh chromosome (Chromosome 7) in human cell; "q" represents the long arm q in Chromosome 7; "3" represents Region 3 in q; and "1.2" represents band 1 and sub-band 2. Since scientists use the locus of the CTRF gene to identify the gene itself, and thus the symbolic tuple <7, q, 3, 1.2> may be used to represent the CTRF gene in the framework of the classical genetics. Accordingly, may we say that the location of a gene individuates the gene? Before answering the question, it is necessary to discuss how classical geneticists did chromosomal location of genes. In other words, what technique is used in the process of locating genes?

chromosomal location of genes is a well-known story. Many philosophers of biology have retold it once again (Darden 1991, Waters 1994, Falk 2009, Griffths and Stotz

[4] "The CFTR gene codes for an ABC transporter-class ion channel protein that conducts chloride and thiocyanate ions across epithelial cell membranes. Mutations of the CFTR gene affecting chloride ion channel function lead to dysregulation of epithelial fluid transport in the lung, pancreas and other organs, resulting in cystic fibrosis. Complications include thickened mucus in the lungs with frequent respiratory infections, and pancreatic insufficiency giving rise to malnutrition and diabetes. These conditions lead to chronic disability and reduced life expectancy." (Wikipedia Encyclopedia: https://en.wikipedia.org/wiki/Cystic_fibrosis_transmembrane_conductance_regulator)

2013). For the purpose of this paper, I introduce a very brief version of this story.

In the 1910s, Thomas Hunt Morgan and A. H. Sturtevant developed a technique to map the linear relations among factors (genes) in linkage groups. They used Mendelian breeding data. Morgan and his team discovered a pair of chromosomes may cross a fragment over with each other in the period of meiosis. *Crossing over* produces a specific ratio of the linked traits. Morgan thought that "the percentage of crossing over is an expression of the 'distance' of the factors from each other." (Morgan et.al. 1915: 61) Sturtevant used percentages of linked characters with crossing over from breeding experiments to calculate the relative positions of the factors to each other. This is the kernel technique for constructing genetic maps. By using genetic maps, they determined the locations of many genes at the four chromosomes of *Drosophila melanogaster* (fruit fly). See figure 2. Given the genetic maps, the classical geneticists assume that no other genes could locate at the same position of a chromosome. As a consequence, *the single location of a gene indicates the individuality of genes*!

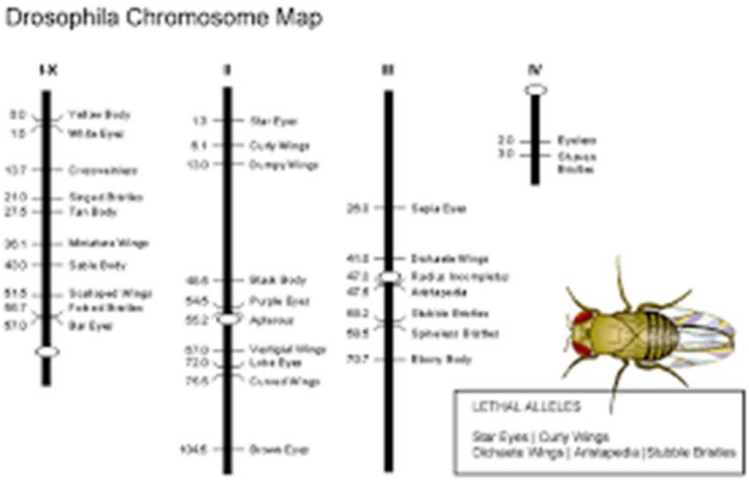


Figure 2: The genetic map of *Drosophila*

Genetic maps by nature are diagrammatic models for loci of genes in chromosomes. They are reasoning consequences from statistical data of breeding experiments. Models represent what are general. When we say that chromosomal location of a gene in a genetic map represents the linear locus of a Mendelian gene at a chromosome, we really mean that it represents the locus of a type of Mendelian gene at an identical type of chromosome in a cell within a kind of organism. Of course, this implies that a token of a type of Mendelian gene at a token of a type of chromosome can be cognitively identified and discerned, because we can distinguish it from other tokens of other genes. Consequently, we can count genes at a cell and cognitively distinguish it from others. The located genes satisfies the two traditional criteria of individuality: *distinguishability* and *countability*.

If all chromosomes were stick-shaped substances constituted by uniform material and had no complicated structure, then chromosomal location of Mendelian genes would be able to genuinely individuate them. According to the knowledge from molecular biology, however, chromosomes are a long chain of double helix DNA molecules that curl themselves up as being twisted. In such a case, we cannot delineate a located

Mendelian gene or depict its contour or boundary, because the chromosomal position at which the gene locates includes a twisted part of the long DNA molecule. Even if one invokes the knowledge from molecular biology, she would still be disturbed by the puzzling problem of defining the molecular gene.

3. Individuating molecular genes as individuals

Ever since the era of molecular biology, the continuously accumulated knowledge out of genetics have not solved the problem of individuation of genes, instead brought more troubles about the definition of the gene concept. Is a gene "a sequence of DNA for encoding and producing a polypeptide"? Should we include the start and stop codons? Should we count those introns deleted during the process of transcription into the investigated gene? The difficulty in defining the gene concept brings forth the impediment in individuating a gene. The transgenic technique of biotechnology developed from molecular biology made breakthrough. It can individuate (molecular) genes as particulars (or individuals), while other experiments in molecular biology have been performed without a clear concept of the gene and a gene. Why can the transgenic technique do so? What condition according to which it can individuate a gene as an individual?

Chen (2016) has argued that the first experiment of bacteria transformation individuated an antibiotic resistance gene by developing a conception of experimental individuality with three attendant criteria: *separability, manipulability, and maintainability of structural unity*.[5] How did the experiment satisfy the three criteria?

According to Chen (2016: 360-363), Stanley Cohen and Herbert Boyer combined DNA of *Escherichia coli* (*E. coli*) in 1973 and 1974 by transferring two different DNA segments encoding proteins for ampicillin and tetracycline resistance into *E. coli*. They thereby realized the transformation of this bacterium. Both DNA segments are called "antibiotic resistance gene." Cohen and Boyer used small circular plasmids (extrachromosomal pieces of DNA) as vectors to transfer a foreign DNA segment into a bacterial cell. The plasmids were made by cutting out a (supposed) antibiotic resistance *gene* from other bacteria with the restriction enzyme *Eco*RI, linking the gene into a plasmid by using another enzyme, DNA ligase, and transferring the plasmid into an *E. coli* cell lacking the ability to resist antibiotics. The result, a modified *E. coli* cell, was able to resist antibiotics and contained the antibiotic resistance gene. In this experiment, the antibiotic gene was separated from its original bacteria and then was manipulated (i.e., linked and transferred). Its structural unity was not broke down, thereby allowing it to be expressed in the other kind of bacteria throughout the process. Scientists thus identify it as *a* gene – an individual biological entity. Therefore, the separated, manipulated, and maintained antibiotic gene was naturally separable, manipulable, and maintainable.

Herein we interpret the performance of the techniques used in transgenic experiments as the general process of individuating transgenes. The process of individuating a gene has five stages in general. Figure 3 depicts a partial process of

[5] In his paper, Chen uses the creation of Bose-Einstein condensates in laboratories as the other example from physical experiments. Chen's intent is to argue that biological entities and physical entities in laboratories share the same criteria of experimental individuality.

experiments by using the transgenic technique.

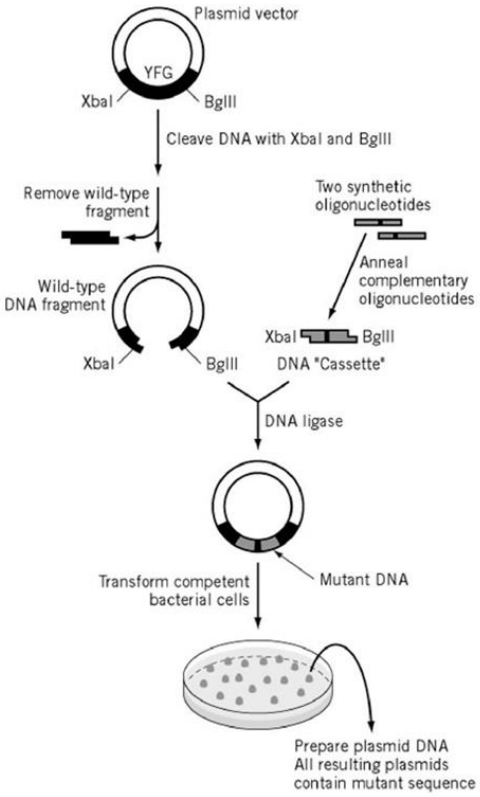


Figure 3: The technological process of cleaving, linking, and cloning a gene

The technological process with five stages can be described as follows:

(1) At the first stage, the technique requires scientists use restriction enzymes to cleave specific segments from recognition sites of long DNA chains. A specific restriction enzyme can cut away a specific DNA segment at a specific site.

(2) At the second stage, the technique requires scientists link the cleaved segment of DNA to a plasmid vector by using DNA ligase. The vector is a circular DNA that may come from a wild type of virus.

(3) At the third stage, scientists try to paste the DNA segment in the vector onto the genome of another organism by injecting the plasmid vector to a cell of the target organism. Of course, they may fail when the intended feature is not expressed.

(4) At the fourth stage, scientists make copies of DNA segments by cloning the cell containing the transferred segment of DNA. The aim of DNA cloning is to copy a segment of interest or a gene from an organism and massively produce its copies.

(5) At the fifth stage, scientists are required to observe the expression of the novel feature that the target organism does not have. If a DNA segment cut from an original organism is successfully pasted on cell of a target organism and the target organism expresses the intended feature that the original organism has, then one can concludes that the segment is a gene.

Herein the first stage is correspondent to the operation of the separation condition, the second, the third and the fourth stages to the manipulation condition, and the fifth stage to the maintenance condition. Accordingly, one can easily see that this cut, linked, transferred, pasted, and copied gene is a particular – an individual, not only because it satisfies the three criteria of experimental individuality but also because it is *single* and *particular*. In other words, the single segment of DNA maintains its structural unity when being separated and manipulated. This is so, because cutting a gene from an original organism is separating it from its environment, and because transferring, pasting and copying a gene is manipulating it. If the gene does express the intended feature in an unoriginal organism, this condition indicates that the unity of its chemical and informational structure is maintained.

This kind of experimental individuation of genes implies a special conception of gene which is defined by the transgenic technique. I call this conception "the transgenic conception of gene."

4. Two kinds of individuation of genes

The previous discussion in section 2 and 3 indicates that two different objects have been individuated in different experimental and theoretical contexts. In the context of classical genetics, scientists used breeding experiments and theoretical inferences to locate a gene at some position in some chromosome. As we have argued, they assumed that no other genes could coexist at the same position and thus individuated genes as types. If one interpret the meaning of "individuation" as "only individuals can be individuated," then the phrase "individuating genes as types" sounds unreasonable. Is it better to say "*unitization* of genes" rather than "individuation of genes"? It is quite right to say the classical geneticists *unitizing* genes as types. In a sense, however, we may reasonably say that we individuate a gene as a type, because the type has tokens or members that are distinguishable and countable individuals. Classical geneticists suppose that all types of genes (genotypes) have corpuscular members, i.e., individuals. In such a sense, to talk of "individuating genes as types" is reasonable. If a kind that has no distinguishable and countable members, then the kind cannot be individuated. That said, we cannot individuate such a kind, for examples, water or air that is expressed by "mass" nouns in the level of non-molecules. Of course, we may individuate a water molecule in the level of molecule. In the cases of experiments using transgenic technology, molecular biologists really individuate a singular, particular, and unique gene. Thus, we claim that scientists experimentally individuating genes as individuals in such a context.

The two individuated objects indicate two different references of the term "a gene" in the literature. As we have seen, many philosophers and scientists have asked "what is a gene" once again. They really refer to a type of gene, when they use "a gene" in discussing the gene concept or the definition of "gene." Similarly, in some contexts of scientific investigation, scientists use "a gene" to refer to a type of gene as the phrase "chromosomal location of a gene". In the context of transgenic experiments, however, we use "a gene" to refer to a genuine individual – a single, particular, and unique gene, because it can maintain its structural unity when being separated and manipulated in the process of experimenting.

The two different objects and references indicate two different kinds of experimental individuation, which are realized by two different techniques: the technique of genetic location and the technique of transgenes. In the two different contexts, scientists appeal to different sets of criteria for individuality.

Experiments by using the technique of genetic location individuate a type of objects whose tokens or members are countable individuals rather than matter referred by uncountable nouns. In this kind of experimental contexts, we emphasize the distinguishability and the countability as the central features of individuals. Experiments by using the technique of transgenes individuate a single, particular, and unique individual. In this kind of experimental contexts, we emphasize the particularity and the uniqueness of individuals in contrast to the universality of types or kinds. We assure the particularity and uniqueness of individuals by the realization of experimental individuality, namely, the joint realization of separability, manipulability, and the maintainability of structural unity. In a summary, different techniques produce different kinds of individuation.

One may wonder: Can the technique of genetic locution individuate a singular, particular, and unique gene in the sense of individuating entities as individuals? I think the answer is negative, because that technique cannot separate and manipulate a gene and maintain its structural unity. On the contrary, one may ask: Can the technique of transgenes individuate a type of gene? I give an uncertain answer. In the sense that scientists suppose that a member of a genotype has been individuated in transgenic experiments, thus, we are allowed to say that the technique also individuate a genotype. However, scientists cannot make sure that the technique of transgenic can be applied to every genotype. In fact, the probability of failure is quite high. Unless the experimentally individuation of a particular gene can be performed repeatedly and stably, then one can say that the gene individual indicate a *general* type of gene and that the genotype has been individuated as an individual. But the object individuated by the technique is not a type of gene, because the technique always requires manipulating a particular segment of DNA -- a gene.

5. Conclusion

I conclude that there are at least two kinds of experimental individuation of genes. Individuating a gene as a type or individuating a gene as an individual depends on the technology used in the experiment of individuation. Furthermore, I claim that different kinds of experimental individuation presuppose different conceptions of gene: the classical conception of Mendelian gene and the transgenic conception of gene. One may wonder what the relation between the transgenic conception of gene and the molecular gene concept is. In addition, one may wonder whether we can unify different conceptions of gene by integrating different experimental techniques, say, the technique of genetic location, the technique of genetic sequencing, and the transgenic technique. Maybe? But the two problems will be left for the future.

References

Baetu, Tudor M., 2012. “Genes after the Human Genome Project.” *Studies in History and Philosophy of Biological and Biomedical Science*, 43: 191-201.
Beurton, P., R. Falk, and H.- J. Rheinberger, 2000. *The Concept of the Gene in Development and Evolution: Historical and Epistemological Perspectives*. Cambridge, UK: Cambridge University Press.
Beuno, Otavio, Ruey-Lin Chen, and Melinda B. Fagan, 2018. “Individuation, Process, and Scientific Practice.” In Otavio Beuno, Ruey-Lin Chen and Melinda B. Fagan (eds). *Individuation, Process, and Scientific Practice*. New York: oxford University Press. (In production)
Bouchard, Frédéric and Philippe Huneman, 2013. *From Groups to Individuals: Evolution and Emerging Individuality*. Cambridge, Mass.: The MIT Press.
Chen, Ruey-Lin, 2016. “The experimental realization of individuality.” *Individuals across the Sciences*, ed. Alexandre Guay and Thomas Pradeu, 348-370. New York: Oxford University Press.
Carlson, E. (1991). “Defining the Gene: An Evolving Concept.” *American Journal of Human Genetics*, 49: 475-487.
Dupré, John, 2012. *Processes of Life: Essays in the Philosophy of Biology*. Oxford: Oxford University Press.
Dupré, John and Maureen O’Malley, 2007. Mategenomics and Biological Ontology. *Studies in the History and Philosophy of the Biological and Biomedical Science* 38: 834-846. Collected in John Dupré. 2012. *Processes of life: Essays in The Philosophy of Biology*, 188-205. Oxford: Oxford University Press.
Dupré, John and Maureen O’Malley, 2009. Varieties of Living Things: Life at the Intersection of Lineage and Metabolism. *Philosophy and Theory in Biology* 1: 1-25. Collected in John Dupré. 2012. *Processes of life: Essays in The Philosophy of Biology*, 206-229. Oxford: Oxford University Press.
Falk, Raphael, 1986. “What is a gene?” *Studies in History and Philosophy of Science*, 17: 133-173.
Falk, Raphael, 2010. “What is a gene – revised” *Studies in History and Philosophy of Biological and Biomedical Science*, 41: 396-406.
Fagan, Melinda B., 2016. “Cell and body: Individuals in stem cell biology.” In Individuals across the Sciences, ed. Alexandre Guay and Thomas Pradeu, 122-143. New York: Oxford University Press.
Griffths, Paul and Karola Stotz, 2013. *Genetics and Philosophy: An Introduction*. Cambridge: Cambridge University Press.
Guay, Alexandre and Thomas Pradeu, 2016. *Individuals across the Sciences*. New York: Oxford University Press.
Kitcher, P. S., 1982. “Genes.” *British Journal for the Philosophy of Science*, **33**: 337-359.
Kitcher, P. S., 1992. “Gene: current usages.” In E. Keller and L Lloyd (eds.), *Keywords in Evolutionary Biology*, Cambridge, MA: Harvard University Press, pp. 128-131.
Lidgard, Scott and Lynn K. Nyhart, 2017. *Biological Individuality*. Chicago: The University of Chicago Press.

Maienchein, J., 1992. “Gene: Historical perspectives.” In E. Keller and E. Lloyd (eds.). *Keywords in evolutionary biology*. Cambridge, MA: Harvard University Press, pp. 181-187.
Morgan, Thomas Hunt, et.al., 1915. *The Mechanism of Mendelian Heredity*. New York: Henry Holt and Company.
Moss, Lenny, 2003. *What Genes Can’t Do*. Cambridge, Mass.: The MIT Press.
Pearson, Helen, 2006. “What is a Gene?” Nature, 441(25): 399-401.
Portin, P. (1993). The concept of the gene: Short history and present status. *Quarterly Review of Biology*, 68(2): 173-223.
Reydon, Thomas, 2009. “Gene Names as Proper Names of Individuals: An Assessment.” *British Journal for the Philosophy of Science*, 60(2): 409-432.
Rheinberger, Hans-Joerg, 2015. “Gene.” *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/gene/
Rosenberg, Alexander, 2006. *Darwinian Reductionism*. Chicago: The University of Chicago Press.
Snyder, Michael and Mark Gerstein, 2003. “Defining Genes in the Genomics Era.” *Science*, 300(5617): 258-260.
Stotz, Karola and Paul Griffths, 2004. “Genes: Philosophical Analyses Put to the Test.” *History and Philosophy of the Life Sciences*, 26: 5-28.
Waters, Kenneth C., 1994. “Genes made molecular,” *Philosophy of Science*, 61: 163–185.
Waters, Kenneth C., 2007. “Molecular genetics,” *Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/molecular-genetics/
Waters, Kenneth C., 2018. “Don’t Ask ‘What *is* an Individual?’” In Otavio Beuno, Ruey-Lin Chen and Melinda B. Fagan (eds). *Individuation, Process, and Scientific Practice*. New York: oxford University Press. (In production)
Wilson, Robert and Matthew Barker, 2013. The Biological Notion of Individual. *Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/biology-individual/

Cognitive Decision Theory and Permissive Rationality

Jaemin Jung

Wonkwang University

1 Introduction

Standard decision theory says that a rational action to perform is one with the greatest expected utility—one such that the expected *practical* utility for how states of the world will turn out, given your performing it, is at least as high as that of any alternative act you might perform. Recently, many philosophers have deployed a decision theoretic approach to justify *epistemic* principles such as Probabilism, Conditionalization, the Principal Principle, and the Principle of Indifference.[1] Let us call such a strategy *Cognitive Decision Theory* (*CDT* for short).[2] A rational doxastic state (or a rational updating policy) to adopt, according to CDT, is one with the greatest expected *epistemic* utility—one such that your expected *epistemic* utility for how states of the world will turn out, given your adopting it, is at least as high as that of any alternative doxastic state you might adopt.[3]

One of the central issues in epistemology is whether epistemic rationality is permissive or not: Some claim that (*Uniqueness*) for any total evidence, there is a unique doxastic state that

[1] For instance, by resorting to a decision theoretic approach, Joyce (1998, 2009), Leitgeb and Pettigrew (2010a, 2010b) and Pettigrew (2016) attempt to justify Probabilism; Easwaran (2013), Greaves and Wallace (2006), Leitgeb and Pettigrew (2010a, 2010b), and Pettigrew (2016) attempt to justify Conditionalization; Pettigrew (2013) attempts to justify the Principal Principle; Pettigrew (2014) attempts to justify the Principle of Indifference.

[2] I am following Greaves and Wallace (2006) in the use of *Cognitive Decision Theory*. Some philosophers (for instance, Richard Pettigrew) calls it *epistemic utility theory*.

[3] In section 2, I will explain expected epistemic utilities in detail. 'A rational doxastic state' rather than 'the rational doxastic state' to allow two or more of the possible doxastic states to have the same greatest expected epistemic utility.

any agent with that total evidence should take[4]; others claim that (*Permissivism*) for some total evidence, there are multiple doxastic states that an agent with that total evidence can take.[5]

How does CDT relate to the debate over permissive rationality? The aim of this paper is to explore this question. In particular, as one way of addressing this question, I will provide and assess an argument against CDT: I will assume that an epistemically rational agent always adopts a doxastic state that *maximizes* her expected *epistemic* utility, and prove that, at least in some possible cases where *Non-Strict Epistemic Immodesty* (described in section 3) holds, the agent who is faithfully represented by CDT can change her doxastic state, absent any new evidence. This violates *Epistemic Conservatism* (described in section 5), which says that one's doxastic states should remain the same, absent any new evidence. This seems to be an unfortunate consequence. However, I will further show that when we clearly distinguish among several versions of Permissivism/Uniqueness, the argument is not a real threat to any cognitive decision theorist (CDTer for short), given that each version of Permissivism/Uniqueness is a viable epistemic principle. Depending on which version of Permissivism/Uniqueness CDTers endorse, they may avoid the argument in one of two general ways. One response appeals to the stability of beliefs over time, while the other allows that the instability of beliefs over time fits naturally with epistemic rationality.

I will proceed as follows. Section 2 introduces the framework for CDT. Section 3, with the framework in hand, explains *Expected Epistemic Utility Maximization* and *Epistemic Immodesty* as general norms of CDT that I will assume throughout our discussion. In particular, I will make a clear distinction between *Strict (Epistemic) Immodesty* and *Non-Strict (Epistemic)*

[4] For instance, see Feldman (2007); White (2005, 2014); Christensen (2007); Levinstein (2015).
[5] For instance, see Kelly(2014); Schoenfield (2014); Meacham (2014).

Immodesty. Section 4 provides a problematic example where an agent endorses a *non*-strictly immodest epistemic utility function and, based on the problematic example, I will offer an argument against CDT. Section 5 discusses two available responses to the argument against CDT, those involving the rejection of Non-Strict Immodesty as a rational constraint, and those involving the rejection of Epistemic Conservatism as a rational constraint. After clearly defining three types of Permissivism (what I call *Permissivism*$_1$, *Permissivism*$_2$, and *Permissivism*$_3$, respectively), I will show that CDTers who endorse Uniqueness or Permissivism$_3$ would reject the argument against CDT because they would reject Non-Strict Immodesty in favor of Strict Immodesty; in contrast, CDTers who endorse Permissivism$_1$ or Permissivism$_2$ would *generally* reject the argument against CDT because they would *generally* reject Epistemic Conservatism in favor of Non-Strict Immodesty. In particular, even though Non-Strict Immodesty is not much defended in the recent epistemic utility literature, I will explain why it is as defensible as Strict Immodesty. Finally, section 6 is the conclusion with some brief remarks.

2 Cognitive Decision Theory

CDT provides a framework for determining one's rational doxastic states. The framework for the theory includes possible states of the world, degree of belief (or credence) functions, epistemic utility functions, and expected epistemic utilities. Let me briefly explicate each of these notions below.

2.1 States of the World

There is a set $\boldsymbol{S}$ of mutually exclusive and jointly exhaustive possible states of the world. An agent has a doxastic state over the power set of $\boldsymbol{S}$, which can be thought of as a set of propositions.[6, 7] The specificity of $\boldsymbol{S}$ will determine how fine-grained our distinctions among doxastic states can be.

2.2 Degree of Belief Functions

It is assumed that, in CDT, an agent's doxastic state can be represented by some degree of belief function *c* from the power set of $\boldsymbol{S}$ to the real numbers $\boldsymbol{R}$.[8] Throughout, let $\boldsymbol{C_S}$ be a set of degree of belief functions from the power set of $\boldsymbol{S}$ to to the real numbers $\boldsymbol{R}$.[9]

2.3 Epistemic Utility Functions

Given a state of the world and an action, a *practical utility* is a *practical desirability* of the outcome of performing that action when that state of the world in fact obtains. An *epistemic utility* is the *epistemic* counterpart of that practical utility: Given a state of the world and a doxastic state, an *epistemic* utility is a *purely epistemic desirability* of adopting that doxastic state when that state of the world in fact obtains. An epistemic utility is concerned with epistemically desirable values, which, many think, include *accuracy*, *informativeness*, *simplicity*, and

---

[6] What I aim to show does not depend on whether $\boldsymbol{S}$ is finite or infinite. For simplicity, however, we will assume that $\boldsymbol{S}$ is finite. Nothing will hinge on this restriction.

[7] I use the set-theoretic notation and syntactic notation interchangeably throughout, depending on which seems more stylistically convenient.

[8] Here we do not have to assume that the agent's doxastic state should always be modeled by a *single* degree of belief function. For instance, when the agent's total evidence is *unspecific*, her doxastic state might be represented by a *set* of degree of belief functions. Many different ways of representing doxastic states could be allowed–so long as there is some way of (precisely or imprecisely) quantifying expected epistemic utilities to maximize them and, in any evidential situation, a single degree of belief function is allowed (explained in section 5). For simplicity, however, we will assume that the agent's doxastic states are modeled by a *single* degree of belief function.

[9] I do not assume Probabilism that says that rational degrees of belief functions are probabilistically coherent. Thus $\boldsymbol{C_S}$ may contain degrees of belief functions that are probabilistically incoherent.

verisimilitude. It would require a substantial philosophical investigation to give a full account of the epistemic values, and there could be various views on what the epistemic values are. However, what I aim to show is compatible with *any* view on the epistemic values—so long as there is a way of quantifying epistemic utilities.

It is assumed that, in CDT, an epistemic utility can be represented by some epistemic utility function u from pairs in $\boldsymbol{C_S} \times \boldsymbol{S}$ to the real numbers $\boldsymbol{R}$.[10] For instance, $u(c, s)$ refers to some real number that represents the epistemic utility of the doxastic state c when s in fact obtains; $u(c_1, s_1) > u(c_2, s_1)$ means that adopting c_1 is strictly better than adopting c_2 when s_1 in fact obtains; and $u(c_1, s_2) = u(c_2, s_2)$ means that adopting c_1 is no better than adopting c_2, and vice versa, when s_2 in fact obtains, from the purely *epistemic* perspective.

There are some conditions that constrain epistemic utility functions. For instance, many believe that an epistemic utility function should satisfy *Extensionality*, which says that the epistemic utility is a function of nothing other than the truth values of $s \in \boldsymbol{S}$ and degrees of belief that $c \in \boldsymbol{C_S}$ assigns to the members of the power set of $\boldsymbol{S}$.[11] Extensionality is not necessarily required for my purposes,[12] but the following two conditions are required:

---

[10] Note that $u(c, s)$ measures how good or bad the overall degrees of belief function c is when s in fact obtains. u is similar with what Leitgeb and Pettigrew (2010 a) call a *global inaccuracy measure*. There are many different epistemic utility functions, and my arguments do not rely on the use of any particular one. Moreover, we do not have to assume that an epistemic utility of a doxastic state should always be modeled by a *single* utility function. For instance, if, given the doxastic state, there are pairs of epistemic values that are *incommensurable*, the epistemic utility of the epistemic state might be represented by the *set* of epistemic utility functions that indicates a range of *possible* epistemic utilities for the doxastic state. Many different ways of representing epistemic utilities could be allowed–so long as there is some way of (precisely or imprecisely) quantifying expected epistemic utilities to maximize them. For simplicity, however, we will assume that the agent's epistemic utilities are modeled by a *single* epistemic utility function.

[11] See Joyce (1998: 591).

[12] For instance, if the epistemic utility is in part based on informativeness, it cannot be extensional because informativeness of a doxastic state is not extensional. However, what I aim to show applies to any version of CDT in which Extensionality holds as well.

Quantifiability: All epistemic utilities can be numerically measured.

Externalism: An epistemic utility of a doxastic state depends, at least in part, on features external to the doxastic state.

I do not have space here to give a full account of *Quantifiability* and *Externalism*, but for my purposes in this paper, it is enough to assume them rather to justify them.

2.4 Expected Epistemic Utilities

An expected *epistemic* utility of adopting a doxastic state $x \in \boldsymbol{C_S}$ with respect to $c \in \boldsymbol{C_S}$ is given by weighting x's epistemic utility at each state of the world $s \in \boldsymbol{S}$ by the degree of belief that c assigns to that state of the world, and summing: $EU_c(x) = \sum_{s \in \boldsymbol{S}} c(s)u(x, s)$. An agent's expected *epistemic* utility of a doxastic state $x \in \boldsymbol{C_S}$ is an *estimation* of x's *epistemic* utility from the perspective of the agent's current doxastic state c.

3 Two General Norms for Cognitive Decision Theory

Among the general norms that may dictate how epistemic utilities constrain the set of rational doxastic states, two are relevant for my purposes: *Expected Epistemic Utility Maximization* and *Epistemic Immodesty*.

3.1 Expected Epistemic Utility Maximization

It is *very intuitive* to say: An agent ought to adopt a doxastic state that has the highest expected *epistemic* utility of all possible doxastic states with respect to the agent's current doxastic state. More formally, with the framework for CDT in hand, where $\boldsymbol{M_c}$ is the set of degree of belief functions that maximize epistemic utility with respect to the agent's current doxastic state c, the norm that I have in mind holds:

> **Expected Epistemic Utility Maximization**: Given an agent's degree of belief function $c \in \boldsymbol{C_S}$, let $\boldsymbol{M_c} \subseteq \boldsymbol{C_S}$ be such that, for all $x, y \in \boldsymbol{M_c}$ and for all $z \in \boldsymbol{C_S} \sim \boldsymbol{M_c}$, $EU_c(x) = EU_c(y)$ and $EU_c(x) > EU_c(z)$. The agent *should* adopt a member of $\boldsymbol{M_c}$ and *could* adopt *any* member of $\boldsymbol{M_c}$.[13]

Note that, given Expected Epistemic Utility Maximization (Maximization for short), when $\boldsymbol{M_c}$ is a *non-empty* and *non-singleton* set, an agent who holds a degree of belief function c would satisfy it, as long as the agent adopts *any* member of $\boldsymbol{M_c}$.

3.2 Epistemic Immodesty

It is also *very intuitive* to say: If an agent adopts a doxastic state x, the agent should expect x to be a *best* one from the perspective of her doxastic state x. More formally, with the framework for CDT in hand, the norm that I have in mind holds:

[13] The notation $\boldsymbol{C_S} \sim \boldsymbol{M_C}$ denotes the set of elements of $\boldsymbol{C_S}$ that are not in $\boldsymbol{M_C}$.

Epistemic Immodesty: For all $x \in \boldsymbol{C_S}$, if a rational agent's current degree of belief function is x, then the agent should take her own degrees of belief, x, to maximize expected epistemic utility. In other words, $\boldsymbol{M_x}$ contains x.

What is the difference between Epistemic Immodesty and Maximization? Note that, for any $x \in \boldsymbol{C_S}$, when an agent's degrees of belief function is x, Epistemic Immodesty constrains the membership of M_x, while Maximization says that the agent should adopt a member of M_x and could adopt any member of M_x. Thus, Epistemic Immodesty does not imply Maximization. To see this, assume that Epistemic Immodesty holds. Then, if an agent's degrees of belief function is x, $x \in \boldsymbol{M_x}$. Suppose further that $x \neq y$ and $y \notin \boldsymbol{M_x}$. Without independently assuming Maximization, the agent may rationally adopt y. Note that as long as $y \in \boldsymbol{M_y}$, Epistemic Immodesty still holds.

Maximization does not imply Epistemic Immodesty either. To see this, assume that Maximization holds. Then, if an agent's degrees of belief function is x, the agent should adopt a member of M_x. Suppose further that the agent is not epistemically immodest.[14] That is, $x \notin \boldsymbol{M_x}$. Note that, given $x \neq y$ and $y \in \boldsymbol{M_x}$, as long as the agent newly adopts y, Maximization still holds.

Let us call the denial of Epistemic Immodesty *Epistemic Modesty*. Thus, given *Epistemic Modesty*, for some $x \in \boldsymbol{C_S}$, $\boldsymbol{M_x}$ does not contain x. Interestingly, Christensen (2013), Elga (2013), and Lasonen-Aarnio (2014) argue that sorts of epistemic modesty are compatible with epistemic rationality. However, I do not think that their versions of epistemic modesty are exactly same with what I take to be *Epistemic Modesty* here. For instance, according to Christensen (2013:

[14] That is, the agent holds a modest epistemic utility function.

90), versions of epistemic modesty that he has in mind "allow evidence that I've made an epistemic mistake in thinking about *P* to affect the degree of confidence it's rational for me to have to have in *P*." Such versions of epistemic modesty, I think, are compatible with what I take to be *Epistemic Immodesty* here, because it is possible that, out of all *fallible* doxastic states that an agent could adopt, she currently has a best one from the perspective of her current *fallible* doxastic state itself. That is, even if her current doxastic state is *fallible*, for every $x \in \boldsymbol{C}_{\boldsymbol{S}}$, it is possible for her to estimate *x*'s *epistemic* utility from her current (*fallible*) doxastic state and take her current doxastic state to be a best one out of all possible (*fallible*) doxastic states that she could adopt.

In this paper, I will assume that, on the version of Epistemic Immodesty that I have in mind, *Epistemic Modesty* is *not* rational. I think that this assumption is very plausible because when an agent adopts a modest doxastic state, as Joyce (2009: 277) points out,

> "She has a prima facie epistemic reason, grounded in her beliefs, to think that she should not be relying on those very beliefs. This is a probabilistic version of Moore's paradox. Just as a rational person cannot fully believe '*X* but I don't believe *X*,' so a person cannot rationally hold a set of credences that require her to estimate that some other set has higher epistemic utility. The modest person is always in this pathological position: her beliefs undermine themselves."

It is noteworthy that, regarding Epistemic Immodesty, we can make a distinction as follows:

Strict Immodesty: For all $x \in \boldsymbol{C_S}$, if a rational agent's current credence function is x, then the agent should take her own degree of belief function to *uniquely* maximize expected epistemic utility. That is, $\boldsymbol{M_x}$ contains *only* x.

Non-Strict Immodesty: For all $x \in \boldsymbol{C_S}$, if a rational agent's current credence function is x, then the agent should take her own degree of belief function to be a member of $\boldsymbol{M_x}$ that *possibly* contains other degree of belief functions too. That is, $\boldsymbol{M_x}$ contains x and *possibly* other degrees of belief functions too.

It is noteworthy that, given Strict Immodesty, for all $x \in \boldsymbol{C_S}$, when an agent's current credence function is x, for any $y \in \boldsymbol{C_S}$ $(x \neq y)$, $EU_x(x) > EU_x(y)$. On the other hand, given Non-Strict Immodesty, for all $x \in \boldsymbol{C_S}$, when an agent's current credence function is x, for any $y \in \boldsymbol{C_S}$ $(x \neq y)$, $EU_x(x) \geq EU_x(y)$.

Many philosophers endorse Strict Immodesty rather than Non-Strict Immodesty.[15] Non-Strict Immodesty is not as much defended as Strict Immodesty in the epistemic utility literature. A few philosophers defend Non-Strict Immodesty.[16] However, in section 5.2, I will show that Non-Strict Immodesty is as defensible as Strict Immodesty, when we leave open possibilities of adopting each type of permissive rationality.

4 An Argument against *Cognitive Decision Theory*

The result that we want to establish in this section is:

[15] For example, see Lewis (1971); Oddie (1997); Joyce (2009); Gibbard (2007); Greaves and Wallace (2006); Easwaran (2013).

[16] For example, see Mayo-Wilson and Wheeler (2016). (They call Non-Strict Immodesty *Mild Immodesty*.)

Given Non-Strict Immodesty and Epistemic Conservatism, CDT is not the correct theory of epistemic rationality.

Note that this result can apply to any decision theoretic approach to epistemic rationality, which assumes *Quantifiability, Externalism, Expected Epistemic Utility Maximization, and Epistemic (non-strict) Immodesty, Epistemic Conservatism*: It does not depend on any particular epistemic utility functions, so long as they satisfy these constraints. Before proceeding further, let me briefly explain what I mean by *Epistemic Conservatism*.

4.1 Epistemic Conservatism

Suppose that you are an astrophysicist, and has of course a fairly extensive body of degrees of belief about the multiverse that you deem best at each time. Suppose further that your degree of belief in *M* (the proposition that the multiverse exists) is 0.8 at t_1 and you learn nothing between t_1 and t_2. Then how *should* you change your degree of belief in *M* at t_2? Many would take it to be very intuitive that your degree of belief in *M* should remain the same at t_2. On this way of thinking, the following principle seems to be assumed:

> ***Epistemic Conservatism***: When a rational agent undergoes *no* learning experience between t_i and t_j ($t_i < t_j$), her degree of belief in any proposition should remain the same during that time.

There are of course a number of different versions of Epistemic Conservatism (Conservatism for short).[17] For instance, the strongest version of Conservatism implies that the mere fact that a proposition is believed provides it with some epistemic justification.[18] As we can clearly see, the version of Conservatism that I examine in this paper is much weaker than the strongest version. According to Conservatism that I have in mind, one is obliged to *continue* to have credences in a proposition in the absence of learning experience.[19] Thus, as opposed to the strongest version, it does not imply that the mere fact that one has a particular degree of belief in a proposition provides some epistemic justification to it.

Note also that the version of Conservatism that I examine in this paper applies only to cases in which an agent undergoes *no* learning experience. So, my version is weaker than the other versions of Conservatism that apply to cases in which an agent undergoes a learning experience that is *evidentially irrelevant* to a proposition in question.

4.2 Argument I

Now let us consider the following argument:

Argument I: an Argument against CDT

P1: Non-Strict Immodesty allows a *rational* agent to hold a non-strictly immodest epistemic utility function.

[17] See Vahid (2004) for various kinds of Epistemic Conservatism.
[18] For instance, see Chisholm (1980).
[19] See Harman (1986) for the similar version of Conservatism.

P2: Given Maximization, for any agent who holds a *non*-strictly immodest epistemic utility function, doxastic state shifts from one degree of belief function to another may be rational, even in the absence of new evidence.

C1: Given Maximization, rational agents are allowed, in the absence of new evidence, to rationally make doxastic state shifts from one degree of belief function to another. (From P1, P2)

P3: Given Epistemic Conservatism, in the absence of new evidence, a rational agent's degree of belief in any proposition should remain the same.

C2: Therefore, CDT is not the correct theory of epistemic rationality.[20] (From C1, P3)

Argument I is valid: If we accept P1, P2, and P3, we should also accept C2.[21] P2 is uncontroversial. For any agent who holds a non-strictly immodest epistemic utility function, when the agent currently holds degree of belief function $x \in \boldsymbol{C_S}$, it is *possible* that there is another distinct degree of belief function $y \in \boldsymbol{C_s}$ such that $x, y \in \boldsymbol{M_x}$, and thus $EU_x(x) = EU_x(y)$. That is, when Non-Strict Immodesty holds, it is *possible* that adopting x and adopting y have *equal* and *optimal* expected epistemic utility from the perspective the agent in the doxastic state x itself. Given Maximization, as Greaves and Wallace point out, "when this occurs, the agent *can* stick to his current [degree of belief function x], but it will be equally consistent with ideal rationality if he chooses to move to [degree of belief function y] on a whim."[22] Note that

[20] This argument is basically the same as one given by Greaves and Wallace (2006: 630). (As Greaves and Wallace point out, Maher (1993: 179) suggests the original argument.) However, even though they seem to see the possibility of applying their argument to *non-strictly immodest* epistemic utility functions (see their footnote 13 (p. 629)), they apply their argument only to *modest* epistemic utility functions.

[21] Note that CDT recommends Maximization.

[22] Greaves and Wallace (2006: 621). The brackets are mine. See Oddie (1997: 537) for a similar point.

Non-Strict Immodesty itself does not recommend the agent who currently holds degree of belief function x to move to another degree of belief function y. However, given Maximization, when x, $y \in \boldsymbol{M}_x$, moving to another degree of belief function y is *as epistemically rational as* sticking to the current degree of belief function x. Since P2 holds, P1 and P3 are the argument's only questionable steps.

5 Responses

There are only two options to reject the argument: those involving the rejection of Non-Strict Immodesty as a rational constraint, and those involving the rejection of Epistemic Conservatism as a rational constraint. Let us first consider the former.

5.1 Non-Strict Immodesty is not rational (Rejection of P1)

As already pointed out, Non-Strict Immodesty allows a *rational* agent to hold a non-strictly immodest epistemic utility function. Is Non-Strict Immodesty a rational constraint? Whether we should accept Non-Strict Immodesty (P1) or not crucially depends on whether there are some possible cases in which a *rational* agent regards two or more distinct doxastic states as having *equal* and *optimal* expected epistemic utility. In fact, it will turn out that there are very intricate connections between the debate over *Strict/Non-Strict Immodesty* and the debate over *Permissivism/Uniqueness*. To see this, let me first briefly explain what *Permissivism* and *Uniqueness* are.

5.1.1 Permissivism and Uniqueness

Permissivism says that for some evidence *E*, there are multiple doxastic states, any one of which a possible agent with that total evidence *E* can rationally take. Uniqueness is the *denial* of Permissivism. As Meacham (2014: 1188) points out, Uniqueness is equivalent to the conjunction of the following two claims:

(1) (Agent Uniqueness) For any possible agent with a total evidence *E*, there is only one permissible degree of belief function for that agent.

(2) (Permission Parity) The same degree of belief functions are permissible for all possible agents who share a total evidence *E*.[23, 24]

Thus, following Meacham (2014), we can distinguish three types of *Permissivism*.

Permissivism$_1$: One that rejects both (1) and (2)

Permissivism$_2$: One that rejects (1) but accepts (2)

Permissivism$_3$: One that accepts (1) but rejects (2)

Permissivism$_1$, Permissivism$_2$, Permissivism$_3$, and Uniqueness are mutually exclusive and collectively exhaustive. Note that Permissivism$_1$ and Permissivism$_2$ reject Agent Uniqueness,

[23] I am following Meacham (2014) in the use of Uniqueness and Permissivism. In particular, *Uniqueness* is what Meacham (2014: 1187) calls *Evidential Uniqueness*, which implies that the evidence *alone* suffices to fix what a rational credal state is. Someone might think that *Evidential Uniqueness* is compatible with the possibility that there is more than one rational stance for a given body of evidence, because there are, for instance, other epistemic standards such as reliability that affect epistemic rationality. However, such a view, I think, already assumes Evidential Permissivism. In this paper, I do not have such an assumption. The aim of this paper is to show how CDT relates to Evidential Uniqueness and Evidential Permissivism.

[24] For the proof of the equivalence, see Meacham (2014: 1188).

while Permissivism$_3$ accepts it. Thus, even though Permissivism$_1$, Permissivism$_2$, and Permissivism$_3$ are all permissive principles, there is a crucial difference between Permissivism$_1$ (or Permissivism$_2$) and Permissivism$_3$: Permissivism$_1$ and Permissivism$_2$ appeal to the following two permissive intuitions:

> Permissive Intuition 1: There are evidential situations in which two different agents can rationally adopt different beliefs.
>
> Permissive Intuition 2: There are evidential situations in which a particular agent can rationally adopt a range of different beliefs.[25]

In contrast, Permissivism$_3$ appeals *only* to Permissive Intuition 1. That is, while Permissivism$_1$ and Permissivism$_2$ have *interpersonal* permissive import and *intrapersonal* permissive import, Permissivism$_3$ has *interpersonal* permissive import and *intrapersonal im*permissive import.[26] In order to clearly distinguish versions of Permissivism that have *intrapersonal* permissive import from one that lacks such import, let us call Permissivism$_1$ (or Permissivism$_2$) and Permissivism$_3$ *Intrapersonal Permissivism* and *Interpersonal Permissivism*, respectively.[27]

In this paper, my purpose is to show how these two types of Permissivism (and Uniqueness) relate to CDT. Some philosophers might think that, in contrast to Interpersonal

[25] The distinction between two permissive intuitions is from Meacham (2014: 1190).
[26] The distinction between *interpersonal* import and *intrapersonal* import is from Weintraub (2013) and Kelly (2014).
[27] *Intrapersonal Permissivism* has interpersonal import as well because it appeals to both Permissive Intuition 1 and 2, but, in this paper, *Interpersonal Permissivism* refers *only* to Permissivism$_3$ while *Intrapersonal Permissivism* refers to Permissivism$_1$ or Permissivism$_2$.

Permissivism, Intrapersonal Permissivism is *prima facie* implausible.[28] Thus, they may think, any claim about the relation between Intrapersonal Permissivism and CDT is not very interesting unless we independently show how Intrapersonal Permissivism fits well with pre-theoretical intuition. However, I think, there are at least some evidential situations in which Intrapersonal Permissivism is sufficiently plausible.[29] Given certain kinds of evidence, I think, Intrapersonal Permissivism is well motivated from an evidential perspective. To illustrate, let us compare the following two cases:

> You are asked to represent your degree of belief that (*N*) car A will beat car B in the Formula One: in the first instance, the only evidence you have that is relevant to *N* is that the objective chance of *N* is between 0.1 and 0.9. Contrast this with the second situation, in which you know the chance of *N* is exactly 0.75.

In the second situation, given Lewis (1980)'s Principal Principle, according to which, one's degrees of belief in propositions about objective chances should constrain her degrees of belief in other propositions[30], your degree of belief in *N* should be 0.75. What about the first situation? Suppose in the first situation, the evidence allows you to adopt a precise degree of belief in *N*. It is intuitive that your precise degree of belief in *N* should be between 0.1 and 0.9. Suppose further

---

[28] As par as I know, Intrapersonal Permissivism is not much defended in the recent permissive rationality literature. However, Kelly (2014: 299-301) briefly explains how Intrapersonal Permissivism could be defended. What I will provide below is basically similar with Kelly's idea.

[29] To be clear, my aim here is not to propose a knock-down argument for Intrapersonal Permissivism. Instead, my purpose here is to provide an evidential situation in which Intrapersonal Permissivism is plausible, and thus to show that it is still philosophically interesting to consider how Intrapersonal Permissivism relates to CDT.

[30] More formally, if *E* is a rational agent's total (admissible) evidence at t, for any proposition *X*, $c(X \mid E \,\&\, \text{chance}_t(X) = x) = x$, given $c(E \,\&\, \text{chance}_t(X) = x) > 0$, where *c* is her (initial) degrees of belief function and $\text{chance}_t(X)$ is an objective chance of *X* at t.

that in the second situation, Interpersonal Permissivism (or Uniqueness) holds. Then your unspecific and incomplete evidence would single out a unique precise degree of belief as the maximally rational option to you. It is noteworthy that in the first situation, as opposed to the second situation, the evidence itself does not privilege a particular degree of belief in *N* above any other between 0.1 and 0.9. However, on the assumption of Interpersonal Permissivism (or Uniqueness), the evidence itself requires you to adopt a particular degree of belief in *N* that is *uniquely* rational to you. Thus, in the first situation, Interpersonal Permissivism seems not to be well motivated from an evidential perspective, because it requires you to adopt the particular doxastic state that must go beyond what your evidence objectively supports.

What about Intrapersonal Permissivism? In contrast to Interpersonal Permissivism, on the assumption of Intrapersonal Permissivism, the evidence itself does *not* require you to adopt a particular degree of belief that is *uniquely* rational to you. On the assumption of Intrapersonal Permissivism, even if the evidence also requires you to have a single degree of belief in *N*, in contrast to Interpersonal Permissivism, your degree of belief in *N* could be rationally on a par with any other in the range that open to you. Thus, even after you end up in a particular doxastic state, you could still rationally think that your own doxastic state is *not uniquely* rational within the range. In the first situation, I think, so long as you end up in some precise doxastic state within the range, you must think that your own doxastic state is on a par with any other in the range. Thus, regarding the first situation, Intrapersonal Permissivism does provide a more proper response to the evidence than Interpersonal Permissivism does, because, given Intrapersonal Permissivism, in contrast to Interpersonal Permissivism, you are *not* required to have a precise doxastic state that has to go beyond what your evidence objectively supports.

Of course, what I claimed above holds only if, in the first situation, you could rationally have a *precise* degree of belief as a proper response to the unspecific and incomplete evidence. And, as well known, some philosophers would claim that, in the first situation, you *should* have an imprecise degree of belief in *N* rather than a precise one, because your total evidence is very unspecific and incomplete. That is, according to them, when your total evidence is unspecific and incomplete, your doxastic state should be represented by a *set* of degree of belief functions rather than by a *single* degree of belief function.[31] If such an imprecise view is correct, you would not be fully (epistemically) rational, so long as you have a single precise degree of belief in *N*. However, whether unspecific and incomplete evidence warrants *only* imprecise degrees of belief is the subject of intense debate.[32] I do not pretend to have complete answers to this issue. But, from an evidential point of view, the supporter of imprecise degrees of belief does not seem to have any good reason that favors the imprecise view over the intrapersonally permissive precise view. The main reason for the imprecise view, I think, is that, the unspecific and incomplete evidence does not simply single out a single degree of belief function over the others. That is, such evidence does not privilege a single degree of belief function above the others. And, as I pointed out above, it is noteworthy that the intrapersonally permissive precise view does not simply single out a precise degree of belief function in such a way: On the assumption of Intrapersonal Permissivism, the unspecific and incomplete evidence requires you to have a single degree of belief function that could be rationally on a par with any other in the range that open to you. Thus, I think, we can at least say that the evidential considerations that motivate the imprecise view does not rule out the possibility of the intrapersonally precise view. Given the

[31] For instance, see Joyce (2010).
[32] For instance, see Schoenfield (2015)

unspecific and incomplete evidence, on the assumption of Intrapersonal Permissivism, precise degrees of belief are at least prima facie plausible option.

5.1.2 Strict/Non-Strict Immodesty and Permissive/Impermissive Rationality

Interestingly, there are very intricate connections between *Strict/Non-Strict Immodesty* and *Permissivism/Uniqueness*. We can prove the followings:

> (I) *Intrapersonal Permissivism* entails *Non-Strict Immodesty* or *the denial of Maximization*.
>
> (II) *Uniqueness* or *Interpersonal Permissivism* entails *Strict Immodesty* or *the denial of Maximization*.

To see this, let me start by proving (I). We can show that Strict Immodesty and Maximization entail Uniqueness or Interpersonal Permissivism. Note that Strict Immodesty says that, for all $x \in \boldsymbol{C}_S$, if a rational agent's current degree of belief function is x, then $\boldsymbol{M}_x = \{x\}$. Thus, assuming Strict Immodesty and Maximization, given a total evidence E, if an agent, α, who holds a strictly immodest epistemic utility function has degree of belief function x, there is a unique rational epistemic state, x, that α with the total evidence E *should* take. And, assuming Strict Immodesty and Maximization, given the same total evidence E, if another agent, β, who also holds a strictly immodest epistemic utility function has degree of belief function y, there is a unique rational epistemic state, y, that β with the total evidence E *should* take. It is a logical truth that, for every E, $x = y$, or, for some E, $x \neq y$. If, for every E, $x = y$, Uniqueness follows. On the other hand, if,

for some E, $x \neq y$, Interpersonal Permissivism follows. Thus we have shown that Strict Immodesty and Maximization entail Uniqueness or Interpersonal Permissivism.[33] Note that Strict Immodesty alone does not entail Uniqueness or Intrapersonal Permissivism, because, as already pointed out in section 3.2, given an agent's degrees of belief function is x, Epistemic Immodesty just constrains the membership of M_x. Without independently assuming Maximization that says that an agent whose degrees of belief function is x *should* adopt a member of M_x and *could* adopt any member of M_x, an agent who holds a strictly immodest epistemic utility function may rationally adopt a doxastic state that is not a member of M_x. Assuming that *only* immodest epistemic utility functions are rational, from the fact that Strict Immodesty and Maximization entail Uniqueness or Interpersonal Permissivism, it follows by contraposition that Intrapersonal Permissivism (Permissivism$_1$ or Permissivism$_2$) entails Non-Strict Immodesty or the denial of Maximization.[34]

Someone may think that (I) is *false* because it is *possible* to hold Intrapersonal Permissivism, Strict Immodesty, and Maximization. However, in contrast to Interpersonal Permissivism, Strict Immodesty and Maximization are not compatible with Intrapersonal Permissivism. To illustrate, suppose that there is a *particular* agent (say John) whose total evidence is E. Suppose further that Strict Immodesty and Maximization hold in John's case. Thus, if John responds to E by adopting x, then $\boldsymbol{M_x} = \{x\}$. And if John responds to E by adopting y, then $\boldsymbol{M_y} = \{y\}$. (Assume that John adopts either x or y.) Of course, in order for x and y both to be rationally permissible for John, it is not required that both x and y are in $\boldsymbol{M_x}$ or in $\boldsymbol{M_y}$. To see

[33] Note that the reverse does not hold because, for instance, given the denial of Maximization, Non-Strict Immodesty is compatible with Uniqueness and Interpersonal Permissivism.
[34] Note that Intrapersonal Permissivism, Interpersonal Permissivism, and Uniqueness are mutually exclusive and collectively exhaustive.

this, note that, from a purely epistemic point of view, John who adopts x (say John_x) is not identical with John who adopts y (say John_y) because they have different doxastic states, respectively. (So John is either John_x or John_y (not both).) Given the total evidence E is interpersonally permissive, John_x and John_y could rationally have different degrees of belief functions, x and y, that are uniquely permissible for John_x and John_y, respectively. That is, Strict Immodesty and Maximization are clearly compatible with *Interpersonal* Permissivism. However, given Strict Immodesty and Maximization, any total evidence is not permissive with respect to the *range* of rational doxastic states open to any particular individual such as John. On the assumption of Strict Immodesty and Maximization, John_x (John_y) cannot rationally adopt y (x) because $\boldsymbol{M}_x = \{x\}$ ($\boldsymbol{M}_y = \{y\}$). (Note that John is either John_x or John_y (not both).) Thus, *Intrapersonal* Permissivism is not compatible with Strict Immodesty and Maximization.

Similarly, we can easily prove (II). We can show that Non-Strict Immodesty and Maximization entail Intrapersonal Permissivism.[35] Non-Strict Immodesty says that, for all $x \in \boldsymbol{C}_S$, if a rational agent's current degree of belief function is x, then $x \in \boldsymbol{M}_x$ that is *possibly* a non-singleton. Thus, assuming Non-Strict Immodesty and Maximization, given a total evidence E, if an agent, γ, who holds a non-strictly immodest epistemic utility function has degree of belief function x, there may be a set of multiple rational doxastic states, $\boldsymbol{M}_x$, anyone of which γ with the total evidence E *could* rationally take. And, assuming Non-Strict Immodesty and Maximization, given the same total evidence E, if another agent, δ, who also holds a non-strictly immodest epistemic utility function has degree of belief function y, there may be a set of multiple rational doxastic states, $\boldsymbol{M}_y$, anyone of which δ with the total evidence E *could* rationally

[35] The reverse does not hold either because, for instance, given the denial of Maximization, Strict Immodesty is compatible with Intrapersonal Permissivism.

take. It is a logical truth that, for every E, $\boldsymbol{M}_x = \boldsymbol{M}_y$, or, for some E, $\boldsymbol{M}_x \neq \boldsymbol{M}_y$. If, for every E, $\boldsymbol{M}_x = \boldsymbol{M}_y$, Permissivism$_2$ follows. On the other hand, if, for some E, $\boldsymbol{M}_x \neq \boldsymbol{M}_y$, Permissivism$_1$ follows. Thus, assuming that *only* immodest epistemic utility functions are rational, it follows by contraposition that Uniqueness or Interpersonal Permissivism (Permissivism$_3$) entails Strict Immodesty or the denial of Maximization.

5.1.3 Argument I and Permissive/Impermissive Rationality

Now let us return to Argument I. To CDTers who endorse Interpersonal Permissivism or Uniqueness, Argument I is not a real threat to CDT because they would reject P1 in favor of Strict Immodesty. For instance, they can provide the following argument for Strict Immodesty:

Argument II: an Argument for Strict Immodesty[36]

O1: Epistemic Immodesty (Strict Immodesty or Non-Strict Immodesty)

O2: Non-Strict Immodesty

O3: Maximization

C'1: the denial of Conservatism (From O2 and O3)

O4: Epistemic Conservatism

C'2: the denial of O2 (From C'1 and O4)

C'3: Therefore, Strict Immodesty (From C2' and O1)

[36] This argument is from Oddie (1997: 535-8).

Note that, as I explained in section 4, given Maximization, Non-Strict Immodesty entails the denial of Epistemic Conservatism. Thus, assuming Epistemic Conservatism, in order to avoid contradiction, we should reject either O2 or O3. CDTers would reject O2, and hence C'3 follows, because CDT recommends Maximization.[37] CDTers who accept Interpersonal Permissivism or Uniqueness would take this strategy and thus reject P1 of Argument I.[38]

However, what about CDTers who endorse Intrapersonal Permissivism? As shown above, given Intrapersonal Permissivism, *Non-Strict Immodesty* or the denial of *Maximization* follows, and hence, to CDTers who endorse Intrapersonal Permissivism, the conclusion of Argument I (C2) appears to follow: If they accept Non-Strict Immodesty (P1), on the assumption of Epistemic Conservatism, C2 follows; if they reject Maximization, C2 also immediately follows because CDT recommends Maximization. Then, is Argument I a real threat to those permissive CDTers? As I will show below, it needs not be, because, on the assumption of Intrapersonal Permissivism, they could *generally* reject Conservatism. That is, they could take the remaining option to reject Argument I: those involving the rejection of Epistemic Conservatism as a rational constraint. We turn now to it.

5.2 Non-Conservatives are sometimes rational (Rejection of P3)

Is there any good *epistemic* reason to accept Conservatism? In particular, should *any* CDTer accept Conservatism? As I will explain below, it seems not.

---

[37] See Oddie (1997: 537).

[38] Someone may claim that it is completely *ad hoc* to reject O2 rather than O3, because Non-Strict Immodesty could be as plausible as Maximization. For instance, see Maher (1993: 179). Without already assuming CDT that recommends Maximization, I think, it might be *ad hoc* to reject O2 rather than O3, because Uniqueness or Interpersonal Permissivism entails Strict Immodesty or the denial of Maximization.

5.2.1 Conservatism and Permissive/Impermissive Rationality

As already pointed out in section 5.1.1, Uniqueness has *inter*personal as well as *intra*personal *impermissive* import; Interpersonal Permissivism has only *intra*personal *impermissive* import and lacks *inter*personal *impermissive* import; Intrapersonal Permissivism lacks *inter*personal as well as *intra*personal *impermissive* import.

If a principle has *intra*personal *impermissive* import, it is naturally compatible with Conservatism. Thus Interpersonal Permissivism and Uniqueness are compatible with Conservatism. To illustrate, suppose that Interpersonal Permissivism holds. Note that Interpersonal Permissivism is motivated *only* by Permissive Intuitions 1(see section 5.1.1): it is permissive across individuals (*inter*personal *permissive* import) but impermissive across time-slices of any particular individual (*intra*personal *impermissive* import). Then there is a possible evidential situation (say total evidence *E*) in which two agents, say John and Paul, who share the total evidence *E* have different doxastic states, D_1 and D_2 ($D_1 \neq D_2$), which are uniquely permissible for John and Paul, respectively (D_1 for John and D_2 for Paul). Suppose further that John and Paul with the total evidence *E* undergo no learning experience between t_1 and t_2. That is, *E* is John and Paul's total evidence at t_1, at t_2 and at every instant of time between t_1 and t_2. Then, we can lay out Paul's doxastic history and John's between t_1 and t_2, respectively, as follows:

Figure 1: Paul's doxastic history between t_1 and t_2

…….. D_1……. at t_2 (Only D_1 is permissible for Paul.)

↑

…….. D_1……...at t_1 (Only D_1 is permissible for Paul.)

Figure 2: John's doxastic history between t_1 and t_2

…….. D_2…… at t_2 (Only D_2 is permissible for John.)

↑

…….. D_2……..….. at t_1 (Only D_2 is permissible for John.)

At each time, Paul and John should have D_1 and D_2 respectively, and thus they naturally satisfy Conservatism between t_1 and t_2. We can easily show that Uniqueness is also compatible with Conservatism in a similar way. Thus, anyone who already endorses Interpersonal Permissivism or Uniqueness can consistently endorse Conservatism as well.

In contrast to Interpersonal Permissivism, Intrapersonal Permissivism is motivated by both Permissive Intuitions 1 and 2 (see section 5.1.1): they are permissive across individuals (*inter*personal *permissive* import) and permissive across time-slices of a particular individual

(*intra*personal *permissive* import) as well. As opposed to Permissive Intuitions 1 that is compatible with Conservatism, Permissive Intuition 2 is *generally* in tension with Conservatism. To see this, following Meacham (2014: 1190), let us divide Permissive Intuition 2 into two parts:

> Permissive Intuition 2-1: There are permissive cases in which a possible agent at *initial* time t_0 ($i = 0$) can rationally adopt a range of different beliefs.
>
> Permissive Intuition 2-2: There are permissive cases in which a possible agent at *non-initial* time t_i ($i > 0$) can rationally adopt a range of different beliefs.

Of course, Permissive Intuition 2-1 is *compatible* with Conservatism because, in any case that invokes *only* Permissive Intuition 2-1, Conservatism is *vacuously* satisfied. (Note that Conservatism is a *diachronic* principle.) However, Permissive Intuition 2-2 is in tension with Conservatism. To illustrate, suppose that Intrapersonal Permissivism holds. Then there is a possible evidential situation (say total evidence *E*) in which a particular agent, say John, with the total evidence *E* can rationally adopt a range of different doxastic states, D_1 and D_2, at time t_0. Suppose further that John begins in initial doxastic state, D_1, and undergoes no learning experience between t_0 and t_1. That is, *E* is John's total evidence at t_0, at t_1 and at every instant of time between t_0 and t_1. Then, we can lay out John's doxastic history between t_0 and t_1 as follows:

Figure 3: John's doxastic history between t_0 and t_1

…….. D_1……. at t_1 (If Epistemic Conservatism holds, only D_1 is permissible for John; if both D_1 and D_2 are still permissible for John, Epistemic Conservatism does not hold.)

↑

…….. D_1……. at t_0 (D_1 and D_2 are permissible for John and Epistemic Conservatism vacuously holds.)

In the case under consideration, at the initial time t_0, Intrapersonal Permissivism that appeals to *Permissive Intuition 2-1* is clearly *compatible* with Conservatism: at t_0, by assumption, D_1 and D_2 are permissible for John, and Conservatism *vacuously* holds because it is a diachronic principle. However, at t_1, Intrapersonal Permissivism that appeals to *Permissive Intuition 2-2* is *no longer compatible* with Conservatism: If Conservatism holds between t_0 and t_1, John should retain his initial doxastic state, D_1, at t_1, because he learns nothing between t_0 and t_1. Thus, at t_1, only D_1 is permissible for John; if Intrapersonal Permissivism still holds at t_1, Conservatism does not hold, because, even though John learns nothing between t_0 and t_1, both D_1 and D_2 are permissible for John at t_1. Thus, at t_1, it is rational for John to make a doxastic state shift from D_1 to D_2, in the absence of new evidence. Note that even if John *actually* retains his initial doxastic state, D_1, at t_1, Conservatism does not hold as long as D_2 is permissible for John at t_1.

As illustrated above, Intrapersonal Permissivism that appeals to Permissive Intuition 2-2 is *incompatible* with Conservatism.[39] Thus, if an agent like John at t_i ($i > 0$) is rational, s/he cannot accept both Intrapersonal Permissivism and Conservatism, on pain of inconsistency. When Intrapersonal Permissivism is incompatible with Conservatism, should we accept the latter rather than the former? As I will explain below, it seems not.

5.2.2 Motivations for Conservatism and Intrapersonal Permissivism

Let me start by briefly considering some motivations for Conservatism that I examine. Some philosophers may claim that Conservatism would be required from a *pragmatic* perspective because, other things being equal, changing one's mind is an energy consuming process.[40] Other philosophers may endorse a learning-preserving motivation for Conservatism: They may claim that Conservatism would be required from an *epistemic* perspective because, given one's degrees of belief properly reflect what she has learned, conservative doxastic attitudes would preserve what she has learned.[41]

Now, I do not have space here to give a full account of how to motivate Conservatism, but I think that we can safely say that commitments that motivate the endorsement of Conservatism are commitments that lead an agent to regard her *past* response to her total evidence (say *E*) as a *new* constraint on a *present* response to *E*, when *no* alteration to *E* has occurred in the meantime. Do such motivations provide an *intrapersonal permissivist* good *epistemic* reason to accept Conservatism? It seems not. Why? From *purely epistemic* perspective,

[39] There might be weaker versions of Conservatism that are compatible with Intrapersonal Permissivism. The version of Conservatism that I examine in this paper is stronger than such weaker versions.
[40] See Sklar (1975).
[41] See Harman (1986).

an intrapersonal permissivist would think that, in any case where a possible agent at *non-initial* time t_i ($i > 0$) can rationally adopt *a range of different beliefs*, Conservatism is *counter-intuitive*, because Conservatism implies that the agent should regard her past *optional* response to her total evidence (say *E*) as a *new* constraint on a present response to *E*, when *no* alteration to *E* has occurred in the meantime and thus *E* is still intrapersonally permissive from the agent's *current* epistemic point of view.[42]

To illustrate, let us consider the learning-preserving motivation for Conservatism. As already pointed out, according to the learning-preserving motivation, in the absence of new evidence, when her doxastic state properly reflects her total evidence, the agent should retain her doxastic state due to the fact that her doxastic state itself is her repository of past learning. Now, suppose that total evidence *E* is permissive with respect to the range of different degrees of belief in *P*, d_1 and d_2, open to a *particular* individual, say Bob, who learns *E* and nothing else between t_0 and t_1. Then, we can say that, from a purely epistemic perspective, adopting d_1 and adopting d_2 equally well reflect what Bob has learned (*E*) between t_0 and t_1. Thus, at t_1, it is *optional* for Bob to adopt which one out of two degrees of belief in *P*. Suppose further that, Bob happens to adopt d_1 at t_1, and undergoes no learning experience between t_1 and t_2. In such a case, according to Conservatism, at t_2, Bob *should* retain his degree of belief in *P* (d_1), which is thus *uniquely* permissible for him at that time. That is, according to Conservatism, a *new* constraint on a proper response to *E* has occurred between t_1 and t_2. In Bob's case, has *any* sort of alteration to *E* occurred between t_1 and t_2? Some might think that the fact that Bob at t_2 is aware of Bob at t_1's

[42] Note that, as already pointed out in section 5.1.1, in this paper, Permissivism is the *evidential one*, which implies that the evidence *alone* suffices to fix what a rational credal state is.

degree of belief in P (d_1) but not the reverse leads them into different evidential situations.[43] However, note that adopting d_1 is *optional* to Bob, and thus, if Bob adopted d_2 rather than d_1 at t_1, Bob would still be as equally (epistemically) rational as Bob actually is. It is hard to see why we should regard the awareness of Bob's past *optional* response to the total evidence E as a sort of *new* evidence about P. So, in Bob's case where Intrapersonal Permissivism holds at t_1, I think, we can safely say that *no* alteration to E has occurred between t_1 and t_2 and so E is still intrapersonally permissive to Bob at t_2, without already assuming conservative principles.

Thus, given Intrapersonal Permissivism, it is counter-intuitive that, even if Bob at t_1 is permitted to have different doxastic states as a rational response to E, Bob at t_2 is not, when *no* alteration to E has occurred in the meantime. Similar points apply to the other motivation for Conservatism. Therefore, if it is admitted that a rational agent can adopt *a range of different beliefs* through time, no epistemically intuitive form of the conservative idea remains.[44]

5.2.3 Cognitive Decision Theory, Conditionalization, and Epistemic Conservatism

Is there any (independent) argument that succeeds in showing that Conservatism is epistemically more fundamental than Intrapersonal Permissivism? If so, we should of course reject Intrapersonal Permissivism in favor of Conservatism. Christensen (1994) argues that *no* account has succeeded in providing an *epistemic* justification for Conservatism. However, one may think that some CDTers have already provided an epistemic justification for Conservatism. For

[43] Note that, if so, we cannot apply Intrapersonal Permissivism to Bob's case because Bob at t_1 and Bob at t_2 do not share the exactly same total evidence E.

[44] Note that, as already pointed out, here I assume a particular version of Conservatism, according to which, one is obliged to *continue to have* credences in a proposition in the absence of learning experience. Thus, some philosophers who have something different in mind might not think that to adopt *a range of different beliefs* through time is in conflict with the conservative idea.

instance, as I have already pointed out in section 1, Greaves and Wallace (2006), Leitgeb and Pettigrew (2010a, 2010b), and Easwaran (2013) apply an approach of CDT to provide an epistemic justification for Conditionalization: they attempt to show that, given that an epistemic utility is *a purely epistemic desirability*, when an agent learns a proposition with certainty and nothing else, Conditionalization is the *unique* updating policy that maximizes expected epistemic utility.[45, 46] If any of those arguments is successful, a sort of epistemic (indirect) justification for Conservatism may follow, because, at least in some cases, Conditionalization implies Conservatism. For instance, suppose that an agent updates her degrees of belief *only* by Conditionalization. Then, it immediately follows that, in the absence of new evidence, she must stick with her current doxastic state. Here let me focus on Greaves and Wallace (2006).

It is noteworthy that Greaves and Wallace (2006) explicitly assumes Strict Immodesty and Maximization for their arguments. And, in fact, given Strict Immodesty and Maximization, we may provide a *direct* justification for Conservatism, because we can easily prove that Conservatism immediately follows from Strict Immodesty and Maximization. Here is the proof: Suppose that you undergo no learning experience between t_i and t_j ($t_i < t_j$) and your doxastic state

[45] Of course, there are crucial differences between those arguments. For instance, while Leitgeb and Pettigrew (2010a, 2010b) and Easwaran (2013) assume that (in)accuracy is the only thing that matters to epistemic utility, Greaves and Wallace (2006) involve no such an assumption. (They just talk about 'epistemic utility'.) Moreover, while Leitgeb and Pettigrew (2010a, 2010b) and Greaves and Wallace (2006) restrict their discussion to finite set of (possible) states of the world, Easwaran (2013) provides an argument for Conditionalization that extends Leitgeb and Pettigrew's result in a way that includes infinite set of states of the world. Lastly, in contrast to Greaves and Wallace (2006) and Leitgeb and Pettigrew (2010b), even though Easwaran (2013: 134) provides and considers an argument for Conditionalization, he himself seems not to endorse it due to 'a notoriously difficult problem' that, as Christensen (1991) points out, diachronic norms face. He (p. 135) says that "Despite the fact that a variety of arguments formally reach diachronic conclusions, Christensen (1991, 246) argues that "without some independent reason for thinking that an agent's present beliefs must cohere with her future beliefs," there can be no more support for a diachronic norm (like conditionalization) than for an interpersonal norm (e.g., the requirement that one's credences must match those of one's spouse."

[46] Of course, as well known, there are (diachronic) Dutch book arguments for Conditionalization. For instance, see Teller (1973) and Lewis (1999). However, many take Dutch book arguments to fail to provide a purely *epistemic* justification for Conditionalization. For instance, see Joyce (1998). So let us set aside Dutch book arguments in this paper.

at t_i and your doxastic state at t_j are c_i and c_j, respectively (c_i, $c_j \in \boldsymbol{C_S}$). By Strict Immodesty, $EU_{ci}(c_i) > EU_{ci}(c_j)$ if $c_i \neq c_j$. Thus, given that you evaluate c_i and c_j from your *prior* epistemic perspective at t_i. in order to maximize your expected epistemic utility, you should retain your doxastic state c_i at t_j from your epistemic perspective at t_i.

Given Strict Immodesty and Maximization, therefore, one could provide the following argument for Conservatism:

Argument III: A direct argument for Conservatism

Q1: Strict Immodesty

Q2: Maximization

Q3: You evaluate relevant (future) doxastic states from your prior epistemic perspective.

[47]

[47] Note that Argument III works only if the epistemic evaluation of doxastic states (c_i and c_j) is done from your *prior* epistemic perspective at t_i. In Standard Decision Theory for practical rationality, it is natural to evaluate present (or future) possible outcomes *only* from *prior* point of view, because it assumes that voluntariness is necessary for the applicability of decision-theoretic considerations. However, CDT does not assume such a sort of voluntarism about beliefs. Given that we understand CDT as an apparatus to evaluate doxastic states or updating policies, as long as c_i 's epistemic evaluation of c_i and c_j is allowed, c_j 's epistemic evaluation of c_i and c_j is as well. And, when you update your doxastic state in a way that violates Conservatism, if the epistemic evaluation of c_i and c_j is done in reverse order (i.e., from your *posterior* epistemic perspective at t_j), Strict Immodesty and Maximization would justify your *non-conservative* updating policy. It is also noteworthy that a similar point applies to Easwaran (2013: 134), Greaves and Wallace (2006: 623-627), and Leitgeb and Pettigrew (2010b: 248-250). For instance, Leitgeb and Pettigrew (2010b) attempts to justify Conditionalization by showing that it follows from what they call *Accuracy (Diachronic expected local)*, which says that, "where an agent has learned evidence between *t* and *t'* that imposes constraints *C* on her belief function $b_{t'}$ at time *t'* or on the set *E* of worlds that are epistemically possible for her at *t'* or both …, at time *t'*, such an agent ought to have a belief function that satisfies constraints *C* and is minimal among belief functions thus constrained with respect to the expected local inaccuracy of the degrees of credence it assigns to each proposition $A \subseteq W$ by the lights of her belief function at time *t*, relative to a legitimate local inaccuracy measure and over the set of worlds that are epistemically possible for her at time *t'* given the constraints *C*." (p. 241) Note that here the epistemic evaluation of 'her belief function $b_{t'}$' is done from her *prior* epistemic perspective at *t*. And, when the epistemic evaluation of 'her belief function b_t' is done from her *posterior* epistemic perspective at *t'*, we cannot justify Conditionalization in a way that Leitgeb and Pettigrew (2010b) suggest. In fact, we can easily prove that, for any diachronic principle *x*, when the *reverse* epistemic evaluation of doxastic states is allowed, Strict Immodesty and Maximization justify any updating policy that violates *x*. See Easwaran (2013: 134-5) for the similar point.

C''1: When you undergo no learning experience, you should retain your current doxastic state from your prior epistemic perspective. (From Q1, Q2, and Q3)

Do the above arguments convince us to endorse Conservatism rather than Intrapersonal Permissivism? In particular, do they succeed in showing that any CDTer should reject Intrapersonal Permissivism in favor of Conservatism? No. Why? Note that even though there are substantial differences between these (direct or indirect) arguments for Conservatism in details, they share the following crucial features: These arguments for Conservatism explicitly assume Strict Immodesty and Maximization.[48] And, as already pointed out in section 5.1.2, Strict Immodesty and Maximization entail Uniqueness or Interpersonal Permissivism. That is, they implicitly assume Uniqueness and Interpersonal Permissivism. Thus, to CDTers who endorse Intrapersonal Permissivism, those arguments for Conservatism would be regarded as begging the question.

To sum up, then, on the assumption of Intrapersonal Permissivism, as far as I know, no argument successfully convinces a CDTer to endorse Conservatism rather than Intrapersonal Permissivism. Conservatism itself alone does not seem to give us epistemic reason to reject Intrapersonal Permissivism.

5.2.4 Argument I and Intrapersonal Permissivism

Now let us return to our original question: Is Argument I a real threat to those CDTers who endorse Intrapersonal Permissivism? No. To possible agents at *initial* time t_0 who hold

[48] Greaves and Wallace (2006: 625).

Intrapersonal Permissivism, the conclusion (C2) of Argument I does not follow because, at t_0, Conservatism vacuously holds; To possible agents at *non-initial* time t_i ($i > 0$) who hold Intrapersonal Permissivism, C2 does not follow because they would reject Conservatism (P3) in favor of Intrapersonal Permissivism. Therefore, Argument I is not a real threat to CDTers who endorse Intrapersonal Permissivism.

To sum up, then, we have considered the argument against CDT (Argument I): On the assumption of Conservatism, the correct theory of epistemic rationality will not endorse non-conservative doxastic state shifts from one degree of belief function to another, in the absence of new evidence. When Non-Strict Immodesty holds, CDT endorses non-conservative doxastic state shifts, in the absence of new evidence. This seems to be an unfortunate consequence. However, when we clearly divide up Permissivism/Uniqueness, we can see that Argument I is not a real threat to any CDTer, as long as each kind of Permissivism is a viable option: To those who endorse Uniqueness or Interpersonal Permissivism, Argument I is not a real threat to them because they would reject Non-Strict Immodesty in favor of Strict Immodesty; to those who endorse Intrapersonal Permissivism, Argument I is not a real threat to them because either it could not be applied to them (when they are at initial time t_0) or they could reject Conservatism in favor of Intrapersonal Permissivism (when they are at non-initial time t_i ($i > 0$)).

6 Conclusion

As I have shown above, depending on which versions of Permissivism/Uniqueness CDTers embrace, they would respond to Argument I in significantly different ways. Which one is better?

In this paper, I have not intended to defend a particular way. My purpose here is to show how strongly CDT relies on a view about permissive/impermissive rationality: On the assumption of Uniqueness or Interpersonal Permissivism that entails Strict Immodesty or the denial of Maximization, a CDTer would expect more stability of beliefs over time, while, on the assumption of Intrapersonal Permissivism that entails Non-Strict Immodesty or the denial of Maximization, a CDTer would allow that the possibility of changing one's beliefs on a whim fits naturally with epistemic rationality.

Many might think that the latter does not fit well with their intuition. And, in fact, among many philosophers who endorse Epistemic Immodesty, many defend Strict Immodesty rather than Non-Strict Immodesty. Non-Strict Immodesty is not much defended in the recent epistemic utility literature. However, it is noteworthy that, as long as Intrapersonal Permissivism is a *viable* position, as I have shown above, Non-Strict Immodesty is a defensible epistemic principle in the context of CDT as well.

References

Chisholm, R. (1980). "A Version of Foundationalism." in Wettstein, et al (eds), Midwest Studies in Philosophy V, University of Minnesota Press, Minneapolis, MN.

Christensen, D. (1991). "Clever Bookies and Coherent Beliefs." *Philosophical Review*, Vol. 100, No. 2: 229-247.

Christensen, D. (1994). "Conservatism in Epistemology." *Noûs* 28: 69-89.

Christensen, D. (2000). "Diachronic Coherence versus Epistemic Impartiality." *The Philosophical Review*, Vol. 109, No. 3: 349-371.

Christensen, D. (2007). "Epistemology of disagreement: The good news." *The Philosophical Review* 116(2), 187–217.

Christensen, D. (2013). “Epistemic modesty defended. In D. Christensen & J. Lackey (Eds.), *The epistemology of disagreement: New essays* (pp. 77–97). Oxford: Oxford University Press.

Christensen, D. (2016). “Conciliation, Uniqueness and Rational Toxicity.” *Noûs* 50(3): 548-603.

Easwaran, K. (2013). “Expected Accuracy Supports Conditionalization—and Conglomerability and Reflection.” *Philosophy of Science*, 80(1): 119-142.

Elga, A. (2013). “The Puzzle of the Unmarked Clock and the New Rational Reflection Principle,” *Philosophical Studies* 164: 127-139.

Feldman, R. (2007). ‘‘Reasonable Religious Disagreements.’’ in L. Antony ed. *Philosophers without God.* (pp. 194–214). Oxford: Oxford University Press.

Gibbard, A. (2007) “Rational Credence and the Value of Truth.” in T. Gendler and J. Hawthorne, eds., *Oxford Studies in Epistemology* vol. 2. Oxford: Clarendon Press.

Greaves, H., and D. Wallace. (2006). “Justifying Conditionalization: Conditionalization Maximizes Expected Epistemic Utility.” *Mind* 115 (459): 607–32.

Harman, G. (1986). *Change in View*. Cambridge, MA: MIT Press.

Joyce, J. M. (1998). ‘‘A Nonpragmatic Vindication of Probabilism.’’ *Philosophy of Science* 65: 575–603.

Joyce, J. M. (2009). “Accuracy and Coherence: Prospects for an Alethic Epistemology of Partial Belief.” in F. Huber and C. Schmidt-Petri (eds.), *Degrees of Belief*. Synthese Library.

Joyce, J. M. (2010). “A defense of imprecise credences in inference and decision making.” *Philosophical Perspectives* 24: 281–323.

Kelly, T. (2014). “Evidence Can Be Permissive.” in M. Steup, J Turri, and E. Sosa, eds., *Contemporary Debates in Epistemology*. Second Edition. Wiley Blackwell.

Lasonen-Aarnio, M. (2014). “Higher-Order Evidence and the Limits of Defeat.”.*Philosophy and Phenomenological Research* 88(2): 314-345.

Leitgeb, H., and R. Pettigrew. (2010a). “An Objective Justification of Bayesianism I: Measuring Inaccuracy.” *Philosophy of Science* 77 (2): 201–35.

Leitgeb, H., and R. Pettigrew. (2010b). “An Objective Justification of Bayesianism II: The Consequences of Minimizing Inaccuracy.” *Philosophy of Science* 77 (2): 236–72.

Levinstein, B. (2015). “Permissive Rationality and Sensitivity.” *Philosophy and Phenomenological Research* (2015 Online First), (DOI) 10.1111/phpr.12225.

Lewis, D. (1971). “Immodest Inductive Methods.” *Philosophy of Science* 38: 54-63.

Lewis, D. (1980). “A Subjectivist's Guide to Objective Chance.” In *Studies in Inductive Logic and Probability*, eds. R. C. Jeffrey, 263-293. vol. 2. Berkeley: University of California.

Lewis, D. (1999). “Why Conditionalize?” in Lewis, D. *Papers in Metaphysics and Epistemology*, CUP.

Maher, P. (1993). *Betting on Theories*. Cambridge University Press.

Mayo-Wilson, C. and G. Wheeler (ms). “Scoring Imprecise Credences: A Mildly Immodest Proposal.” (forthcoming in *Philosophy and Phenomenological Research*).

Meacham, C. (2014). “Impermissive Bayesianism.” *Erkenntnis* 79: 1185-1217.

Oddie, G. (1997). “Conditionalization, Cogency, and Cognitive Value.” *Brit. J. Phil. Sci.* 48: 533-541.

Pettigrew, R. (2011). “Epistemic Utility Arguments for Probabilism.” *The Stanford Encyclopedia of Philosophy* (2015 Edition), Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/entries/epistemic-utility/>.

Pettigrew, R. (2013). “A New Epistemic Utility Argument for the Principal Principle.” *Episteme*, 10, 1: 19–35.

Pettigrew, R. (2014). “Accuracy, Risk, and the Principle of Indifference.” *Philosophy and Phenomenological Research* 91(1) : 1-30.

Pettigrew, R. (2016). *Accuracy and the Laws of Credences*. Oxford University Press.

Quine, W.V.O. (1951). "Two Dogmas of Empiricism," in *From a Logical Point of View*, 2nd ed. New York: Harper & Row, 1961.

Schoenfield, M. (2014). “Permission to believe: Why permissivism is true and what it tells us about irrelevant influences on belief.” *Noûs* 48(2), 193–218.

Schoenfield, M. (2015). “The Accuracy and Rationality of Imprecise Credences.”

Sklar, L. (1975). "Methodological Conservatism," *The Philosophical Review* 84: 374-400.

Teller, P. (1973). “Conditionalization and Observation.” *Synthese* 26(2): 218-258.

Vahid, H. (2004). “Varieties of Epistemic Conservatism.” *Synthese* 141(1): 97-122.

Weintraub, R. (2013). “Can Steadfast Peer Disagreement Be Rational?” *The Philosophical Quarterly* 63: 740-759.

White, R. (2005). “Epistemic Permissiveness.” *Philosophical Perspectives* 19: 445-459.

White, R. (2014). “Evidence cannot be permissive.” In M. Steup, J. Turri, and E. Sosa (Eds.), *Contemporary Debates in Epistemology*. Wiley-Blackwell.

Backtracking analysis and causal ascription of singular historicals

Richard Hou
Department of Philosophy
National Chung Cheng University

Abstract

One task of historians is to construct causal ascription of singular historicals between eminent historical events. For instance, the controversy resulting from the confusing butterfly ballot of Florida's year 2000 presidential election cost Gore his presidency. However, to research into these matters is inevitably to appeal to counterfactual deliberation in an epistemic fashion because the past is fixed. One standard idea is Max Weber's, Weber causation: "f was a cause of φ" is assertable iff "$\neg f \,\square\!\!\rightarrow \neg\varphi$" is assertable. Reiss (2009) gives an exceptionally good analysis of this topic and outlines historians' reasoning, claiming that backtracking analyses of counterfactual conditionals employed in historical thought experiments is the signature of historical study of causal ascription of singular historicals. Nevertheless, he concludes that it is very difficult to reach an uncontroversial ascription for this sort in most cases. For this reason, he proposes to find difference-making relations that will suffice. The objective of this paper is to provide a more fine-grained, intervention-based, backtracking analysis of counterfactual conditionals upon which a more satisfactory account of causal ascription of singular historicals can be given. Reiss' account of difference-making relation will be shown to be unsatisfactory. Moreover, a formal ground of the epistemology of historical thought experiments can be given, along with the constraints of this account resultant from the semantic features of non-transitivity and strong centring of counterfactual conditionals. Finally, some epistemological points of causal ascription of singular historicals and historical thought experiments will be given.

Key words: backtracking counterfactual analysis, historical thought experiment, intervention, causal ascription of singular historicals, Weber causation

I. Introduction

One may ask, "Why did Peter go to Hong Kong?" Another may answer, "He went because his best friends, Paul and Mary, are getting married". This mundane conversation involves a simple causal ascription that were Paul and Mary not getting married, Peter would not have gone to Hong Kong. If this is not the case, then the above answer is wrong and the causal ascription is false.[1] Causal ascription of

[1] While this is correct in many cases of causal ascription, it is not so in every case. The complication

singular historicals (CASH) is a more sophisticated matter regarding the evaluation of the gravity of significant historical events or the significance of historical figures or their decisions or actions. Some solid research is needed to vindicate this particular kind of historical research.

During the decade before the turn of the century, the employment of counterfactuals and/or thought experiments in historical study was considered as akin to doing virtual history or fiction. This is because the future depends on the present but the past does not depend on the present. It follows that the counterfactual conditionals employed have to undertake backtracking analysis. The problem of backtracking analyses of counterfactual conditionals is reflected by two asymmetries. The first asymmetry is causal, viz. no backward causation: causes always precede or are simultaneous with their effects. The second is future openness, viz. the future seems open and the past is fixed. For Lewis, we should only use non-backtracking counterfactual conditionals for which we hold the past fixed up to the point when the antecedent is supposed to obtain. The fixedness of the past thus prescribes to the historical counterfactual conditionals the backtracking analyses feature and the non involvement in causal dependency feature. These asymmetries are the reasons, semantically and metaphysically, why standard Lewisian semantics forbids backtracking analysis of counterfactual conditionals, considering them to be non-standard and unstable (Lewis 1979). This also explains why many historians hold that the utilisation of counterfactual conditionals and thought experiments in studying history is fictional.

Nevertheless, of recent, interest in historical counterfactual conditionals and thought experiments has been growing. This rise did not result from the change of the attitude of historians toward virtual history—that is, backtracking analysis of counterfactual conditionals in cases of CASH do reveal the exact causal relations between historical events—but from a different attitude toward backtracking analysis. Backtracking analysis is not to be seen as a tool of exploring the metaphysical aspect of history but only the epistemic aspect, constructing and testing hypothetical CASHs.

In this respect, the importance of finding an epistemology of historical thought experiments is urged by many, that without the employment of counterfactuals and/or historical thought experiments causal ascriptions of singular historicals can hardly make sense.[2] Such an ascription is not a simple matter of putting two historical

---

will be discussed further in the third and fourth sections.

[2] For instance, Daniel Nolan, “Why historians (and everyone else) should care about counterfactuals”

events together to *claim* a historical causal connection between them. In order to support a particular historical causal ascription, historians who are sympathetic to this sort of *virtual*, or rather *counterfactual*, history view may raise many historical facts as evidence. However, the exact structure of this ascription is unclear.

One task of historians is to connect different historical events in a way to render them into some sort of *singular* causal relation, relying on, for instance, counterfactual dependency.[3] The way to carry out the above tasks is studied by philosophy of history and is termed by Reiss (2009) as the implementation of historical thought experiments. Similar to the study of other types of thought experiments, what concern us are the reasoning structure(s) and the epistemology. Different from other types, the peculiar features of the historical ones are their empirical character and the reliance on backtracking analysis of counterfactual conditionals, which are the main concerns of this paper.[4]

By means of counterfactual dependency, Max Weber proposed an account of CASH, call it *Weber causation*, which will be discussed shortly. Showing the difficulty in ascertaining whether or not certain counterfactual dependency is satisfied, Reiss (2009) disagrees with Weber's idea and suggests a *difference making* relation to replace Weber causation. In view of Reiss' account of backtracking analysis of counterfactual conditionals, I agree with his criticism of Weber causation. However, with a more fine-grained account I will argue for a modified version of Weber causation and shed further light on why Reiss' difference making account is not acceptable. In order to show this, there are two steps. The first is to study, through some examples, the reasoning structure of historical thought experiments, especially the backtracking analysis of the respective (historical) counterfactual conditional(s). Based on the outcome of this study, the second step is to delineate the formal ground of historical thought experiments and the epistemic meaning of the execution of historical thought experiments. Note that my objective is not to do historical research but to do philosophical and methodological study of a particular branch of historical study, viz. the study of CASH. Therefore, how to historically justify a CASH or the assertability of the respective counterfactual conditional goes beyond the compass of the current discussion.

(*Philosophical Studies* (2013) 163:317–335).

[3] For a more detailed discussion, please see the fourth section.

[4] Employing counterfactual conditionals means this account has to inherit the semantic features of non-transitivity and strong-centring. The fifth section will be explained why this account is compatible with them.

The starting point is the outline of the relevant issues of CASH discussed in Reiss 2009 in the following section, which for the sake of discussion focus is taken for granted**.** Reiss also explains why in most cases, if not all, a CASH seems to be too difficult to construct. Although the starting point is Reiss' analysis, and one particular example from his discussion, it will be shown in the third and fourth sections that a more fine-grained analysis of backtracking counterfactual conditionals sheds light on both why Reiss reached his conclusion and why it is not correct. In the following, we are going to find out what is essentially the philosophical importance of backtracking analysis of counterfactuals to CASHs and what is purely historically intricate of finding relevant historical facts, as evidence, to support a construction of a CASH, and, moreover, why lacking a refined account of backtracking analysis leads to a conflation of these two. But first, the discussion platform needs to be set up by introducing some of Reiss' basic ideas.

II. Reiss' basic ideas

Three main issues of CASHs are what sort of causal structure we are after, even just epistemically, what sort of analysis of relevant counterfactual conditional(s) is needed, and how to make sense of them. The third issue consists of two parts: first, what is the structure of a judgement of a CASH, and second, why a given counterfactual conditional is relevant to such a judgement. A brief explanation of each of them is presented in the following, given which the formal structure and epistemological questions can be addressed.

1. Weber causation

Weber causation, originally proposed by Max Weber ([1905] 1949, 171), is as follows:

> **(WC)** "f was a cause of φ" is assertable iff "$\neg f \,\Box\!\!\rightarrow \neg\varphi$" is assertable (Reiss 2009: 721),

where f and φ are actual historical events and '$\Box\!\!\rightarrow$' is the symbol for counterfactual conditional. There are three things of note here. First, to talk about assertability is to talk about the epistemic standards satisfying of which the proposition that f was a cause of φ is allowed to be asserted.[5] Second, although a CASH claim, such as "f was a cause of φ", is of course metaphysical, the whole point of pondering a CASH is the expectation of learning something from an epistemic evaluation of history because the idea of manipulating and thus experimenting on historical events is absurd. This

[5] Except truth normativity, all other accounts of assertion normativity are about different epistemic features of the normativity of assertion making.

explains the formulation of (WC), the truth of which relies on whether there is evidence supporting the assertion of the counterfactual conditional that $\neg f \,\Box\!\!\rightarrow \neg\varphi$. Third, normally the assertability of "$\neg f \,\Box\!\!\rightarrow \neg\varphi$" cannot guarantee that *f* was a cause of φ. Counterfactual dependency does not suit the need of explaining causation in every case. The same applies to epistemic consideration of counterfactual dependency. However, it can only be adumbrated here with an indication that a suitable account of counterfactual backtracking analysis is capable of clearing all the sceptical air. Further discussion and arguments will be given in the third and fourth sections.

2. The assertability of a historical counterfactual conditional

Naturally what makes a counterfactual conditional of (WC) assertable is the next thing to explain. Whether or not it is assertable depends on a historical judgement that is made by a historian in terms of the following reasoning pattern.

> (1) Let *H* be a historian with beliefs about the relevant evidence and causal generalizations *B* and *C* a historical context such that $\neg f$.
> (2) Relative to *B*, the counterfactual "$\neg f \,\Box\!\!\rightarrow \neg\varphi$" is assertable iff
> **f*, φ obtained,
> *$\neg f$ is historically consistent and precise enough to allow of a judgment regarding φ,
> **H* judges $\neg\varphi$ to obtain in *C*. (Reiss 2009: 720)

(1) and (2) require further clarification respectively.

For (1), almost everything could be *made* possible in history if some other historical events were changed drastically. So historians accept a *minimal rewrite rule* in terms of which a counterfactual antecedent is required not to falsify too much of what we know of the actual course of historical events. The implementation of the rule also depends on what causal generalisations—social, political, or economical, and so on—historians accept (Reiss 2009: 719).[6] From Tetlock and Belkin 1996 (23) we get three further ideas of this rewrite rule; viz. the pondering of CASH starts with the real world right before the counterfactual is asserted, it is unnecessary and even incoherent to disentangle the historical past in order to rewrite counterfactually long stretches of history, and, the knowledge of what the historical figures in question believes or the goal they want to reach is not to be disturbed too much.

[6] Strictly speaking, this is not a *ceteris paribus* condition. The reason is clearly that backtracking analysis of relevant counterfactual conditional(s) of a given CASH is by default to backtrack to the precondition of the counterfactual antecedent, and if necessary to backtrack further. The process of backtracking is limited but not so limited to just amending the antecedent or its precondition.

The issues that are important to (2) are the following. First, regarding a counterfactual conditional, such that $\neg f \,\square\!\!\rightarrow \neg\varphi$, the sensibility of a historical judgment, the ease with which causal ascriptions can be justified, depends on how assertable the counterfact that $\neg f$ is.[7] The way historians do the evaluation is to assess the likelihood of the antecedent obtaining in the course of history while certain causal conditions were present. For an antecedent that was likely to obtain as such, Reiss calls it *historically consistent* (Reiss 2009: 720).[8] Evidently, the minimal rewrite rule delineates the basic condition following which the integrity of the historical consistency of the construction of a CASH is at least partially preserved.

Second, the way we evaluate whether the antecedent is historically consistent is to *backtrack* to its precondition. This is majorly different from the general line of opinion according to which if counterfactual conditionals are used to capture or used as stands-in for causal claims, they have to be non-backtracking. Lewisian semantics requires the antecedent to be implemented by bringing the alternative event about by a miracle, with no cause event of its own. Historians apparently would be humiliated by this requirement, demanding them to base their historical judgements on some miracles. Fortunately, what is at issue here is not the semantics of counterfactual conditionals employed by historians but how a backtracking analysis of counterfactual conditionals serves to outline the formal structure and the epistemological features of a CASH.[9]

3. Historical thought experiments

Since actual historical events are not manipulable, no empirical experiments can be constructed. The way historians evaluate a counterfactual conditional, say, that $\neg f \,\square\!\!\rightarrow \neg\varphi$, is to construct a thought experiment in which imagination is employed to psychologically remove f from the actual course of history and to inquire into whether or not this removal would make a difference to the occurrence of the target event φ.

[7] Some may disapprove of the idea of a counterfact, holding it to be nonsense. Since it is rather a terminology issue and thus inessential to the points and arguments presented below, please find your own suitable idea or term(s).

[8] Some may propose that, different from logical consistency, historical consistency rather comes in degrees. After all, it seems to make sense to say that the target counterfact is in certain degree to be consistent with the actual course of history. However, I hold historical consistency to be a property that a counterfact is either with or without. What comes in degree is the likelihood for it to obtain and the quality of evidential support, not historical consistency. Even in some time one may say that within the same evaluation context one the counterfact of one hypothesis is more historically consistent than the other, it is just a loose talk. What one says is still that one counterfact is more likely to obtain than the other or the quality of evidential support is better.

[9] I do believe causal model semantics is the right semantics for historical thought experiments and historical backtracking analysis. However, due to the limit of this paper, this will be left to further research.

The best way to illustrate how historians employ a thought experiment to evaluate whether a given CASH is acceptable is to examine such an instance. In the following, I borrow the Chamberlin case from one of Reiss' examples, which originally is in Khong (1996). The case is about Chamberlin's appeasement policy and Hitler's Sudetenland demands before the outbreak of World War II. Because the purpose of employing this case is to explain how historians rely on thought experiments to evaluate whether a CASH is appropriate, and further to explain how a backtracking analysis of the related counterfactual conditional(s) is carried out, the actual complication of historical facts, or even counterfacts, regarding how to appropriately reason which is the precondition of what, does not concern the following discussion.

It is probably common opinion that Chamberlin's appeasement policy brought about the realisation of Hitler's rapacious ambition, even contributed to World War II. To make sense of this claim, it is necessary to show it is likely to be the case that, *had Chamberlin adopted an anti-appeasement policy* (AA_{CH}), Hitler *would have backed down from his rapacious Sudetenland demands* (B_H). The Weber causation of this is as follows, where the *italic*, including symbols, indicates a counterfactual event and the regular typeface, also including symbols, indicates a historical fact:

> $(WC)^{CH}$ "Chamberlin's appeasement policy (A_{CH}) was a cause of Hitler's not backing down from his Sudetenland demands ($\neg B_H$)" is assertable iff "*Had Chamberlin adopted an anti-appeasement policy* (AA_{CH}), Hitler *would have backed down from his rapacious Sudetenland demands* (B_H)" is assertable.

Put the counterfactual conditional in symbols,

> (CH) $AA_{CH} \mathrel{\square\!\!\rightarrow} B_H$.

To be clear, the left hand side of a Weber causation is the assertability of a CASH, and the right hand side is the necessary and sufficient conditions of the assertability of the CASH. The assertability of the respective counterfactual conditional, such as (CH), is employed to state the necessary and sufficient conditions of the assertability of the target CASH. The historical thought experiment is in turn the backtracking scenario the historian constructs to see whether the counterfactual conditional is assertable. The thought experiment of $(WC)^{CH}$, or rather the counterfactual conditional, conducted by the historian thus has two aims. The first is for the historian to psychologically remove the historical fact of A_{CH} and to see whether or not AA_{CH} is historically consistent. The second is to see whether or not, along with other related historical facts and necessary background generalisations, it bears sufficient

significance to bring about the counterfact B_H. Of course, to bear sufficient significance to bring about some counferfact, when it is said, it does not mean anything metaphysical but epistemic. It means that by means of the related chuck of historical evidence and causal generalisations, the historian judges that, in the relevant context, the counferfact of the antecedent is likely to make the counterfact of the consequent happen, even though it did not. The whole idea of backtracking is the same, that based on the evidence and the generalisations the precondition backtracked to is likely to bring about the counterfact backtracked from. It is all epistemic.

The focus of this paper is to account for the backtracking analysis of counterfactual conditionals, such as (CH), employed in historical thought experiments. For this purpose, the complexity and elaborateness of how historians do consider these matters do not concern the discussion in the following section. Whether the historical setup in the analysis of (CH), as an instance, is adequate is simply assumed for the sake of simplicity. In other words, in the explication of the backtracking analysis, the historical counterparts in all cases are simply assumed with no intention to defend any related historical judgements. More specifically, what concern the study of the backtracking analysis is only about *the philosophical part* of the counterfactual conditional in a Weber causation or its modification. Philosophical analysis aside, whether the conditional is assertable depends on historians' expertise. In other words, the assertability of the conditional is dependent on historians' evaluation of whether the antecedent is historically consistent and whether there is sufficient evidential support for the connection between the antecedent and the consequent. The philosophical part only characterizes the structural feature the historian has to respect while making such an evaluation.

III. Backtracking analysis of counterfactual conditionals

To consider whether (CH) is assertable, and to see whether it is sensible to the respective CASH, there are two steps: first, to see whether AA_{CH} is historically consistent, and second, should it be consistent, to see whether B_H follows from it. The backtracking account matters to both, though the focus of the discussion is mainly on the part of historical consistency. To illustrate, the following is a simplified analysis of (CH):

The simplified analysis

hawkish cabinet member(s) (Eden, Cooper, and Churchill)

◤

the hawkish cabinet

◤

the rearmament was sufficed

◤

(CH) $AA_{CH} \boxdot\!\!\rightarrow B_H$

↳ other historical facts,

where '◤' (or the like) indicates backtracking to the precondition and the **boldface** indicates the historical fact that would anchor the backtracking process starting from the counterfactual antecedent AA_{CH} in the present case. An anchoring fact is essential to historical consistency, without which the backtracking process continues indefinitely, and the assertability of the conditional and the truthfulness of the respective Weber causation are pending for good. I will resume discussion of this issue in the fifth section. Different from Reiss', I insert one more counterfact, that the cabinet was hawkish, to make the case or the discussion more coherent. The discussion below will show the necessity of this.

Historians do not aim at *any* possibility, but at some historical consistent one, for otherwise it is certainly metaphysically possible that Chamberlin was temporarily insane and thereby held an anti-appeasement policy. Lewisian semantics is applicable accordingly. The significance of constructing a CASH with historical facts as evidence is completely undercut. According to Reiss and Khong, the historian backtracks to the precondition of AA_{CH}, the sufficient rearmament since World War I, by which AA_{CH} would be brought about.[10] However, the actual rearmament was insufficient to confront the German demands, which means the backtracking process has to go on to the precondition that would bring about the sufficient rearmament—that is, whether or not the cabinet was hawkish to support a much faster rearmament. This backtracking also reaches a counterfact, and a further backtracking is necessary to ground whether or not there were hawkish cabinet members.[11] In fact, there were three, viz. Eden, Cooper, and Churchill, and one of them might even replace Chamberlin as the prime minister. We may say the fact that there were three hawkish cabinet members (HCM), who occupied important positions, anchors the historical consistency of AA_{CH}, that it was likely to obtain in the actual course of history.[12] An anchoring historical fact is extremely important to each backtracking process, for if there were no such a fact, a counterfactual precondition would be not

[10] The relevant historical pondering, but not the philosophical part, all comes from either Reiss or Khong, or both at times. This will not be further noted.

[11] Obviously, with the insertion of the counterfact of *the hawkish cabinet* the explanation of why the rearmament is insufficient is more sensible. Since in the actual course of history it is a fact that there were hawkish cabinet members *and* it is also a fact that the rearmament was insufficient, the existence of hawkish cabinet members cannot be the anchoring fact of the counterfact of sufficient rearmament.

[12] The same abbreviation, say, HCM, that is not in boldface just means that it is a historical fact.

much different from any given metaphysical possibility. Without HCM as the anchoring fact, the sufficient rearmament, though more germane to AA_{CH}, is just another metaphysical possibility whose status is the same as Chamberlin's temporary insanity.

What concerns Reiss is this. A given instance of Weber causation shows a counterfactual dependency between two events, *f* and *φ*, and had *f* not been, *φ* would not have been. This means that the realisation of *f* makes a difference to the realisation of *φ*. However, it is difficult, and sometimes incredibly difficult, for historians to ascertain that *f* is really *a* cause of *φ*, that any one counterfactual antecedent, as a cause, can be isolated and thereby justify its attribution of causal weight to its consequent. For instance, it may be the case that for Hitler to back down, the implementation of anti-appeasement policy has to go through rearming first, and it is unclear whether rearmament sufficed to refrain Hitler from proposing his Sudetenland demands. Considering this, Reiss suggests that in the case of (CH), as far as the difference-making relation between AA_{CH} and B_H can be found, whether it is the military threat and the diplomacy jointly, or *either* the military threat *or* the diplomacy, making the difference does not matter (Reiss 2009: 722). In other words, Reiss holds that what is important to this sort of historical judgements is not to epistemically construct a suitable CASH but to scrutinize whether there was any significant difference-making relation to be found.

Except how complicated it may be to make reasonable historical judgements of CASHs, let's focus on the causal graph of the simplified analysis:

(CG1) Causal Graph 1

$\neg$HC (the cabinet was not hawkish)

↘

$\neg$RS (the rearmament was not sufficed)

↘

$A_{CH} \rightarrow \neg B_H$,

where '↘' or '→' (or the like) indicates a causal connection.[13] Note that the above graph does not suggest that the fact $\neg$HC is a cause of the fact A_{CH}, or the fact $\neg$RS is a cause of $\neg B_H$—if there were such a causal connection, there would be another arrow to indicate it.[14]

[13] The discussion of a given Weber causation, such as $(WC)^{CH}$, mainly focuses on how to justify the assertability of the respective counterfactual conditional, such as (CH). For the sake of simplicity, the talk of assertability is often omitted, but the point is the same.

[14] For further discussion, please see the following two sections.

To understand a causal graph we need to base it on the corresponding backtracking structure. Corresponding to (CG1), there is a backtracking structure as follows:

(BS1) Backtracking Structure 1

HCM (**hawkish cabinet member(s)**)

↖

HC (*the hawkish cabinet*)

◤

RS (*the rearmament was sufficed*)

◤

$AA_{CH} \boxrightarrow B_H$,

where '↖' (or the like) indicates a historical consistency anchoring. The justification of the CASH given in (CG1) is by means of (BS1), in which some causal connections presented by successful backtrackings are shown and AA_{CH} is made historically consistent accordingly, whereas the relation between the antecedent and the consequent still requires further justification. In (BS1), there are two new counterfactual conditionals, that $HC \boxrightarrow RS$ and that $RS \boxrightarrow AA_{CH}$. The anchoring historical fact HCM makes *HC* historically consistent with the actual course of history and *HC*, and HCM too, in turn make *RS* historically consistent. Undoubtedly, each '→' of a causal graph corresponds to a particular backtracking analysis of the target counterfactual conditional.

Reiss is bothered by how to ascertain (CG1) as more likely to be the case than the following:

(CG2) Causal Graph 2[15]

$\neg$HC

↘

$\neg$RS

↙ ↘

A_{CH} $\neg B_H$

According to (CG2), for Reiss, $\neg$RS is the common cause of both A_{CH} and $\neg B_H$, which means $(WC)^{CH}$ is false because, while (CH) is true in this case, A_{CH} is not a cause of $\neg B_H$. What brought about Hitler's proposing his Sudetenland demands, as per this analysis, is the insufficient rearmament of Great Britain. Since for Reiss there is no philosophical way to distinguish (CG1) from (CG2), he dissents from the idea

[15] This graph is a reasonable interpretation of Reiss' idea that Hitler might back down because of the military threat alone.

that finding Weber causation is essential to CASHs and historical judgements.

Although Reiss appears to have a point here, in the next section I will show that it is the result of lacking a more fine-grained theory of backtracking counterfactual analysis.

IV. CASH and the intervention-based backtracking analysis

There are two kinds of counterfactual backtracking analysis and only one of them suits the need of constructing a CASH. Consider the following popular instance:

> **Barometer**
> If the height of the column of mercury were different in the barometer, then there would be a different amount of rainfall. (Woodward & Hitchcock 2003: 7)

This conditional is normally considered false in the so-called non-backtracking manner of evaluation. For Woodward (2011: 412), the emphasis of the special status of the antecedent is that the counterfactual difference in the consequent does not rely on any difference issued from the precondition by which the difference in the antecedent is brought about, but only on the difference from the antecedent alone.[16] Suppose the precondition required to bring about the antecedent of the Barometer is to sabotage the barometer in order to change the column of mercury. This change would not bring about a different amount of rainfall.

Another way to read this counterfactual conditional appears to show that it is true. When observing the difference between the heights of the column of mercury, we backtrack to its precondition, the different atmospheric pressure, for the difference to arise. And the change in atmospheric pressure leads to a different amount of rainfall. The design of barometer is based precisely on this feature to help us foresee the change in the amount of rainfall by means of which we make a practical decision, say, whether to bring an umbrella.

In terms of the two readings, there are two respective causal graphs of Barometer:

(BCG1) Barometer's Causal Graph 1

sabotage barometer

↙ (↗)

[16] Wang and Hou 2014.

different height of mercury →?? different amount of rainfall

(BCG2) Barometer's Causal Graph 2

Different atmospheric pressure

↙(↗) ↘

different height of mercury different amount of rainfall

Now we can distinguish our backtracking symbols: '⇗' (or the like) represents an *i-backtracking* and'↗' (or the like) an *n-backtracking*. The common place of the two readings is to *backtrack to the precondition bringing about the antecedent*. The difference is obviously that they backtrack to different kinds of preconditions. If a counterfactual conditional is employed to focus specifically on the *causal effectiveness* between the antecedent and the consequent, it is *intervention-based*, and the precondition of the antecedent has no extra causal effectiveness on the consequent. Call this *i-backtracking*. For those counterfactual conditionals whose antecedents have extra causally effective preconditions, we may call this *n-backtracking*.[17] Clearly, (BCG1) is the *i-backtracking* reading and (BCG2) is the *n-backtracking*. And the Barometer counterfactual conditional represents the sort of special cases that both readings can work for them.

The distinction between the *i-backtracking* reading and the *n-backtracking* reading suits the present discussion extremely well. For any CASH, the right sort of analysis of the counterfactual conditional, such as (CH), embodied in the respective thought experiment is *i-backtracking*. This fits the discussion in the second section, that the historian employs her imagination to psychologically remove the historical event, which is the negation of the antecedent of the target counterfactual conditional, from the actual course of history and to see whether the consequent was likely to obtain. Obviously, the reasonableness of removing a particular historical event psychologically is the historical consistency of its counterfactual counterpart, and only an *i-backtracking* analysis suits the judgement of that. Now there is a more precise way to formulate historical consistency:

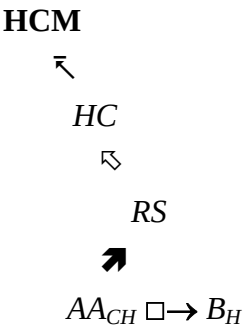
> A counterfactual antecedent is historically consistent, if and only if it is with an anchored *i-backtracking* analysis,

[17] For a more detailed discussion, see Wang and Hou 2014. Underpinning these discussions is the idea of causal model semantics, but strictly speaking, the presentation here is not and need not be causal modelling. Causal model semantics may contribute to the clarification of historical thought experiments and CASHs, which is worthy of further research, but the whole idea here is that the simpler current account already sufficed to do this job.

where an anchored *i-backtracking* analysis is an analysis the backtracking process of which ends with an anchoring historical fact.

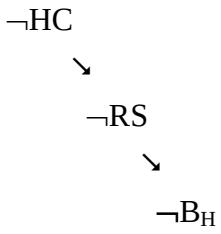
Based on the idea of an *n-backtracking*, we can see the corresponding backtracking structure of (CG2) as follows:

(BS2) Backtracking Structure 2



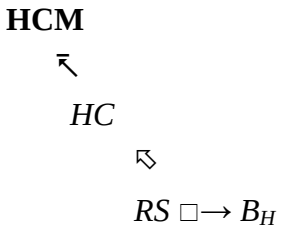
By doing the psychological removal, if the historian backtracks to a precondition that also brings about the consequent of the target counterfactual conditional—that is, it is an *n-backtracking* analysis—such as BS2, then the whole historical thought experiment so employed is defeated because the respective Weber causation is compromised. An *n-backtracking* analysis of (CH) is thus totally wrong-headed. She should rather replace her original conditional with a new one whose antecedent is the very precondition she backtracked to. The appropriate CASH is as follows:

(CG21) Causal Graph 2-1



And the backtracking structure is:

(BS21) Backtracking Structure 2-1



Therefore, Reiss' worry concerning $(WC)^{CH}$ in particular and Weber causation in

general results from a lack of a more fine-grained account of backtracking counterfactual analyses.

While the above account explains *what* Reiss' worry is and throws some light on the difficulty of ascertaining that a historical event is another's cause, some may still contend that it has little to say about *how* to ease the worry. The contention results from confusing Reiss' twofold worry. On the one hand, in terms of the relevant historical facts and the relevant background it may be very difficult to decide which causal structure presented by the respective backtracking analysis of the counterfactual conditional embodied in the thought experiment was the more likely—for instance, (CG1) or (CG2). On the other hand, because of this sort of difficulties, finding the Weber causation for a CASH is impractical, if not mistaken.

The first part is purely historical, and sometimes the gathered historical evidence is insufficient to determine whether or not the target Weber causation successfully captures a historical causal ascription. For instance, $(WC)^{CH}$ is false if the backtracking structure is (CG2). But it will be true if (CG1). The second part is philosophical. The proposal of difference making relations results from the employment of a less sophisticated account of backtracking analysis. The above more refined account of backtracking analysis provides a means, if not the means, to explain the philosophical aspect, and thus the philosophical prospect, of historical thought experiments and CASHs. Clearly, in the above instance, it is the *n-backtracking* feature of (CG2) that makes false $(WC)^{CH}$ and it is the *i-backtracking* feature of (CG1) that makes it true. A philosophical way is presented to ease Reiss' worry—that is, to construct a historical thought experiment and a corresponding Weber causation is to find a suitable *i-backtracking* analysis of the embodied counterfactual conditional(s). With a less sophisticated backtracking theory, the resultant conflation and thus the confusion seem inevitable. But once the theory is upgraded, the contention is shown to be groundless.

Given that the account of backtracking analysis is more fine-grained, (WC), and $(WC)^{CH}$ accordingly, needs a new formulation:

(WC-*i*) "*f* was a cause of φ" is assertable iff "$\neg f \,\Box\!\!\rightarrow \neg\varphi$" is *i*-assertable.

By '*i*-assertable', I mean the assertability of "$\neg f \,\Box\!\!\rightarrow \neg\varphi$" rests on an *i-backtracking* analysis.

Since my purpose is not to do historical study but to do philosophical research on the way to do a particular branch of historical study, CASH, namely, the focus is thus not about how to historically justify a CASH or the assertability of the target counterfactual conditional, but about how, in constructing a CASH relative to a particular context of history, the *i-backtracking* analysis contributes to the pondering of whether or not a given *i*-Weber causation is true, to the clarification of the exact causal structure of the CASH. To make the point clear, the backtracking analysis of an *i*-Weber causation has to be unconditionally *i-backtracking*. On this point, please consider the following backtracking structure:

(BS3) Backtracking Structure 3

HCM

↖

HC

↖

RS

↖

$AA_{CH} \square\!\!\rightarrow B_H$

The corresponding causal graph is as follows:

(CG3) Causal Graph 3

¬HC

↙ ↘

¬RS → A_{CH}

↘

$\neg B_H$

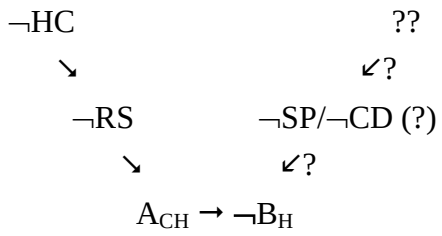
The (WC-*i*)^{CH} is false because (CH) is not *i*-assertable. The backtracking from *RS* (*the rearmament was sufficed*) to *HC* (*the hawkish cabinet*) is an *n-backtracking*, which means ¬HC turned out to be the common cause of ¬RS and A_{CH} (Chamberlin's appeasement policy). The correct analysis in this case is similar to the way of (CG21) and (BS21), and the resultant *i*-Weber causation will be true.[18]

[18] Another way to understand (CG3), it may seem to someone, is to take that ¬HC and that ¬RS to be some sort of compound cause of that A_{CH}. This is not the case. Put aside the above discussion of *n-backtracking*, what is crucial to the obtaining of that A_{CH} is that ¬HC. If we take off that ¬HC from the causal graph, since it is the cause of that ¬RS, it is gone too. Hence, that A_{CH} does not obtain. Moreover, if we take off that ¬RS from the causal graph, the causal connection from that ¬HC to that

As repeatedly emphasized, the historical entanglement does not concern the study of the account of backtracking analysis of counterfactual conditionals. The above comparison is meant to provide a basis for discussing how a backtracking analysis of counterfactual conditional deals with causal complication. Different from (CG2), where ¬RS is the common cause of both A_{CH} and $\neg B_H$ (Hitler's not backing down from his Sudetenland demands) and $(WC\text{-}i)^{CH}$ is thereby falsified, when the historian uncovered the backtracking to *RS* from AA_{CH} (*Had Chamberlin adopted an anti-appeasement policy*) to be defeated, she has yet to falsify her starting hypothesis of $(WC\text{-}i)^{CH}$. If she successfully established a new *i-backtracking* route from AA_{CH} to *HC* in (BS3), whose anchoring fact is still HCM (hawkish cabinet member(s)), the tottering $(WC\text{-}i)^{CH}$ was redeemed. Of course, if no such *i-backtracking* analysis can be made sense of, then it is indeed falsified. Although the historical part for finding and explaining an *i*-Weber causation may be often thorny, the philosophical part is not.

Apart from (CG1) to (CG3), there may be a case of overdetermination.[19] The discussion below assumes certain historical causal construction to explain how the *i-backtracking* account deals with this case. Consider the following:

(CG4) Causal Graph 4



The difference between (CG1) to (CG3) and (CG4) is apparently the introduction of the German political and military reality at that time into the consideration.[20] "¬SP" represents that no Hitler's political opponent intended to seize power and "¬CD" represents that no German generals conspired a *coup d'état*. Historians certainly can pursuit different routes to study possible causal connection between given historical events relative to a particular context of history. Perhaps in the present case, $\neg B_H$ was partially caused by ¬SP and ¬CD,

---

A_{CH} is intact. So causal-wise, only that ¬HC matters.

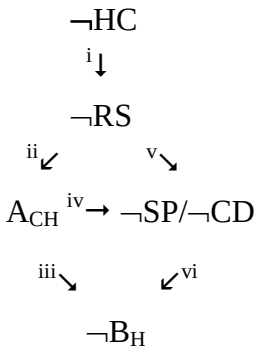
[19] Although an overdetermination is metaphysical, the whole point of discussing this pondering is entirely epistemic.

[20] The following discussion concerning how the delicacy of the political and military reality of Germany at that time might be triggered by Great Britain's stance towards Hitler's demands comes from Reiss 2009. But the supposition in the discussion is mine. For the sake of simplicity, it is seen that ¬SP and ¬CD, or the respective counterfacts, occurred jointly with no separate causes.

which were caused in turn by another historical event or events. Does this sort of overdetermination cause trouble for the backtracking account, such that (CH) is assertable because of the wrong reason, that AA_{CH} and *SP/CD* are both historically consistent? The answer is no. As far as their *i-backtracking* analyses are mutually *independent*, the worry would not arise in the first place. Note that Weber causation and CASH are just about *a* cause of a historical event. Whether or not (CH) is assertable is sealed in its own evaluative context consisting of its *i-backtracking* analysis and the respective historical judgement on the occurrence of the consequent in this context. The judgement on (CH) would not be contaminated by other parallel evaluative contexts. Of course, if overdetermination occurs within a backtracking analysis, then at least one *n-backtracking* can be found.

History is complicated. Perhaps most CASHs are not as simple as $(\text{WC-}i)^{\text{CH}}$. Nevertheless, the *i-backtracking* counterfactual analysis and *i*-Weber causation are supposed to be able to explain them. Consider (CG5) as an example:

(CG5) Causal Graph 5



Under possible stimulations of the rapid rearmament of Great Britain and Chamberlin's unyielding diplomatic attitude, Hitler might back down from his demands because his political rival might intend to seize power and some of his generals might conspire a *coup d'état*. Constructed as such, it is still a case of overdetermination. (i)-(ii)-(iii) is the usual $(\text{WC-}i)^{\text{CH}}$. The causal structure of (ii)-(iv)-(v) is similar to (CG3) and should rather be replaced by one that is similar to (CG21). So settled, the supposed causal relation between A_{CH} and ¬SP/¬CD is cancelled and the overdetermined causes of $\neg B_H$ are A_{CH} and ¬SP/¬CD. The respective causal structures are (i)-(ii)-(iii) and (i)-(v)-(vi). The respective hypothetical counterfactual conditionals are (CH) and that *SP/CD* $\square\!\rightarrow B_H$. And finally the two respective *i*-Weber causations are $(\text{WC-}i)^{\text{CH}}$ and $(\text{WC-}i)^{\text{SPCDBH}}$ ("No political and military rivals (¬SP/¬CD) is a cause of Hitler's not backing down from his Sudetenland demands" is assertable if and only if

"*SP/CD* □→ B_H" is *i*-assertable).

A historical causal structure between historical events may be complicated, may be overdeterminated, pre-empted, or trumping, and so on. If for instance (CG5) is *one* construction of the causal structure of all those mentioned historical events, then (ii) and (v) indicate an *n-backtracking*, and overdetermination, and thus make false the respective *i*-Weber causation. Even if (i)-(ii)-(iii) and (i)-(v)-(vi) are constructed independently through respective *i-backtracking* analyses, the historian(s) may still wonder whether or not they are really independent in the actual course of history and will therefore conduct further research regarding their relation(s). The account of *i-backtracking* analysis of counterfactual conditionals and *i*-Weber causation can assist in building up the historical consistency of historical events (or rather counterfacts) occurring in the entire backtracking process piece by piece epistemically. That which is complicated is history, not the philosophical account of how to construct backtracking analysis of counterfactuals employed in historical study.

V. Some epistemological remarks

1. For Woodward and Hitchcock, for instance, the *i-backtracking* analysis is to backtrack to the precondition of the counterfactual antecedent by which the antecedent is brought about in an intervention manner. Interventions as such are not psychological. Perhaps it is more appropriate to consider the above account to be about *quasi-i-backtracking* analysis rather than *i-backtracking*.

2. The searching for a particular CASH is professional and *hypothetical*, as is with the research of other fellow social scientists. In judging whether or not "Chamberlin's appeasement policy was a cause of Hitler's not backing down from his Sudetenland demands (or even encouraging it)", or "A_{CH} was a cause of $\neg B_H$," Reiss holds that, because *RS* may be the common cause of AA_{CH} and B_H, $(WC)^{CH}$, the CASH between Chamberlin's appeasement policy and Hitler's satisfaction of Sudetenland demands, is false. This is correct if *RS* indeed was the common cause. However, the target CASH in this case is just a hypothesis and the respective thought experiment is to exercise the historian's constructing power and her mastering of relevant historical facts to test this hypothesis. When her conclusion is that *RS* was the common cause, she indeed rejects the hypothesis that A_{CH} was a cause of $\neg B_H$. But in the meantime, she finds herself a new hypothesis through the backtracking process, and a new *i*-Weber causation accordingly, that the insufficient rearmament was a cause of Hitler's proposing his Sudetenland demands. This is supposed to be the normal case of

academic research, history included, and the account of *i-backtracking* analysis of counterfactuals provides the needed means. In this respect, this account can be seen as providing the formal ground for the epistemology of historical thought experiments.

3. The difference-making relations promoted by Reiss are too weak and sometimes misleading. Take the case of $(WC)^{CH}$ again. If the aspiration of finding a $(WC)^{CH}$ is discouraged in favor of finding a difference-making relation, then no distinction can be made between (CG1), (CG2), and (CG3). In a difference-making sense, they are indistinguishable. Even if insufficient rearmament were a cause of Hitler's not backing down, and the anti-appeasement policy would not be implemented without first establishing the rearmament, Reiss still holds that it is inessential to find a Weber causation and that a difference-maker is sufficient. All *i*-Weber causations entail some respective difference-making relations. In the present case, (CG1), (CG2), and (CG3) make the exact same difference for Reiss, viz. Chamberlin adopting anti-appeasement policy and Hitler backing down from his Sudetenland demands (B_H). But they tell different causal stories. The only difference making relation allowed by *i*-Weber causation is (CG1). According to the present account, (CG1) is acceptable because its corresponding backtracking structure is *i-backtracking*, whereas (CG2) and (CG3) are not acceptable because of their respective *n-backtracking* structures. Nevertheless, relative to the modifications of their backtracking structures, the modified *i*-Weber causations of (CG2) and (CG3) can be constructed respectively.

CASHs have many functions. Even if historical evaluation of a particular historical event or a particular character is the main purpose of constructing a CASH, different evaluations are presumably given, at least partially, by (CG1), (CG2), or (CG3). It would be sheer coincidence that Chamberlin was responsible for slowing the rearmament and for adopting the appeasement policy. And therefore it would be sheer coincidence that a particular difference-making relation suffices to make an appropriate evaluation of him. Suppose Chamberlin was solely responsible for the appeasement diplomacy while a different politician was responsible for the insufficient rearmament, and further suppose that what would make a difference to Hitler's stance on the Sudetenland demands was Chamberlin's diplomacy—that is, Hitler would not have backed down even if Great Britain had sufficient military force—then it would be (CG1), but not (CG2) or (CG3), that suitably capture *the* difference-making relation. Of course, if the case is different in that it is the sufficient rearmament that would make Hitler back down, then it is (CG2). However, Reiss' account tells us differently. Prominently, in term of the *i-backtracking* account, the different causal structures given by the different backtracking analyses of them makes

a difference here. The prospect of the reasonableness of Reiss' difference-making account is evidently dawned to get worse when the scale of the historical thought experiment is enlarged.

4. The account of CASHs, *i*-Weber causation, significantly shapes the way historical thought experiment(s) is constructed for the purpose of testing some historical causal hypotheses—that is, some target counterfactual conditional(s). Except the constraint of how far-off a backtracking analysis of counterfactual conditionals can extend, normally the more far-off the less reasonable, *i-backtracking* analyses further restrict the way historians do their backtracking analyses. Any finding of *n-backtracking* means failing a hypothesis. Since this account is thoroughly after *a* causal structure of historical events, relative to a particular context of history, constructed by the employment of historical thought experiments to test various interlocked counterfactual conditionals, if the forgoing has its limit of how far-off the backtracking process is tolerated, this also applies to the target counterfactual conditional of the target *i*-Weber causation, such as (CH). In other words, the causal connection between its antecedent and its consequent cannot be too far-off either.

Take (CG1)/(BS1) as an example. The historical consistency of AA_{CH}, the antecedent of (CH), depends on the respective two counterfactual conditionals, that $RS \mathrel{\square\!\!\rightarrow} AA_{CH}$ and that $HC \mathrel{\square\!\!\rightarrow} RS$, and the two respective *i*-Weber causations, $(\text{WC-}i)^{\text{RSAACH}}$ and $(\text{WC-}i)^{\text{HCRS}}$. The too far-off causal connection between *RS* and AA_{CH} or between *HC* and *RS* would cost the historical consistency of AA_{CH}. Considering as such, taking an *i*-Weber causation as a hypothesis would be inappropriate if the potential causal connection between the antecedent and the consequent is too far-off. One of Reiss' three cases is the following:

> Had the Greeks not won against the Persians at Salamis, Western civilization would not have become dominant in the world.[21] (Reiss 2009: 713)

No *i*-Weber causation can be found for it because *i-backtracking* analysis is completely trivialized. No precondition of the antecedent can possibly have extra causal power on the consequent. No causal connection between the antecedent and the consequent can be constructed, except for a far-fetched story.

5. Relying on counterfactual conditionals to construct *i*-Weber causation has its

[21] This example comes from Victor Davis Hanson, "A Stillborn West? Themistocles at Salamis, 480 BC" (in Philip Tetlock, Richard Ned Lebow, and Geoffrey Parker (eds.), *Unmaking the West: "What-If?" Scenarios That Rewrite World History* (2006). Ann Arbor: University of Michigan Press, 47–89).

semantic consequences. Two semantic features of counterfactual conditionals affect the implementation of historical thought experiments. The first pertains to non-transitivity. Consider (CG1) and (BS1). The sensibility of (CG1) is vindicated by three *i*-Weber causations, (WC-*i*)^{CH}, (WC-*i*)^{RSAACH}, and (WC-*i*)^{HCRS}. Their truth in turn relies on the assertability of three respective counterfactual conditionals, that $AA_{CH} \mathrel{\square\!\!\rightarrow} B_H$, that $RS \mathrel{\square\!\!\rightarrow} AA_{CH}$, and that $HC \mathrel{\square\!\!\rightarrow} RS$, and two successful *i-backtracking*s. Since there is no transitivity for counterfactual conditionals, even if the above three counterfactual conditionals are taken to be true, that $RS \mathrel{\square\!\!\rightarrow} B_H$ and that $HC \mathrel{\square\!\!\rightarrow} AA_{CH}$ cannot be derived. This means no further construction of the causal structure, no more *i*-Weber causation, can be made. Moreover, this is actually the natural consequence of the *i-backtracking* account. For instance, it does not follow from that it is assertable that $AA_{CH} \mathrel{\square\!\!\rightarrow} B_H$ and that it is assertable that $RS \mathrel{\square\!\!\rightarrow} AA_{CH}$, that it is assertable that $RS \mathrel{\square\!\!\rightarrow} B_H$. The assertability of counterfactual conditionals is not transitive. Furthermore, if the counterfactual conditionals of that $RS \mathrel{\square\!\!\rightarrow} B_H$ or that $HC \mathrel{\square\!\!\rightarrow} AA_{CH}$ can be constructed in the backtracking analysis of (CH), then some *n-backtracking* must occur in the backtracking structure of (CH), and (WC-*i*)^{CH} is falsified accordingly.

The consequence of this is the non-transitivity of singular historical causation. That is, for instance, although according to (WC-*i*)^{CH}, A_{CH} is a cause of $\neg B_H$, and according to (WC-*i*)^{RSAACH}, $\neg RS$ is a cause of A_{CH}, $\neg RS$ is not a cause of $\neg B_H$. Lewis (2000) tried to fix the causal bridge broken by counterfactual dependency to redeem causal transitivity.[22] However, even if metaphysically speaking there is a proper way to explicate what a causal chain is, epistemically and historically we cannot make equal sense of it. For the theory of justification, it is a controversy over the acceptance of the closure principle that if it is justified that $p \rightarrow q$ and if it is justified that $q \rightarrow r$, and because that $p \rightarrow q$ and that $q \rightarrow r$ logically imply that $p \rightarrow r$, it is justified that $p \rightarrow r$. Nevertheless, it is not a controversy for the present case but impossibility. Even though it is justified that $p \mathrel{\square\!\!\rightarrow} q$ and it is justified that $q \mathrel{\square\!\!\rightarrow} r$, but because of the non-transitivity of counterfactual conditionals—that is, that $p \mathrel{\square\!\!\rightarrow} q$ and that $q \mathrel{\square\!\!\rightarrow} r$ do not imply that $p \mathrel{\square\!\!\rightarrow} r$—it is not justified that $p \mathrel{\square\!\!\rightarrow} r$. This means that, metaphysically speaking granted, $\neg RS$ is a cause of $\neg B_H$ because A_{CH} is a cause of $\neg B_H$ and $\neg RS$ is a cause of A_{CH}, it is in principle not justified that $\neg RS$ is a cause of $\neg B_H$ for the reason just explained. In other words, given the above account of

[22] The basic idea of that *c* causes *e* rests on a chain of stepwise influence from *c* to *e*. And by 'a chain of stepwise influence', it is saying that a range of c_1, c_2… of different not-too-distant alterations of *c* (including the actual alteration of *c*) and a range of e_1, e_2… of different not-too-distant alterations of *e* are such that, if c_1 had occurred, e_1 would have occurred, and if c_2 had occurred, e_2 would have occurred, and so on. Even if it is successful, it does not affect the above point.

backtracking analysis, it is in principle not for a given historian to know. Put in another way, the assertability of historical causal ascriptions is not transitive. Moreover, if in the above case that $\neg$RS is a cause of $\neg B_H$ is known (or justified) and therefore assertable, then according to the backtracking account the original backtracking analysis of (WC-*i*)^{CH} is not *i-backtracking* and thus is in need of revision.

Some may propose a modified version of Reiss' difference-making account, claiming that the precondition of a counterfactual antecedent makes a difference to the consequent. That is, for a given causal ascription there is a chain of difference making. However, to facilitate this proposal a chain as such has to be *i-difference-making*, and the exact problem stated above recurs. This finalises the discussion of non-transitivity, that my backtracking account works well with it and that the metaphysical conception of a chain of causes does not have its counterpart in historical causal ascriptions.

Secondly, counterfactual conditionals are strong-centring, non-monotonic, meaning that adding more information to the antecedent may render an assertable counterfactual conditional unassertable. Suppose *RS* brings about AA_{CH} with no extra causal effectiveness on B_H and suppose (CH) is, therefore, assertable. The assertability of (CH) may be ruined if, for example, the antecedent were to be enriched by the following: (1) no political rival of Hitler was able to seize power and (2) all German generals were loyal to Hitler. For no trouble back at home might actuate Hitler to pursuit the achievement of his rapacious ambition even in the face of the British military and diplomatic threat. In other words, even if Chamberlin's appeasement policy were a cause of Hitler's not backing down according to the respective *i*-Weber causation, it would not be such a cause if both (1) and (2) were featured in the consideration. This is not surprising. With an enriched antecedent, along with further historical facts taken as evidence, the original *i*-Weber causation, as a hypothesis, can of course be changed.

Any given CASH is epistemic, and its reasonableness rests on the historian's judgements on what historical facts are relevant, what generalisations are utilised in the background, which counterfactual conditional is hypothesized, and whether or not the counterfactual conditional is *i*-assertable. Non-transitivity and strong-centring show the constraints of this epistemic route of historical causal constructions.

6. Last but not least, the epistemology of CASHs, including the construction and evaluation of historical thought experiments and *i*-Weber causations, is empirical. Any

new relevant evidence may defeat a particular *i-backtracking* analysis of the target counterfactual conditional of the respective *i*-Weber causation or help delete some competing but yet unsettled *i-backtracking* analyses. Moreover, CASHs of personal histories can be similarly analysed and constructed. Finally, an empirical semantics of *backtracking* analysis may be given by means of causal model semantics.

References:

Khong, Yuen Foong (1996). Confronting Hitler and Its Consequences, in Philip Tetlock and Aaron Belkin 1996: 95–118.

Lewis, David (1973). Causation. *Journal of Philosophy* 70 (8): 556–567.

Lewis, David (1979). Counterfactual Dependence and Time's Arrow. *Noûs* 13 (4): 455–476.

Lewis, David (2000). Causation as Influence. *Journal of Philosophy*, 97: 182–97

Reiss, J. (2009). Counterfactuals, Thought Experiments, and Singular Causal Analysis in History. *Philosophy of Science* 76: 712-723.

Tetlock, Philip, and Belkin, Aaron (1996) Counterfactual thought experiments in world politics: logical, methodological, and psychological perspectives, in Tetlock and Belkin eds. (1996), *Counterfactual Thought Experiments in World Politics: Logical, Methodological and Psychological Perspectives*. Princeton, NJ: Princeton University Press: 1-38.

Wang, Linton & Hou, Richard (2014), "Conditionals and Deliberations". Manuscript.

Weber, Max ([1905] 1949), "Objective Possibility and Adequate Causation in Historical Explanation", in *The Methodology of the Social Sciences*. Edited and translated by Edward Shils and Henry Finch. Glencoe, IL: Free Press, 164–188. Originally published in *Archiv für Sozialwissenschaft und Sozialpolitik.*

Woodward, James (2011), "Mechanisms Revisited," *Synthese*, 183, 409-427.

Woodward, James & Hitchcock, Christopher (2003), "Explanatory Generalizations, Part I: A Counterfactual Account," *Noûs*, 37, 1-24.

The notion of interaction and the nature-nurture debate

Abstract: The method of heritability analysis used in biometrics of ascribing causal responsibility to the genes has long been criticized by interactionism. This is usually called "the nature-nurture debate". James Tabery proposes that the controversy arises from an unnoticed explanatory divide between biometrics and interactionism, which results in two concepts of interaction. This paper aims to examine whether the clarification of the explanatory divide can resolve the disputes surrounding the debate. First, the statistical basics of heritability analysis and how its results are interpreted to represent the magnitude of genetic cause is briefly presented. Second, I distinguish three notions of interaction in biometric context, namely, trivial, vernacular and statistical interaction, according to which different disputes might arise. I argue that Tabery's position works only for the disputes over trivial interaction and vernacular interaction, but not for that on statistical interaction. Finally, I revisit Tabery's case study between Hogben and Fisher, and argue that they concern exactly the notion of statistical interaction. Their dispute amounts ultimately to an empirical question of whether the causal interdependence between genes and the environment reflected by statistical interaction is common in nature, which is still unresolved. The conclusion is that apart from the disputes arising from the explanatory divide proposed by Tabery, there is another deeper dispute over the notion of statistical interaction in the nature-nurture debate.

Key words: nature-nurture debate; ANOVA; heritability analysis; statistical interaction

1. Introduction

The nature-nurture debate has been raised for several decades in philosophy of biology. It concerns mainly with the method of heritability analysis used in classical quantitative genetics. The basic idea of the method is to partition the genetic causes from environmental causes for the phenotypic variation in a population. By using the analysis of variance (ANOVA), heritability, which represents the genetic proportion of the total phenotypic variance, is estimated (Falconer and Mackay 1996). The controversy lies in the interpretation of the results in causal terms.

A famous example of the nature-nurture debate is between Arthur Jensen and Richard Lewontin happened in 1970s. In the article "How much can we boost I.Q. and scholastic achievement?", Jensen, by using heritability analysis, claims that "something like 80 percent of the variance in IQ has a genetic basis" (Jensen 1969, p.7). He then infers that it is a fact that intellectual ability is largely determined by heredity, and education policies should be made on this fact in order to boost intelligence in the population. Jensen's view has been criticized vigorously, among which Lewontin's article "The analysis of variance and the analysis of causes" (1974 [2006]) represents a classic. Lewontin argues that the heritability analysis based on ANOVA fails to reflect the relative importance of genes whenever there is non-additive relation between genetic and environmental effects (ibid. p.524). Since ANOVA is applied to particular populations sampled, we cannot presume that *overall* the additivity assumption is met. Hence the heritability estimates cannot reflect the actual genetic influence on phenotypic variation in general.

Here, the focal question is, "for a certain trait such as human IQ, is it mostly determined by genes or by the environment?" The debate is on whether this question can be answered by heritability analysis. In biometric fields including classical quantitative genetics, behavioral genetics and epidemiology, heritability analysis is widely used to study measurable biological traits, which is thought to be able to answer the question. As Jensen, a representative for those who called "biometricians", would claim, if the heritability of IQ (the phenotype) is high, then it is mainly determined by the genes or the nature. However, others disagree and argue that heritability analysis is not a valid method to reveal the actual causes for phenotypic variation. Lewontin has been a representative for so-called "interactionism" supported mostly by developmental biologists and molecular biologists (Perbal 2013). The interactionists claim that genes and the environment are interacting in determining the phenotype, and the idea of separating two causes must be misleading.

More recently, James Tabery's book "Beyond versus: the struggle to define the interaction of nature and nurture" (2014) seems to shed a new light on the nature-nurture debate. To recapitulate his idea, suppose someone wants to know why some people develop a complex trait such as depression (example borrowed from Griffiths & Tabery 2008, p.9). This inquiry incorporates two different research questions. First, what are the causes of variation responsible for *individual differences* in a particular population? Second, what are the causes of the mechanism responsible for the *individual development* of depression? Different questions require different explanations. The first question is the focus of biometricians. They want to know why certain people while not others in the population have depression, especially the magnitude of the importance of the genes. The second question which is the focus of interactionists aims at understanding how a particular genotype, given a certain environment, produce depression during developmental processes. The goal is to find out how different factors interact in development. In this case, the genetic and environmental factors should not be considered separately.

Tabery thus claims that there is an unnoticed explanatory divide: biometricians adopt a "variation-partitioning approach" searching for the causes of variation at the population level, while interactionists adopt a "mechanism-elucidation approach", focusing on the causal mechanisms in individual development (Tabery 2014, p.5). He also argues that this explanatory divide "*shaped* divergent understandings of the interaction of nature and nurture" (emphasis added) (ibid. p.6), namely, a biometric concept for the variation-partitioning approach and a developmental concept of interaction for the mechanism-elucidation approach (ibid.). The former is defined by statistical features and is regarded as "an obstacle" to properly using ANOVA, while the latter is defined by mechanistic features and is regarded as "a valuable insight" in illustrating developmental mechanisms (ibid.). Although both biometrics and interactionism use the same word "gene-environment interaction", they express different concepts of meaning and have been talking past one another.

Tabery's idea is insightful. But from the above quote and elsewhere in his book[1], it is not quite clear what his position would lead with regards to the nature-nurture debate. In other words, could his clarification of the explanatory divide and the distinction of two concepts of interaction resolve

[1] The remaining of his book is mostly on how to integrate these two approaches.

the disputes surrounding the debate? Tabery's book with the title "Beyond versus" seems to give a strong impression that the answer is "Yes". As he said, "Members of the variation-partitioning approach and members of the mechanism-elucidation approach consistently conceptualized interaction differently, promoted different means of investigating it, and offered different assessments of its prevalence in nature." (Tabery 2014, p.7) However, this is highly doubtful. It is possible that there were controversies from conceptual misunderstandings of the notion of interaction between two parties, but it is hard to imagine that all this time, with many experts involved, the different concepts of interaction are not clarified, which consistently lead to opposite attitudes toward the nature-nurture question.

In fact, in the 1974 article just mentioned, Lewontin distinguishes two problems of causation at the beginning of this article. One problem is to regard genes and the environment as "two alternative and mutually exclusive causes" (p.520) and to ask which one causes the phenotype. The other problem is to recognize that all phenotypes are the results of two "interacting causes" of genes and the environment, and to ask the relative importance of these two factors. As Lewontin remarks, "the pseudo-question of 'nature versus nurture' arose precisely because of the confusion between these two problems in causation" (ibid.). After clarifying this confusion, Lewontin focuses on the second problem for the rest of the article. This means that the real target of Lewontin's article is a more sophisticated question, which can be constructed as "for a certain trait such as human IQ, is the variation of the trait in the population mostly determined by genes or by the environment?" And a more sophisticated debate is rather on whether this question can be answered by heritability analysis.

For Lewontin and those who commented on his article in 2006 when it was reprinted on *International Journal of Epidemiology* (Rose 2006; Taylor 2006; Vreeke 2006; Freese 2006; Lewontin 2006), they seem to understand fully that the heritability analysis is within what Tabery called the variation-partitioning approach. This indicates that Tabery's diagnosis of the nature-nurture debate is hardly enough and may be even misleading. This paper aims to examine whether Tabery's diagnosis is correct, in particular, to see if there is a fundamental problem regarding the notion of interaction for heritability analysis in answering the sophisticated nature-nurture question. The overall goal is to render the notion of interaction clearer as well as to clarify the role it plays in the nature-nurture debate.

One thing should be noted. According to James Woodward's manipulation theory of causation, a causal relation is defined when the value of one variable makes a difference to the value of another (Woodward 2003). In classical quantitative genetics, a genetic cause is identified when the variation in a gene lead to phenotypic variation in a population (Waters 2007, p. 7). In developmental biology, the genetic cause of a trait is identified by intervening the genes to see the difference it made in experimental conditions. Therefore, both correspond to the difference-making relation of causation. I will use this theory of causation throughout the paper.

The paper is constructed in three parts. First, the basic rationales behind heritability analysis and how it can reflect the magnitude of genetic cause on the phenotypic variation is briefly presented.

Then I distinguish three notions of interaction in the biometric context, namely, trivial, vernacular and statistical interaction. Given two variables influencing the value of a phenotype, trivial interaction refers to the statistical measure that the change of one variable *would not affect* the effects brought up by the other variable, vernacular interaction with additively affect, and statistical interaction with non-additively affect. I argue that Tabery's diagnosis only works for the disputes on trivial interaction and vernacular interaction, but not for that on statistical interaction. Finally, I revisit Tabery's case study of Hogben and Fisher. I argue that rather than adopting two concepts of interaction, the dispute between Hogben and Fisher is on the interpretation of statistical interaction, and ultimately on an empirical question of whether statistical interaction is common in nature. The conclusion is that apart from the confusion arising from the explanatory divide proposed by Tabery, there is another dispute over the interpretation of statistical interaction and over its validity to answer the more sophisticated question.

2. ANOVA and heritability analysis

The method of ANOVA was first introduced by R. A. Fisher with his studies on crop variation around 1923. It has become the basic method for classical quantitative genetics. According to the standard model of quantitative genetics, the phenotypic variance (V_P) of a population can be explained by two components, its genotypic variance (V_G) and its environmental variance (V_E). In its simplest form, we have:

$$V_P = V_G + V_E \quad (1)$$

From there broad-sense heritability (H^2) is defined. It measures how much of the total phenotypic variation is due to genotypic variation:

$$H^2 = \frac{V_G}{V_P} \qquad (2)$$

Here, V_G can be partitioned into the additive genetic variance (V_A), the dominance genetic variance (V_D) and the epistasis genetic variance (V_I), where V_A is the variance due to alleles being transmitted from the parents to the offspring that contribute to the phenotype. V_D is the variance due to combinations between alleles at one locus for diploid organisms, and V_I is the variance due to combinations between alleles from different loci. Since genotypes of sexual organisms recombine at each generation via reproduction, the effects of combinations of genes, namely dominance and epistasis effects are not transmitted across generations. By only taking V_A into account, narrow-sense heritability (h^2) that "expresses the extent to which phenotypes are determined by the genes transmitted from the parents" (Falconer and Mackay 1996, p.123) is defined:

$$h^2 = \frac{V_A}{V_P} \qquad (3)$$

Narrow-sense heritability (h^2) is used in breeding studies and by evolutionary theorists who are interested in making evolutionary projections, while broad-sense heritability (H^2) is most used by behavioral geneticists and psychologists (Downes 2015). Since the nature-nurture debate concerns mostly human behaviors and human mental traits, I use "heritability" to refer to broad-sense heritability here.

Traditionally, before the genome-wide association studies (GWAS), there is no direct way to measure the variation of genotypes. The data for doing ANOVA are measureable phenotypic values with assumed resemblance of genotypes. Supposing an ideally simply example, the individuals of a plant population have difference heights, and they are determined by the genotype of a certain locus and the temperature of the environment. Table 1 gives a set of hypothetical data of height values for the combinations of two genotypes (G1 and G2) and three temperatures (E1, E2 and E3), as well as means of genotypes (Gm), means of environments (Em) and the overall mean (M).

Table 1. A standard example

	E1	**E2**	**E3**	**Em**
G1	5	10	15	10
G2	15	20	25	20
Gm	10	15	20	**M =15**

To apply ANOVA, V_P is calculated by squaring every individual deviation from the overall mean, adding them up, and dividing the result by the number of individual cases (six combinations). To calculate V_G, we square the deviations of "Gm" from "M", add them up, and divide it by the number of three (three Es). Similarly, we get the value for V_E by squaring the deviations of "Em" from "M", adding them up and dividing it by two (two Gs). The calculations are as follows:

$$V_P = [(5 - 15)^2 + (10 - 15)^2 + \ldots + (25 - 15)^2]/6 = 41.7$$

$$V_G = [(10 - 15)^2 + (15 - 15)^2 + (20 - 15)^2]/3 = 16.7$$

$$V_E = [(10 - 15)^2 + (20 - 15)^2]/2 = 25$$

The heritability can be estimated by applying the numbers to Equation (2):

$$H^2 = \frac{V_G}{V_P} = 16.7/41.7 = 0.4$$

For the population sampled, the contribution from the genotypic variance to the phenotypic variance is 0.4 or 40 percent. That is to say, the height variation in the population is explained as being caused 40 percent by the difference of genotypes or the nature. As to the question of "how much of the height variation is caused by genetic influence in the population", the answer here is 40 percent.

Three points should be noted. First, ANOVA is applied in the population level and hence the heritability estimated is not a measure of the contribution of genotype in any individual level. That said, the heritability analysis provides no direct causal explanation for individual developmental processes. This point is emphasized by Tabery when he distinguishes the variation-partitioning approach on the population level from the mechanism-elucidation approach on the individual level. Confusions surrounding this point can thus be cleared by Tabery's diagnosis.

Second, heritability analysis is relative to the population studied in specific environments, that is, it is fundamentally *local*. ANOVA was originally used for breeding experiments by Fisher in order to improve crop yields, and subsequently became popular in selective breeding in agriculture and laboratories. There are usually specific genotypes and environments that experimenters are interested in, and the goal is to find out the best combination(s) of the genotype and the environment in order to get the optimal phenotypic production. For this kind of scientific practice, it is not required for a general or *overall* explanation across all populations. If someone makes a further generalization and applies it to other populations such as Jensen did, then this would be a mistake. This point has been one major claim made by Lewontin (1974 [2006]). The lesson is that the nature-nurture question, if it can be answered by the heritability analysis, should be specified to particular populations. Hence, the precise question should be "how much of the height variation is caused by genetic influence in the targeted population".

Thirdly, the heritability analysis assumes that the contribution of genes and the environment are additive. If there are non-additive effects, then the causal interpretation of the results of ANOVA would be groundless. This point is extensively discussed by Lewontin (1974[2006]), and it is the other major claim made to criticize people like Jensen. To dig deeper into this point, I focus on the notion of interaction in next section. Because the debate mainly concerns the legitimacy of biometric methods, I will analyze the notion of interaction in the biometric context.

3. Three notions of interaction

"Interaction", according to Merriam-Webster dictionary, means mutual or reciprocal action or influence. To apply this definition in ANOVA, interaction means that the phenotypic variation in a population results from the mutual influence of two variables of genes and the environment. Recall the ideally simply example in Section 2, the plant heights of a population are determined by a certain gene and the temperature of the environment. Assuming that there is a well-defined range of temperature, two hypothetical genotypes and phenotypic measurements. If the change of one variable makes no difference to the phenotype given the other variable, then there is no mutual influence between these two variables, hence no real interaction in biometric context. An illustration using reaction norm graphs is presented in Figure 1.

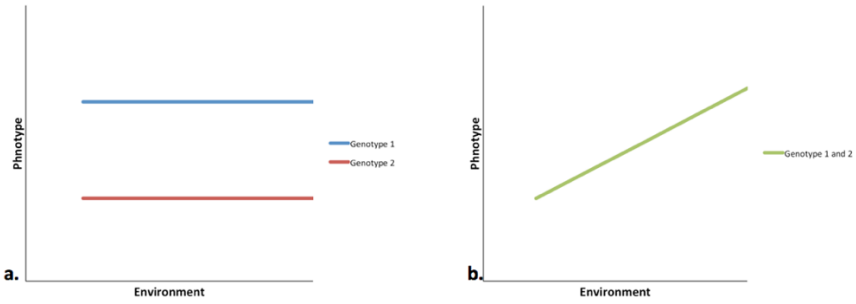


Figure 1. Trivial Interaction

For both Genotype 1 and Genotype 2, changes in the environmental variable has no phenotypic effects (Figure 1a). Given a range of environment, the change between Genotype 1 and Genotype 2 brings no effect on the phenotype, either (Figure 1b). Figure 1a and 1b represent two deterministic cases, the former for genetic determinism and the latter for environmental determinism, each of which suggests that the phenotypic variation in the population is influenced by only one variable.

This, however, as mentioned before, does not mean that the phenotype is caused solely by one variable in individual development. The nature-nurture debate may be seen as an unsophisticated confusion ("the pseudo-question of 'nature versus nurture'" in Lewontin's words) when by gene-environment interaction one means that both genes and the environment are necessary for the production of a phenotype. But surely, no sensible biologists will suggest that genes alone can give rise to a biological trait. If interactionists understand interaction in this sense, then the dispute becomes trivial. That is why I term this notion "*trivial interaction*".

Although trivial interaction is not a meaningful interaction notion for biometricians, interactionists who are taking the mechanism-elucidation approach values this notion more. In the description of a mechanism, DNA replication for example, the specific location, the cellular environment and the functional proteins are all required to be identified (Machamer et al. 2000). It is in the sense that identifying the necessary components for a mechanism that those components "interact" with each other in producing a phenomenon. The different attitudes taken by biometricians and interactionists toward trivial interaction do sometimes raise such a confusion. Here, Tabery's diagnosis can work to clarify this confusion. This is also the reason that the authors involved in the discussion of the nature-nurture debate usually make explicit clarifications in their writings (for example see Jensen 1969; Lewontin 1974 [2006]; Sesardic 2005).

A conceptually richer form of interaction in biometric context is that the change of one variable can bring an additive change to the phenotype regarding to the other variable. Following Sterelny and Griffiths (1999, p.15), I term it "*vernacular interaction*". An example of using reaction norm graph is as in Figure 2. Given a genotype, the change of the environmental variable makes a difference in the phenotype. In this case, the phenotypic value is influenced by both genotypic and environmental variables. The two lines in Figure 2 are parallel, which means that the effects of genotypic and environmental variables are additive.

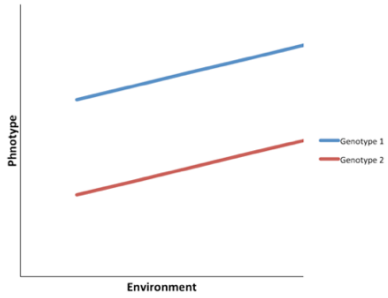


Figure 2. Vernacular Interaction

In fact, the lines of Figure 2 correspond to the data of the standard example of heritability analysis showed in Table 1. Recall the standard example, given the phenotypic value of (G1, E1). When the environmental variable changes from E1 to E2, the height value is raised by 5 points. And the genotypic change from G1 to G2 leads to the increase of the phenotype by 10 points. The sum of two changes is 15 points, which equals exactly to the result of the number of (G2, E2) minutes (G1, E1). The equal relation applies to any combinations of one environmental change and one genotypic change in the example. In this case, the effects of two variables are said to be additive.[2] As we already did ANOVA for this example in Section 2, V_P is 41.7, V_G is 16.7 and V_E is 25. The sum of V_G and V_E equals exactly to V_P, which corresponds perfectly to Equation (1).[3]

As we can see, vernacular interaction meets the requirement of additivity assumption. It is the basic biometric relation of genes and the environment for interpreting the results of heritability estimates. Hence, the notion of vernacular interaction does not posit a challenge for biometric methods in the more sophisticated debate. The criticism from interactionism who relies on the notion of vernacular interaction can be dismissed by asserting that, as Tabery has done, biometricians use heritability analysis to find the causes of variation in the population, and this approach is different from the mechanism-elucidation one adopted by interactionism (a similar strategy is proposed by Sesardic (2005, Section 2.1)).

However, if the effects of genes and the environment are non-additive, the situation changes dramatically. Non-additive interaction is the notion that has been used in the biometric tradition for population genetics (Fisher and Mackenzie 1923), quantitative genetics (Falconer and Mackay 1996; Sesardic 2005), and genetic epidemiology (Ottman 1996), as well as in the developmental tradition (Griffiths and Tabery 2008; Lewontin 1974[2006]), and following those authors I call it "*statistical interaction*".

To slightly change the example, Table 2 gives the height values of six G-E combinations, and Figure 3 shows the norm of reaction graph accordingly. In this example, given the phenotypic value of (G1, E1), when E1 changes to E2, the phenotypic value is increased by 5 points. When

[2] To be noted, the standard example represents linear relations between each variable with phenotypic values, but linearity between one variable and its phenotypic effect is not a requirement for the additivity of the effects of two variables. If the curves (or the lines) of the reaction norms are parallel (they are always the same distance apart), then there is vernacular interaction.
[3] A third variable should be added when there is a gene-environment correlation. Because it is a different kind of complication for biometric analysis of variance, I assume that there is no gene-environment correlation in the examples. For details see (Sesardic 2005, chap.3).

G1 changes to G2, it leads the increase of the phenotype by 22 points. The sum of these two effects is 27 points, but the combination of these two changes, that is from (G1, E1) to (G2, E2), is rather 15. This means that the effects of two variables are non-additive. Accordingly, the two lines of Figure 3 are not parallel.

Table 2.Statistical Interaction

	E1	E2	E3	Em
G1	5	10	15	10
G2	27	20	13	20
Gm	16	15	14	**M =15**

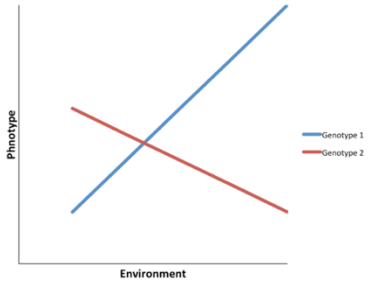


Figure 3. Statistical Interaction

To apply ANOVA again. The phenotypic variance V_P is 49.7, V_G is 0.7 and V_E is 25. Here, the sum of V_G and V_E is smaller than V_P. This means a breaking down of the additivity of V_G and V_E in Equation (1) and requires a modification as follows:

$$V_P = V_G + V_E + V_{G*E} \qquad (4)$$

Here, V_{G*E} can be easily calculated as 24. But it is not easy to interpret it. A literal reading of Equation (4) would be that the phenotypic variation is caused by the gene, the gene-environment interaction and the environment, separately. The causal graph is like Figure 4a, where each of three factors independently contributes to the total phenotypic variation. However, as Wahlsten claims (Wahlsten 2000), the factors are conceptually inseparable. A non-nil V_{G*E} means that the genetic cause and environmental cause affect the phenotypic variation interdependently. The causal graph should be as Figure 4b. That said, the phenotypic variation in the population is explained by interdependent causes of the gene and the environment.

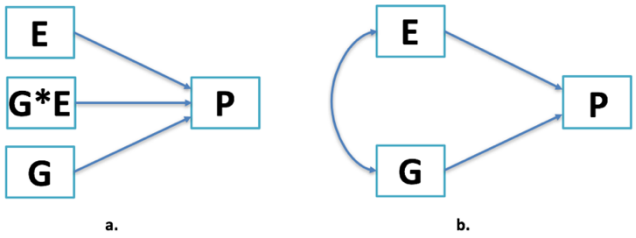


Figure 4. Causal graphs

In other words, the number of statistical interaction V_{G*E} does not correspond to the degree of interdependence between genotypic and environmental variables, but a sign of the two factors being interdependent in causing phenotypic difference. Therefore, theoretically speaking, if V_{G*E} is significant (not caused by a systematic error), then the gene-environment partition collapses at

first place, and the interpretation of a separation of two causes for V_G and V_E would be incorrect.[4] For the question of "how much of the height variation is caused by genetic influence in the targeted population", the heritability analysis can no longer provide a valid answer.

4. The dispute on statistical interaction

Given that the notion of statistical interaction posits a fundamental problem for heritability analysis, the next step is to examine the specific dispute on it. In this section, I revisit Tabery's case study between Hogben and Fisher (Tabery 2014, chap.2). Different from Tabery's diagnosis that the dispute is caused by adopting different concepts of interaction, I show that their dispute is exactly on statistical interaction, and ultimately on an empirical question of whether the causal interdependence between genes and the environment reflected by statistical interaction is common in nature.

As mentioned before, Fisher's study on crop variation sets the foundation for using ANOVA in analyzing the causes of variation. In 1932 and 1933, Lancelot Hogben points out a problem in interpreting the results of ANOVA. Hogben's concern builds on Krafka's experiment of *Drosophila*, in which two strains (the low-bar and the ultra-bar) incubated at different temperatures (15° and 25° C) and the number of their eye facets are observed. The resulting data is showed in Table 3 (cited in Tabery 2014, p.29).

Table 3. Krafka's experiment

	E1(15° C)	E2 (25° C)	Em
G1 (Low Bar)	189	74	131.5
G2 (Ultra Bar)	52	25	38.5
Gm	120.5	49.5	**M = 85**

Given the value of (G1, E1), when the temperature increases from E1 to E2, the phenotypic value decreases 115 points. And when the genotype changes from G1 to G2, the phenotypic value decreases 137 points. The sum of two changes is 252 points, which is *not* equal to the number of (G1, E1) minuses (G2, E2). This means that the effects of the genotype and the environment are non-additive. If a graph of reaction norm being drawn, the lines will be unparalleled. Applying ANOVA to Table 3 results with a significant V_{G*E}. This means that there is statistical interaction. If we estimate the heritability anyway, H^2 is 0.32, meaning that 32 percent of the phenotypic variance is due to genetic influence.

Hogben focuses on the interpretation of the statement x percent of the variance is due to heredity. He claims that it does not follow that the variance would be reduced by x percent if there were no genetic difference. There are two ways to eliminate the genetic difference, either by killing off all

[4] Practically speaking, if V_{G*E} is small compared to V_G and V_E, one can loosely interpret V_G and V_E as the genotypic and environmental effects on phenotypic variation, and the heritability as the contribution made by the genes.

ultra-bar flies, or by killing all low-bar flies. But these two ways result in two reductions of the phenotypic variance, 15% in the first case and 95% in the latter case, neither is the same as the heritability estimated. This is called by Hogben a "lack of singularity in the problem". Hogben donates a third class of variability "which arises from the combination of a particular hereditary constitution with a particular kind of environment" (Hogben 1932, p.98). This also motivates him to emphasize the specification of the environment for individual developments before ascribing causes to phenotypic variations.

Hogben's problem would not arise for vernacular interaction. Recall the standard example in Section 2 (showed in Table 1). There are also two ways to abolish the genetic difference, either by killing all of G2, or by killing all of G1, the reductions of phenotypic variance are both 16.7, which equals to the value of V_G. Therefore, the lack of singularity problem is due to statistical interaction. Fisher's reply confirms this idea, as he writes, "Dear Hogben, I think I see your point now. You are on the question of non-linear interaction of environment and heredity. The analysis of variance and covariance is only a quadratic analysis and as such only considers *additive* effects" (Quoted from Tabery on Fisher, 2014, p.33, emphasis added). Here, Fisher admits that if the effects are non-additive, applying of ANOVA would be groundless.

Tabery seems to recognize the real dispute between Fisher and Hogben when he remarks, "[w]hen it came to judging the empirical evidence of interaction in nature, Fisher and Hogben evaluated the situation quite differently. Where Fisher saw a phenomenon of 'possible, but unproved importance,' Hogben saw a phenomenon he took to be common in nature" (ibid. p.34). But then Tabery draws our attention back to the explanatory divide. He claims that Fisher is working in the variation-partitioning approach and holds the interaction of nature and nurture as "an absence of additivity between the main effects", while Hogben in mechanism-elucidation approach and defines interaction as "a presence of the third class of variability" (p. 38-39). Tabery thus calls the former "biometric concept of interaction" and the latter "developmental concept of interaction" (p. 38-39).

This diagnosis, at least in the context of the nature-nurture debate, is quite misleading. Instead of adopting two concepts of interaction, Fisher and Hogben would both agree that the lack of singularity problem arises because of *non-additive* effects of genetic and environmental effects, that is, because of statistical interaction. Their dispute lies on their attitudes towards an empirical question of whether statistical interaction is common in nature[5]. If statistical interaction arises often in nature as hold by Hogben, then it would be wrong to use ANOVA assuming *a priori* an additive relation of the gene and the environment; if it is not, as suggested by Fisher, then it is not a problem for heritability analysis.

To be clear, I do not deny that biometricians and interactionists take roughly different approaches of asking and answering questions. However, I object the idea that the disputes surrounding the nature-nurture debate can be resolved by distinguishing and clarifying two approaches with two

[5] Technically speaking, at least two factors might influence the detection of statistical interaction even in ideal experimental designs with two or more values of both genotypic and environmental variables. They are the scale of measurement (Khoury et al. 1986) and sample size (Wahlsten 2000). The details are beyond the scope of this paper.

concepts of interaction. The position taken by Tabery, does largely resolve the disputes on trivial and vernacular interaction, but not that on statistical interaction. A more sophisticated version of nature-nurture debate remains: if there is a statistical interaction in applying ANOVA, then the heritability estimated cannot be interpreted to answer the more sophisticated question of how much the phenotypic variation in the population is caused by genotypic variation.

5. Conclusion

To conclude, apart from the confusion arising from the explanatory divide proposed by Tabery, there is a fundamental problem regarding to the notion of statistical interaction in heritability analysis. With regards to the more sophisticated version of nature-nurture debate, biometricians might have a way out: if statistical interaction rarely happens in nature, the answer of genetic causal responsibility provided by heritability analysis is still valid.

As to the question of whether the statistical interaction is common in nature, it is unresolved. Tabery's survey on the studies of human depression shows that three meta-analyses have different outcomes about the detection of statistical interaction, of which two of them are negative and one is positive (Tabery 2015). Epidemiologists are also cautious about statistical interaction that they usually accept a particular statistical interaction only if it can be reproduced in two or more studies (Hunter 2005). There are also disputes in human behavioral genetic analyses over whether statistical interactions can be simply ignored or the additivity assumption can be made *a priori* (Moffitt et al. 2006). This means that the nature-nurture debate remains, of which the notion of statistical interaction plays the central role.

Although the notion of statistical interaction hinders biometricians from properly answering the more sophisticated question, recent researches in behavioral genetics and epidemiology have been considering it as an important notion that is worth further investigation (Baye et al. 2011; Darling et al. 2016). This transition is an advance beyond the traditional biometric approach, and is thought to be a step to understand more complex causations of phenotypic variation in populations, especially for human complex diseases and mental disorders. That said, researchers are more interested in the interplay of genes and environmental factors influencing the risk of diseases, which, with the knowledge from molecular biology and developmental biology, would provide information for elucidating the mechanisms of phenotypic production in individual development (McGue and Carey 2017). Only in this sense, the versus between nature and nurture is said to be overcome.

References:

Baye, Tesfaye M, Tilahun Abebe, and Russell A. Wilke. 2011. "Genotype–environment Interactions and Their Translational Implications." *Personalized Medicine* 8 (1):59–70.

Darling, Katherine Weatherford, Sara L. Ackerman, Robert H. Hiatt, Sandra Soo-Jin Lee, and Janet K. Shim. 2016. "Enacting the Molecular Imperative: How Gene-Environment Interaction Research Links Bodies and Environments in the Post-Genomic Age." *Social Science & Medicine* 155:51–60.

Downes, Stephen M. 2015. "Heritability." http://stanford.library.usyd.edu.au/entries/heredity/.

Falconer, Douglas S., and Trudy F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th edition. Harlow U.A.: Pearson.

Fisher, Ronald A., and Winifred A. Mackenzie. 1923. "Studies in Crop Variation. II. The Manurial Response of Different Potato Varieties." *The Journal of Agricultural Science* 13 (3):311–320.

Freese, Jeremy. 2006. "Commentary: The Analysis of Variance and the Social Complexities of Genetic Causation." *International Journal of Epidemiology* 35 (3):534–536.

Khoury, Muin J., Terri H. Beaty, Carol A. Newill, Steve Bryant, and Bernice H. Cohen. 1986. "Genetic-Environmental Interactions in Chronic Airways Obstruction." *International Journal of Epidemiology* 15 (1):65–72.

Griffiths, Paul E., and James Tabery. 2008. "Behavioral Genetics and Development: Historical and Conceptual Causes of Controversy." *New Ideas in Psychology* 26 (3):332–352.

Hogben, Lancelot. 1932. "Genetic Principles in Medicine and Social Science." http://philpapers.org/rec/HOGGPI.

Hunter, David J. 2005. "Gene–environment Interactions in Human Diseases." *Nature Reviews Genetics* 6 (4):287–298.

Jensen, Arthur. 1969. "How Much Can We Boost IQ and Scholastic Achievement." *Harvard Educational Review* 39 (1):1–123.

Lewontin, Richard C. 1974 [2006]. "The Analysis of Variance and the Analysis of Causes." *International Journal of Epidemiology* 35 (3):520–525.

Lewontin, Richard. C. 2006. "Commentary: Statistical Analysis or Biological Analysis as Tools for Understanding Biological Causes." *International Journal of Epidemiology* 35 (3):536–537.

Machamer, Peter, Lindley Darden, and Carl F. Craver. 2000. "Thinking about Mechanisms." *Philosophy of Science*, 1–25.

McGue, Matt, and Bridget E. Carey. 2017. "Gene–environment Interaction in the Behavioral Sciences: Findings, Challenges, and Prospects." In *Gene-Environment Transactions in Developmental Psychopathology*, 35–57. Springer.

Moffitt, Terrie E., Avshalom Caspi, and Michael Rutter. 2005. "Strategy for Investigating Interactions between Measured Genes and Measured Environments." *Archives of General Psychiatry* 62 (5):473–81.

Ottman, Ruth. 1996. "Gene–environment Interaction: Definitions and Study Designs." *Preventive Medicine* 25 (6):764.

Perbal, Laurence. 2013. "G*E Interaction and Pluralism in the Postgenomic Era." *Biological Theory* 7 (3):266–274.

Rose, Steven P. R. 2006. "Commentary: Heritability Estimates—long Past Their Sell-by Date." *International Journal of Epidemiology* 35 (3):525–27.

Sesardic, Neven. 2005. *Making Sense of Heritability*. Cambridge University Press.

Sterelny, Kim, and Paul E. Griffiths. 1999. *Sex and Death: An Introduction to Philosophy of Biology*. Chicago and London: University of Chicago Press.

Tabery, James. 2014. *Beyond Versus: The Struggle to Define the Interaction of Nature and Nurture*. Cambridge, Massachusetts, and London, England: MIT Press.

Tabery, James. 2015. "Debating Interaction: The History, and an Explanation." *International Journal of Epidemiology*, dyv053.

Taylor, Peter. 2006. "Commentary: The Analysis of Variance Is an Analysis of Causes (of a Very Circumscribed Kind)." *International Journal of Epidemiology* 35 (3):527–531.

Vreeke, Gert-Jan. 2006. "Commentary: The Attainability of Causal Knowledge of Genetic Effects in Complex Human Traits." *International Journal of Epidemiology* 35 (3):531–534.

Wahlsten, Douglas. 2000. "Analysis of Variance in the Service of Interactionism." *Human Development* 43 (1):46–50.

Waters, C. Kenneth. 2007. "Causes That Make a Difference." *The Journal of Philosophy* 104 (11):551–579.

Woodward, James. 2003. *Making Things Happen: A Theory of Causal Explanation*. Oxford University Press.

List of Speakers (Order by Last Name)

Chan, Man-Ho (The Education University of Hong Kong, Hong Kong);
Chen, Ruey-Lin (National Chung-Cheng University, Taiwan);
Chiu, Lynn (University of Bordeaux/CNRS, France);
Di Iorio, Francesco (Nankai University, China);
Dumouchel, Paul (Ritsumeikan University, Japan);
Gotoh, Reiko (Hitotsubashi University, Japan);
Hajek, Alan (Australian National University, Australia);
Hassan, Ibrahim N. (The National University of Malaysia, Malaysia);
Hou, Richard W. T. (National Chung Cheng University, Taiwan.);
Hricko, Jonathan (National Yang-Ming University, Taiwan);
Hsiao, Ming-Yuan (Soochow University, Taiwan);
Huang, Tsung-Ren (National Taiwan University, Taiwan);
Hung, Tzu-Wei (Academia Sinica, Taiwan);
Iseda, Tetsuji (Kyoto University, Japan);
Jhang, Jhih-Hao (National Chung-Cheng University, Taiwan);
Jung, Jaemin (Wonkwang University, Korea);
Ko, Insok (Inha University, Korea);
Lin, Hanti (UC Davis, USA.);
Lu, Qiao-Ying (Sun Yat-sen University, China);
Morita, Kunihisa (Kyushu University, Japan);
Othman, Mohd. Yusof Hj. (Universiti Kebangsaan Malaysia, Malaysia);
Otsuka, Jun (Kobe University, Japan);
Park, Ilho (Chonbuk National University, Korea);
Radiman, Shahidan (Universiti Kebangsaan Malaysia, Malaysia);
Rhee, Young E. (Kangwon National University, Korea);
Samian, Abdul L. (Universiti Kebangsaan Malaysia, Malaysia);
Shan, Ya-Feng (Durham University, UK.)
Shim, Jiwon (Inje Unversity, Korea);
Tsai, Meng-Li (National I-Lan University, Taiwan);
Wang, Linton (National Chung-Cheng University, Taiwan);
Wang, Wei (Tsing-Hua University, China);
Xu, Zhu (East China Normal University, China);
Yan, Karen (National Yang-Ming University, Taiwan);
Yang, Shi-Jian (Xiamen University, China);
Yeh, Hsiao-Fan (National Formosa University, Taiwan);
Yen, Chun-Ping (National Tsing-Hua University, Taiwan) ;
Yoshida, Kei (Waseda University, Japan);
Yu, Feng (East China Normal University, China).

List of Conference Assistants (Order by Last Name)

Chang, Ching-Ya;
Chen, Ciou-Mai;
Chen, Listen;
Choong, Aaron;
Chang, Jung-Hong;
Jiang, Huei-Min;
Kwok, Henry;
Lee, Bodinho;
Li, Lin;
Lin, Shi;
Liu, Yu-Chun;
Tsang, Eric;
Wu, Huang-yi.

All people are from the Department of Philosophy, National Chung-Cheng University, Taiwan.