Undergraduate Philosophy Journal of Australasia

Volume 3

December 2021

UPJA

Founded in 2019, the *Undergraduate Philosophy Journal of Australasia* (UPJA) is the first undergraduate philosophy journal run by students from the Australasian region. We publish two issues and host two virtual conferences annually: one mid-year and another at the end of the year. Our calls for papers for each issue open roughly in March and September, respectively. We aim to be an inclusive and diverse journal that welcomes submissions on any philosophical topic attempting to make a substantive contribution to contemporary debate. Submissions from women and other members of underrepresented groups in philosophy, including those for whom English is not their first language, are particularly encouraged.

Undergraduate Philosophy Journal of Australasia – UPJA

Undergraduate Philosophy Journal of Australasia

Mailing list: https://forms.gle/we6ZeihAx1hDKf6E6

contact@upja.online

www.upja.online

ISSN: 2653-3146

2020 - 2021 Editorial Team

Editors-in-Chief

Allan Bechaz (University of Melbourne)

Racher Du (Nanyang Technological University)

Associate Editors

Will Cailes (Monash University)

Thomas Spiteri (University of Melbourne)

2021- 2022 Editorial Team

Editors-in-Chief

Will Cailes (Monash University)

Thomas Spiteri (University of Melbourne)

Associate Editors

Jack Hawke 何健平 (Australian National University)

Jessica Sophia Ralph (Massey University)

Faculty Advisors

Associate Professor Stephanie Collins (Australian Catholic University)

Assistant Professor Sandra Leonie Field (Yale-NUS College)

Dr Carolyn Mason (University of Canterbury)

Referees

Ranier Abengana is a lecturer in the University of the Philippines Baguio and a postgraduate student in University College Dublin (Ireland). He is the co-founder of The Critical University Studies Project and an active member of Minorities and Philosophy (Dublin Chapter). His research interests include German Idealism, critical theory, critical university studies, contemporary political theory, sociology of time, and systems theory.

Dr Snita Ahir-Knight (BA (Hons) Social Work, MA Philosophy, PhD Philosophy) is a visiting scholar in philosophy at Te Herenga Waka—Victoria University of Wellington, Aotearoa New Zealand. She also has a current and established career as a social worker and child and adolescent therapist.

David M. Batt is a recent graduate in History and Philosophy of Science at the University of Melbourne where he also studied mathematics and physics at a graduate level. He is interested in studying the nature of science from the perspectives of nineteenth- and twentieth-century continental philosophers like Nietzsche, Heidegger, and Foucault, specifically the role that history plays in understanding something in the present. His current research centres around the history of the knowledge practices associated with the making, managing, and modifying of sovereign monetary systems in the early-modern period.

Abigail Bergeron is an undergraduate student currently pursuing both a degree in philosophy at Trent University and a degree in law at Swansea University. She is particularly interested in the philosophy of animal minds, feminist epistemologies, subjectivity, and neurodiversity.

Vlademire Kevin D. Bumatay is an Assistant Professor of Philosophy from the University of the Philippines-Baguio, Republic of the Philippines. His main areas of research are General Philosophy of Science, Metaphysics, and Epistemology.

Monte Cairns completed his BA (Hons) in History and Philosophy of Science at the University of Melbourne in 2020. He is now enrolled in the MPhil program in History and Philosophy of Science and Medicine at the University of Cambridge, graduating 2022.

David Chandler is an undergraduate philosophy student at King's College London. His current interests include metaphilosophy, logic, and the philosophy of language.

Benny Chee is a philosophy student at Nanyang Technological University, Singapore. Benny's research interests include epistemology and normative ethics.

Chua Yi, Jonathan is a philosophy Master's student at the Nanyang Technological University, Singapore. His research interests include Confucian political philosophy, hermeneutics and environmental ethics.

Anna Day is an Honours candidate at the University of Sydney, recently graduating with an undergraduate degree in philosophy and politics. She is particularly

interested in the moral imagination and the psychological implications of philosophical children's literature.

Sam Dickson is pursuing his PhD in Philosophy at the University of York. His research interests lie mainly in metaphysics and the philosophy of mathematics, with recent work focused on the nature of mathematical objects and their potential for causal powers, as well as discussions of whether or not counterpossibles are trivially true and the extent to which they can be used in explanation.

Harry Drummond is currently a PhD researcher at the University of Liverpool, United Kingdom, having obtained his BA and MA from Durham University. His research investigates the role of social cognition in aesthetic experience.

James Eder is a COVID-19 contact tracer at New Zealand's Ministry of Health and is completing his Master's thesis in philosophy at the University of Canterbury. His thesis analyses how voluntary uses of technology to enhance human cognitive/behavioural capacities can undermine or amplify one's personal autonomy.

Alejandro Gracia Di Rienzo is a PhD student at the University of Santiago de Compostela, where he is writing a thesis on mathematical structuralism and the metaphysics of quantity. His main interests are logic, metaphysics and the philosophy of mathematics.

Abi Gunning graduated with Honours in French and philosophy from the University of Edinburgh. Her interests include feminist philosophy, social and political philosophy, epistemology, and philosophy of language.

Isaac Hadfield is a student on the BPhil in philosophy at the University of Oxford, having recently completed his BA in Philosophy, Politics, and Economics also at Oxford. He is particularly interested in social philosophy, ethics, epistemology, and philosophical methodology.

Levi Haeck has recently commenced a PhD at Ghent University, and specialises in Kant's transcendental logic and how this discloses a subtle yet fundamental relation between logic and metaphysics.

Frank Hernandez is a first-year MA philosophy student at the Central European University. He graduated from the University of Texas at El Paso with a double major in philosophy and multimedia journalism. His main interests are social epistemology, philosophy of language and philosophy of science.

Brylea Hollinshead is pursuing her Honours degree in philosophy at the University of Canterbury. Her research interests include topics in ethics, philosophy of emotion, and metaphilosophy.

Bhanuraj Kashyap is a philosophy master's student at Deakin University. They are interested in moral philosophy, political philosophy, and the philosophy of race. In their spare time, they enjoy reading, cooking, and running.

Johnny Kennedy is currently studying a Bachelor of Arts and Laws at the University of Sydney. Next semester, Johnny will be commencing his Honours in philosophy, writing a paper synthesising Heidegger and Kuhn.

Beau Kent recently graduated from an Arts undergraduate from the University of Melbourne majoring in philosophy. His philosophical interests include postmodern literary theory and hermeneutics as well as the philosophy of race, speech act theory, and phenomenology. He will be attempting to write an honours thesis on the phenomenology of transracial adoption over the next year and a half.

Vincent Lee graduated from Yale-NUS College in Singapore and has recently begun his PhD in Philosophy at the University of Toronto. His research interests lie in the history of early modern European and Indian Buddhist philosophies, especially where they pertain to substancehood and selfhood.

Italo Lins Lemos is a temporary Professor at the State University of Maringá, Brazil. He's recently finished his Ph.D at the Federal University of Santa Catarina, Brazil. He was a visiting scholar at Dartmouth College, USA, from 2018-19. His philosophical research focuses on Metaphysics, Philosophy of Language and Epistemology, especially when it comes to issues related to the nature of fiction.

Baruday Logan is a philosophy Honours student at Deakin University, Australia. He is especially interested in teaching philosophy to young people, and his philosophical research focuses on Psychoanalytics, Mimetic Theory and Critical Studies of Men and Masculinities.

Sam McGuiggan is currently pursuing a Bachelor of Arts at the University of Queensland, majoring in Philosophy and Creative Writing. He hopes to pursue a postgraduate in Philosophy and has a particular interest in Deleuzian post-structural thought.

Kai Milanovich is completing a dual-degree program in music and philosophy at the University of Missouri-Kansas City. They are primarily interested in bioethics, social epistemology, and queer/feminist philosophy. They are planning to continue their philosophical studies in graduate school in 2022.

Ariane Moore is a PhD candidate in philosophy at the University of Tasmania. Her primary interests include philosophy of religion, aesthetics and atheism studies.

Angelo Antonio Pires de Oliveira is a PhD candidate in Ancient Philosophy at the University of Campinas. He was a Masters (Oct. 2016 - Feb. 2017) and PhD (Sep. 2019 - Sep. 2020) visiting student at the Munich School of Ancient Philosophy (MUSAΦ), Ludwig-Maximilians-Universität (LMU) München, Germany. Currently, he is a member of the research group Metaphysics, Science and Dialectic in Aristotle (MEZA), University of Campinas, Brazil. His research is focused on issues related to Moral Motivation in the Nicomachean Ethics.

Samuel Parsons recently completed an Honours in Philosophy from the University of Adelaide. His primary research interests are in political philosophy and metaphysics.

Lauren Pedersen-Hor recently completed her undergraduate studies at Deakin University, Melbourne, majoring in philosophy. Her research interests include social philosophy and the philosophy of film, with a particular focus on theories of power and social control.

Henry Phillips is a Masters student at the University of Adelaide. He is interested in moral psychology, normative theory, metaethics, philosophy of action and the way that all these fields intersect. When he is not doing philosophy, he enjoys playing the piano, running, and learning languages.

Katie Prosser is a second-year student pursuing a BPhil in Philosophy at the University of Oxford. Her main research interests are animal ethics and climate ethics, primarily focusing on the intersection of these topics, and approaching them from a notably feminist viewpoint.

Lewis Rosenberg is a Masters student at Western Sydney University. He is an alumnus of Flinders University and the Australian National University. His research interests are hermeneutics, phenomenology, Ancient Greek philosophy, and environmental philosophy.

Stephen Rowe is a Master's student in psychology at the University of Canterbury. He has a keen interest in how the descriptive can inform the normative, particularly in relation to debates about mental illness and how we can live meaningful lives.

Mitchell Stirzaker recently graduated from the University of Sydney with an honours degree in Philosophy. His interests include the traditions of Ancient Greek Philosophy, Buddhist Philosophy, and Existentialism. His current focus is the examination of how Philosophy as a Way of Life traditions combine theory and praxis.

Zachary Swindlehurst is a PhD candidate in philosophy at the Australian National University. He is interested in philosophy of mind, epistemology, and metaethics.

Kramer Thompson is currently a philosophy PhD student at the Australian National University. He has a broad interest in philosophy, but focuses on theory of action, ethics, metaphysics, and epistemology.

Olivia Turner is a fourth year student at the ANU studying a double degree of Politics, Philosophy and Economics, and International Security Studies. Olivia has always loved philosophy and is particularly interested in political and social philosophy and its intersection with policy.

Melchior Vacheron-Raywood is currently reading Mathematics and Philosophy at the University of Durham, UK. He specialises in logic and in the history of mathematics, also having interests in the philosophy of science.

Danny Wardle is a PhD candidate at the Australian Catholic University's Dianoia Institute of Philosophy and a longtime member of the Australasian Association of Philosophy's Postgraduate Committee. He is particularly interested in topics at the intersection of metaphysics, philosophy of language, and philosophy of physics.

Xisheng Wang is a graduate student at Sciences Po and the London School of Economics. He earned his B.A.s at Michigan State University majoring in philosophy and political science-prelaw. He is interested in epistemology, political theory, and Kant's transcendental idealism.

Wen Xianda recently graduated with an Honours (Distinction) in Philosophy at the National University of Singapore. His philosophical research focused on the intersection between Philosophy of Art and Media Studies, exploring the philosophical potential of internet memes. He will also be pursuing a Juris Doctor at Singapore Management University this year.

Patrick Williamson is an MPhil candidate at the Australian National University, working on moral decision making.

Henry Youè is an Honours student at the University of Melbourne. His research interests include philosophy of language and pragmatics.

Editors' Note

It's been a very productive year for the 2021 UPJA editorial team. The first half of the year was headed by Co-Editor-in-Chiefs Alan Bechaz and Racher Du and was a period in which several changes were initiated. Namely, the decision to move to one annual volume in 2021. This allowed for more papers to accrue over the year which enhanced the quality of submissions and likelihood of further engagement from those typically underrepresented in philosophy. Following a suggestion by then associate editor Will Cailes, the team worked toward creating a new segment of the journal, *Conversations from the Region,* which has since become a prominent feature of the journal's identity. This is a series of interviews with academic philosophers with strong ties to Australasia and provides a forum for experienced philosophers to share their ideas and advice with students. To date, we have interviewed Professor Graham Oppy, Assistant Professor Sandra Leonie Field, Associate Professor Hiu Chuk Winnie Sung, Professor Frank Jackson and Professor Bill Fish with some highlights being included below. Additionally, we will soon be publishing interviews with Peter Godfrey-Smith, Phillip Pettit and Kate Manne. We would like to thank these philosophers for their time and thoughts reflecting on their work and undergraduate experience.

In the second half of 2021, Will Cailes and Thomas Spiteri were appointed Co-Editors-in-Chief and were joined by current associate editors Jack Hawke 何健平 and Jessica Sophia Ralph. Together we have continued to grow the journal's scope and ambitions outlined by its founders. The team's work has culminated in expanding the biannual Australasian Undergraduate Philosophy conference to include two keynote speakers; deepening our affiliation with the Australasian Association of Philosophy (AAP); gaining an ISSN number in hope of greater dissemination; and the publication of the journal's third volume.

This year, we have selected five papers for publication. We have Monte Cairns of the University of Melbourne who wrote his paper on *Contingent and Inevitable Elements of Quantum Mechanics*; Christian Carbonell of The University of Barcelona who wrote *Actions of Trust and their Cognitive Motivation*; Tiago Carneiro da Silva of the Federal University of Rio de Janeiro who wrote *The Method of Wide Reflective Equilibrium and Self-evidence*; Yunlong Cao from John Hopkins University, whose paper is entitled *Beyond Soul and Virtue*; and Martin Walter Niederl of The University of Vienna whose paper is on *Practical Animal Reasoning*. We are proud to announce and congratulate the winners of our two prizes are Monte Cairns (Best Paper) and Christian Carbonell (Best Paper from a member of an Underrepresented Group in Philosophy). The awards are generously funded by the AAP, and we are grateful for their ongoing contributions to undergraduate philosophy.

Acknowledgments

We had two conferences this year, which were both great successes. The first conference was held in June where Dr Carolyn Mason was the keynote speaker while our recent conference in November had two keynote speakers: Associate Professor M R. X. Dentith and Dr Claire Benn. We would like to thank each keynote speaker for enriching our conference with their wonderful presentations. Over these two conferences, we were also pleased to have Christian, Martin, Tiago and Yunlong from this issue present their work, as well as many others.

Additionally, there are many people and organisations working in the background to support and assist UPJA's goal of facilitating excellent undergraduate philosophy scholarship. We are grateful for the support of the numerous student philosophy societies and Minorities and Philosophy chapters for circulating our call for papers; Justin Weinberg of Daily Nous and the AAP for sharing our content; our three faculty advisors — Associate Professor Stephanie Collins, Assistant Professor Sandra Leonie Field, and Dr Carolyn Mason — for their ongoing assistance and advice; and most importantly, our team of referees for providing such detailed and constructive reports. The present volume would not have been completed without their time and effort.

Finally, we would like to thank those who have served alongside us as editors. We thank Alan Bechaz and Racher Du for their philosophical knowledge and commitment to scholarship, as well as their near limitless patience training us during the first half of the year. Equally we would like to thank Jack Hawke 何健平 and Jessica Sophia Ralph for their hard work and philosophical insight during the second half of 2021, and for consistently being motivated to improve and expand the journal's mission. We are excited to see how they develop UPJA in 2022.

Will Cailes and Thomas Spiteri

Co-Editors-in-Chief

December 2021

Conversations From the Region

"Do you have any advice for students?"

Professor Graham Oppy (Monash University)

I guess it depends what we're thinking about. If you're doing a class, the important thing is not to give up. No matter how hard it seems to you, that's going to be a measure of how hard it seems to pretty much everyone, I would think. If we're thinking more broadly about the possibility of a career in philosophy, it's probably worth knowing this: the job market in philosophy in Australia has been terrible since about 1975. So, if you're thinking about a career in philosophy, you may well end up somewhere other than Australia, and you should know that before you go into it. Because, if you kind of get to a point where you have to give up because you're not prepared to leave Australia, and you're not finding one of the very rare positions that comes up in Australia, it's better to avoid that disappointment from the beginning, I think. The reason why 1975— well, there was a big explosion of places in Australian universities from the beginning of the 60s and all of the new universities were around and had filled up their staff profiles by about the mid 70s. Since then, there have hardly been any new departments. There have been a couple, ACU and Notre Dame, Bond maybe, but there have been hardly any. Also, there's been some shrinkage in the size of departments. So, there's just very few positions that become available. That doesn't mean that you should think that you can't do it. After all, I'm proof that you can, and there's a bunch of other people just like me all around the place. And I will be shuffling off fairly soon, I don't have that many years left. So, it's not that you can't do it, it's just that you can't take anything for granted, right? It may well be that in order to find the job that you want, if you go on, you might have to be prepared to, as some of my students have done, go to Taiwan or Hong Kong. Or some of them who are better than me, the US and the UK.

Assistant Professor Sandra Leonie Field (Yale-NUS)

Writing is always a struggle. But at least for me, after one or two times having that really hard paper where it just doesn't make sense, and you're not sure whether it's going to come together, and then it comes together; then the next time when you have that terrible feeling, you're like, I've had this feeling before, I just have to keep going! I just have to keep going, and I will get there. I don't know that it goes away, at least for me, it doesn't. You always have that feeling you don't know what you're doing. Just keep going. That's my advice for students.

Associate Professor Hiu Chuk Winnie Sung (Nanyang Technological University)

Academically, I think I struggled the most with writing. English is not my first language. And philosophical writing is so different from writing essays for other courses. In the beginning when I received low grades, I did not know whether the problem was with my English, or with my philosophical writing, or both. So whenever there was an essay due, I was too busy worrying about my English instead of thinking carefully about the philosophical question. I wish I could re-write all these undergraduate essays now. My own voice wasn't really there in these essays. I was too focused on the language bit. If I could go back in time, I would be bolder and just express my view, and then come back and tidy up the language. I would advise the same for students in English-language institutions whose native language is not English. When working on an essay, it might help to first focus on thinking about and addressing the philosophical problems instead of worrying about the quality of my writing.

Emeritus Professor Frank Jackson (Australian National University)

Advice to students? Philosophy is a wonderful subject. Only a small number of you will become philosophers in the sense of joining a philosophy department, but you will be philosophers in a wider sense for the rest of your lives. For the rest of your lives, you will know what it is to develop an argument for a position, to make the key distinctions needed to clarify an issue, how to spot dodgy arguments and rhetorical tricks, and how to think about hard, abstract questions like the relation between mental states and brain states and whether we should be retributivists or consequentialists about punishment.

Professor Bill Fish (Massey University)

When it comes to advice for students, one big thing is to not worry if you don't feel like you completely get something the first time around. I'm still finding new ideas in things I have read many times before, so don't feel like you're not succeeding if you feel there are things you don't fully understand. Probably the most important thing, though, is just to enjoy it – when you study philosophy, you get to think about a range of really interesting things. Find the fun in that, and try not to let it feel like a chore (even though I know that's tough when essay deadlines are looming). And talk to people – not only your classmates, but also your lecturers and philosophers at other universities. As an undergraduate, I remember going to seminars where I could engage with the people who taught me on the same level, rather than as a student to their teacher, and realising that they're real people who are not that dissimilar to me made me feel that I could possibly be a philosopher too.

Table of Contents

* Winner of Best Paper

† Winner of Best Paper (Member of an Underrepresented Group in Philosophy)

Undergraduate Philosophy Journal of Australasia www.upja.online
Vol. 3, December 2021

Contingent and Inevitable Elements of Quantum Mechanics: The Lacking Tension of the Contingency/Inevitability Debate

Monte Cairns[1]

University of Melbourne

Abstract

This paper addresses the contingency/inevitability debate, which asks essentially whether scientific theories are epistemically guaranteed given successful inquiry, or partially determined by unguaranteed social and psychological factors. In other words, participants in the debate ask whether changes to historical conditions might feasibly have resulted in the development of scientific theories alternative to our own. Here, I address the issue in regard to the extant alternative non-relativistic quantum theories of the standard and de Broglie-Bohmian quantum models. The existence of these long-term viable, alternative theories regarding the same phenomena has made quantum mechanics a seminal case study for the debate. Taking James Cushing and Lena Soler as representative of contingentists, and Ian Hacking and Steven Weinberg as representative of inevitabilists, I argue that contrary to appearances contingentists and inevitabilists are not in substantive disagreement regarding quantum mechanics. Contingentists hold that quantum ontologies are contingent, whereas inevitabilists hold that empirical results and the nomological structure provided by Schrödinger's equation are inevitable. These views are mutually sustainable. Thus, the philosophical tension of the debate evaporates, leaving us with a surprisingly large degree of contingency that is nonetheless consistent with inevitabilist claims.

[1] *Monte Cairns has recently completed his BA (hons) in History and Philosophy of Science at the University of Melbourne and is presently completing an MPhil in History and Philosophy of Science and Medicine at the University of Cambridge. He is interested in integrated approaches to HPS, with an aim towards resolving philosophical issues through historical and historiographic work.*

ISSN: 2653-3146

1. Introduction

The contingency/inevitability debate is a recalcitrant problem in history and philosophy of science. The contingentist Lena Soler frames the problem by asking: "Could our taken-as-established scientific conclusions, theories, experimental data, ontological commitments, and other scientific 'results' (in a broad sense of result) have been significantly different?"[2] The heterogeneity of the philosophical targets indicated by Soler — and the heterogeneity of the sciences themselves — suggest that her question may have no single answer. This is borne out by the diversity of scope and target represented in the positions that constitute the debate, which cannot be addressed in full here.[3] But in general terms, the debate can be summated in that contingentists take certain aspects of science to partially depend on historical factors that are *unguaranteed*, i.e. extant only by coincidence. This might include social or psychological factors, e.g. the intellectual climate of a given period. If the results (in Soler's broad sense of the term) of scientific research partially rely on such factors, then it stands to reason that such results might feasibly differ had these unguaranteed factors differed. Inevitabilists, meanwhile, typically hold that the results of scientific research are determined by immutable factors. Most typical of inevitability arguments is the claim essentially that 'how the world is' determines the content of scientific claims. Thus, regardless of historically unguaranteed factors, any successful scientific enterprise will return inevitable results. Resolving this tension is held to have important ramifications for how we consider scientific claims — e.g. whether as a unique account of nature, and hence (traditionally formulated) 'knowledge', or as a mutable means of producing accurate predictions, with an unclear relationship to 'how the world is'. Here, in regard specifically to quantum mechanics, I argue that nominal inevitabilism permits a surprisingly high degree of contingency, on the basis that contingency/inevitability claims aim at alternative philosophical targets, rendering their arguments mutually consistent.

In addressing contingency/inevitability, I take up four participating theorists: James Cushing and Lena Soler as representative of contingency, and Ian Hacking and Steven Weinberg as representative of inevitability. Taking these four as representative of the debate as a whole is necessarily an oversimplification, but their views together confer an impression of the breadth, if not the depth, of the issue in general. My argument is essentially that a lack of conceptual clarity within the contingency/inevitability debate has resulted in an *appearance* of philosophical conflict, between 'contingentists' and 'inevitabilists,' when in fact these supposedly

[2] Soler, L. (2015). 'Introduction: The Contingency/Inevitability Debate.' *Science as It Could Have Been: Discussing the Contingency/Inevitability Problem*, p. 1.
[3] For a well-realised summation of these diverse positions, see Kinzel 2015. The issue, essentially, is that speaking to 'science' in general terms might render the question insoluble, due to the enormous disparities between activities typically accounted for as constituting scientific research.

oppositional parties fail to contradict one another. This is the result of a failure to be explicit about the specific elements of 'science' that they target in their arguments.

I develop this thesis in particular reference to Cushing's 1994 study of quantum mechanics. Quantum mechanics provides a particularly fruitful example when discussing contingency/inevitability, in that there are two extant and arguably underdetermined[4] alternative quantum theories — the standard quantum model (henceforth SQM) and the de Broglie-Bohm quantum model (henceforth BQM). Thus, my argument here is that, *in the case of quantum mechanics*, contingentists and inevitabilists are not in substantive or significant disagreement. Through careful comparison of the consequences of the arguments considered here, it becomes apparent that contingentists are chiefly proposing contingency in regard to scientific *ontology*, wherein SQM and BQM display significant and perhaps irreconcilable differences. Inevitabilists emphasise *empirical claims and nomological structure*, which are shared between SQM and BQM. If these views are mutually sustainable — and they appear to be — the philosophical tension of the contingency/inevitability debate in regard to quantum mechanics more or less evaporates. Interestingly, this means that the claim that scientific ontologies — attempts to make descriptive sense of predictive success regarding unobservables — might be historically contingent, and therefore epistemically unguaranteed, appears uncontested within the debate.

2. Quantum Mechanics as Contingent

In this section, I first discuss Cushing's case for contingency, before recounting Soler's approach, which constitutes essentially accounting for Cushing's contingency as a permanent feature of the unfolding of research. Regrettably, I am unable to provide a complete summary of Cushing's historical work regarding quantum mechanics here. A brief sketch must suffice. First, it is important to note that Cushing is primarily interested in the processes underlying the acceptance or rejection of theory within the scientific community. Regarding scientific acceptance, Cushing holds that certain *non-evidential* factors play a role. In instances of actual underdetermination — wherein empirical recourse fails to decide between two theoretical options — Cushing argues that these non-evidential factors, characterised as typically social or psychological, become decisive for the historical course of scientific acceptance and continuing research.[5] Cushing's study of the history of quantum mechanics is thus intended to present a case of actual underdetermination, and then to show which non-evidential factors contributed to the scientific community's preference for one underdetermined option over another.

[4] 'Underdetermined' scientific accounts, as the term is used here, refers to instances wherein empirical evidence fails to decisively favour one competing theoretical alternative over another.
[5] Cushing, J. (1994). *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony* pp. 7-8.

These apparently underdetermined alternatives, SQM and BQM, can be understood in simplistic terms as follows. The two alternatives are means of interpreting empirical data regarding quantum systems. They thus constitute descriptive accounts of these unobservable systems, extrapolated from empirical evidence. SQM, being the traditional approach to quantum mechanics, holds that quantum systems are *discontinuous* (the so-called quantum postulate). Thus, micro- and macro-systems are non-analogous, and microsystems cannot be described in the terminologies of experience. The epistemological consequence of this is that quantum systems cannot be described unless they are observed. Thus, SQM posits 'superpositions,' multiple states in which quantum systems are held to simultaneously exist, prior to observation. Note that accounting for this statistical model as a descriptive account is sometimes rebuffed on the basis that the discontinuity of quantum systems under SQM is in fact a *lack* of description resulting from a reluctance to make non-empirical claims. However, the common sentiment that microsystems cannot be described in continuous terms belies this.

BQM, meanwhile, produces a causal and hence continuous description of the microlevel, therefore able to be expressed more or less in the terminologies of experience. This is at the expense of accepting non-local correlation or entanglement between quantum systems. Although perhaps problematic, it has been shown by Bell's theorem that maintaining both determinacy and locality in quantum mechanics is untenable. Moreover, the no-signalling theorem indicates that non-local entanglement does not constitute non-local causality. Although it has been argued that, due to the empirically unfounded character of BQM's descriptive interpretation, the model constitutes something like a metaphysics rather than a recognisably scientific account, SQM's indeterminacy is no more empirically demonstrable.[6]

So, as per Cushing, in the case of quantum mechanics, we have available two empirically equal yet divergent quantum models in SQM and BQM. SQM enjoys significantly more currency amongst the scientific community both historically and currently. Schrödinger's equation, the formalism which facilitates reliable prediction regarding quantum systems, is shared between the two. But interpretatively and ontologically, they differ significantly. SQM is statistical and indeterminate, whereas BQM is explicitly causal.[7] The indeterminate ontology of SQM seems to make fewer ungrounded suppositions but leaves us "effectively stranded with the formalism and its predictions,"[8] in that it lacks descriptive content. BQM, meanwhile, is able to produce a coherent — yet empirically undemonstrated — description of particle trajectories at the microlevel. If the two are empirically equal, then the widespread

[6] Bricmont, J. (2016). *Making Sense of Quantum Mechanics*. p. 179.
[7] Cushing 1994, *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, pp 53-5.
[8] Ibid., p. 82.

acceptance of SQM rather than BQM — and the arguably subsequent increased elaboration of SQM at the expense of a causal quantum model — is not demanded by logic or evidence. Note that this equivalence is somewhat exaggerated by Cushing, as it seems that BQM exhibits certain *practical* disadvantages.[9] However, in that valuing factors such as fruitfulness remains at least plausibly non-epistemic, and empirical equivalence between SQM and BQM as non-relativistic quantum theories is evident, I proceed in alignment with Cushing.

Cushing subsequently develops an account of the sociological and psychological factors that 'tipped the balance' in favour of SQM over BQM, resulting in SQM's ascendancy. In historical terms, the emergence of SQM as the dominant model was *contingent* in that it was sensitive to the initial sociological conditions wherein the early history of quantum mechanics occurred, due to an absence of decisive (insensitive) epistemic factors. Had these sociological conditions differed, Cushing argues that the historical unfolding of quantum mechanics might subsequently have differed also, in that BQM would be a) significantly more developed than it is, more indeed than SQM; and b) the quantum model most widely accepted by the scientific community.[10]

Significantly, Cushing does *not* appear to support any claim to the effect that the mathematical structure underpinning both SQM and BQM is contingent.[11] His contingency thesis is limited to the theoretical interpretations that were constructed around the formalism of Schrödinger's equation. Speaking generally to the contingency/inevitability problem, then, it appears that Cushing's claim is that in cases of actual underdetermination (i.e., wherever evidence fails decisively to recommend one theory over another), contingent psychological and sociological factors may prove decisive for theory acceptance. Ultimately, however, standards

[9] For instance, BQM suffers from difficulties regarding its extension into the relativistic domain, whereas SQM allows for this with comparative ease. Following Wolfgang Pauli's (1927: 619) observation that a relativistic extension was necessary for SQM in order to account for electron spin, Paul Dirac (1928) was able to produce such in under a year. In contrast, despite an early assurance from David Bohm (1952) that an analogous account was possible under BQM, the causal formulation of Dirac's equation did not begin to emerge until more than three decades later (Bohm et al 1987). Even in contemporary literature, there are questions regarding whether or not BQM has actually been, or even *could* be, made relativistic (see e.g. Durr et al 2014). On the other hand, there are some indications that BQM has the capacity to make predictions in circumstances where SQM cannot (Cushing 1994: 55). There are also certain experimental issues that seem more easily accounted for under the descriptions provided by BQM, especially in the case of quantum cosmology (Falciano et al 2015). But these possible advantages of BQM have not been empirically demonstrated. The upshot is that SQM and BQM must be suspended from broader resultant theory, and treated as isolated non-relativistic quantum theories, in order to be considered equivalent. It is not at all clear that theories ought to be isolated in such a fashion. However, addressing this problem would require engaging in a broader debate that I cannot address here, regarding whether the empirical domain ought to be extended to address factors such as fruitfulness.

[10] Cushing 1994, *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, pp. 185-6.

[11] Ibid., p. 100.

internal to science prove decisive for the development of basic predictive formalisms.

Soler makes no such explicit concession, although she gives no specific attention to formalisms and emphasises the importance of theoretical interpretations and scientific ontologies to her contingency thesis.[12] Soler argues that SQM and BQM are contingent scientific alternatives on the basis that they both a) appear genuinely scientific, insofar as their modelling and investigative practices conform to the norms of the scientific community; b) seek to provide explanations for the same set of phenomena; and c) insofar as they are approached as non-relativistic quantum theories independent of related theory,[13] are empirically equivalent and constitute actively progressive programs.[14] Moreover, Soler takes SQM and BQM having been co-existent, active, and progressive programs at least since 1952 to indicate that scientific alternatives can be independently successful in the *long-term*. Soler maintains that there is no reason to expect that the differences between SQM and BQM will be resolved, framing them as "irreducibly different" and "contradictory."[15] Thus, Soler concludes that SQM and BQM might be independently viable *indefinitely*. This would mean that Cushing's historical contingency expresses itself as a permanent feature of the unfolding of quantum mechanics.

As stressed by Steven French, the problem of theory individuation is relevant here.[16] Many attempts to refute a contingency thesis made on the basis of BQM, and also to dispute BQM itself, rely on attacking BQM's status as a legitimate alternative *theory.* Instead, BQM is often framed as a mere *reformulation* of SQM. An early example is Werner Heisenberg's assertion that BQM is SQM's "exact repetition in a different language."[17] The key evidence for claims such as this is the shared formalism of SQM and BQM. Also relevant are the symmetries that Soler's criteria appeal to, especially that SQM and BQM are empirically equivalent and share comparable goals. The challenge for contingency is therefore that SQM and BQM may not constitute legitimate scientific alternatives. If this is the case, then Soler's claim that SQM and BQM are evidence of the potential for long-term contingent scientific progress is undermined, in that the continued progressiveness of SQM and BQM is only a result of their being the same theory.

---

[12] Soler, L. (2008). 'Revealing the Analytical Structure and Some Intrinsic Major Difficulties of the Contingentist/Inevitabilist Issue.' *Studies in History and Philosophy of Science.* Vol. 30, no. 2, p. 233.
[13] So that e.g. we do not treat quantum field theory as an element of SQM.
[14] Soler 2015, 'Contingentists Should Not Care About the "Put-Up-Or-Shut-Up" Demand,' p. 70.
[15] Ibid., p. 70.
[16] French, S. (2008). 'Genuine Possibilities in the Scientific Past and How to Spot Them.' *Isis*. Vol. 99, no. 3, pp. 568-575. 'Theory individuation' is just the means by which scientific theories are characterised in order to be individually distinguished.
[17] Heisenberg, W. (1955). 'The Development of the Interpretation of the Quantum Theory.' *Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of his Seventieth Birthday.* P. 18.

ISSN: 2653-3146

Such arguments rely upon an instrumentalist approach, i.e. on the assertion that theories are defined by the relationship between techniques and results. Thus, SQM and BQM might be accounted the same theory in that they employ the same formalism and make the same predictions. But this seems problematic, as an instrumentalist approach to theory individuation risks making physical ontology redundant, and more or less "neglect[s] the intellectual contents" of 'theory' in general.[18] Also problematic is that there appears to be inconsistency insofar as how such an instrumentalist approach is applied. For instance, if we assume instrumentalism, then the measurement problem[19] for SQM evaporates. However, amongst those that seek to dismiss BQM as a reinterpretation of SQM, the measurement problem is not thus dismissed.[20] Although this inconsistency does not fully repudiate an instrumentalist approach, it does indicate a reluctance to render physical ontology completely redundant — a necessary step if SQM and BQM are to be understood as interpretations of the same theory. This reluctance is reasonable, since it seems uncontroversial to assert that theories ought to be considered meaningfully descriptive. At the least, descriptive elements such as physical ontologies ought to be acknowledged as theoretical elements.

Despite this, it seems likely that decisive theory individuation is unreachable, on the basis that such is clearly value-reliant regarding "what essentially matters in physics."[21] But even if SQM and BQM *are* alternative formulations of the same theory, questions regarding the contingency or inevitability of these formulations remain. The divergent ontologies of SQM and BQM, in particular, might suggest that scientific practice, and thus at least potentially the unfolding of quantum mechanics in general, could be influenced by contingent preference for one formulation over another. With this caveat in place, I will continue to refer to SQM and BQM as distinct theories.[22]

---

[18] Levy-Leblond, J. (2015). 'On the Plurality of (Theoretical) Worlds.' *Science as It Could Have Been: Discussing the Contingency/Inevitability Problem*. P. 341.

[19] The descriptive gap between the superpositions posited by SQM and the exact values produced when measurement occurs, a significant conceptual issue if research on the basis of SQM is taken to be capable of producing a complete description of quantum systems.

[20] See e.g. Legget, A. (2002). 'Testing the Limits of Quantum Mechanics: Motivation, State of Play, Prospects.' *Journal of Physics: Condensed Matter*. Vol. 14, no. 15, pp. R415-451; Leggett, A. (2005). 'The Quantum Measurement Problem.' *Science*. Vol. 307, no. 5711, pp. 871-872.

[21] Soler 2015, 'Contingentists Should Not Care About the "Put-Up-Or-Shut-Up" Demand,' p. 73.

[22] Theory individuation is not so problematic for Cushing. Cushing's thesis is that in cases of actual underdetermination, contingent sociological and psychological factors can play a decisive role regarding which underdetermined alternative is accepted by the scientific community. This thesis remains intact regardless of whether 'theory' is supplanted by 'formulation', or 'interpretation.' Soler, while referring to these sociological and psychological factors as a *source* or mechanism of contingency, asserts that SQM and BQM are evidence of the potential for alternative theories to be equally successful programs in the long-term. That is, Soler sees the divided state of quantum mechanics as providing evidence for the contingent nature of science in general, that contingent factors can and will lead to alternative theories and ultimately shape the course of scientific progress.

Regardless, the important difference between Cushing's claim and Soler's utilisation of his case-study can be characterised as follows: Cushing proposes that, in cases of actual underdetermination, whichever alternative is ultimately accepted by the scientific community might be contingent. Soler, meanwhile, asserts that Cushing's case study is an example that can serve as evidence for a broader thesis regarding contingency — that contingent factors might always or often have the potential to produce alternative theories and research programs, and that these alternatives have long-term ramifications for scientific courses of events. Soler thus views Cushing's case as indicative of a general principle about scientific progress — she appears to hold the view that contingent initial conditions are always, or at least often, relevant to the historical progression of science.[23] In contrast to Cushing, for Soler, this claim appears to hold firm regardless of whether or not contingent alternatives can be specifically identified.

3. Inevitabilism and its Contingent Implications

In this section, I detail Hacking and Weinberg's accounts of inevitability in order to demonstrate how inevitabilism functions in the context of SQM versus BQM, and also to show how inevitabilism appears cogent with contingentism as presented above. If Hacking and Weinberg's arguments are accepted, even Soler's more general and long-term contingency thesis is affected only to the extent that (relatively minor) limitations upon *what* can be contingent in the long-term must be imposed.[24] In accomplishing this objective, I first identify what Hacking and Weinberg *take to be inevitable*. I then infer what seems permitted as contingent on that basis.

Hacking and Weinberg are representative of strong inevitabilism, based on Hacking's scale of contingency, which ranges from one (strong inevitability) to five (strong contingency).[25] Hacking places his arguments at two on this scale, and notes that Weinberg makes even stronger claims.[26] Presumably, Hacking would rank Weinberg's arguments as one on this scale. This seems plausible since Weinberg himself asserts that science follows the "pull of reality"[27] and appears to adhere to

---

This can be observed in Soler's (2015: 98) conviction that a "pluralist regime" could produce multiple scientific alternatives that are currently stymied through scientific monism

[23] Soler 2008, 'Revealing the Analytical Structure and Some Intrinsic Major Difficulties of the Contingentist/Inevitabilist Issue,' p. 232.

[24] Soler does not *specifically* advocate for what I take to be her 'unlimited' account, but she is somewhat non-specific regarding what she takes to be contingent. That being so, I frame my proposed reformulation of long-term contingency as amendments to Soler's account. Soler (2008: 233) does emphasise contingent ontologies as being particularly important to her argument. Since the amendments that I propose (below) specifically maintain that ontology can be radically contingent in the long-term, I anticipate that Soler would not object to the limitations that I place on her account.

[25] Hacking, I. (1999). *The Social Construction of What?* P. 99.

[26] Ibid., pp. 89-92.

[27] Weinberg, S. (2001a). *Facing Up: Science and Its Cultural Adversaries*. P. 103.

convergent realism, as indicated by the title of his 1994 book, *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*.[28] Weinberg is arguably "as boldly inevitabilist as it gets."[29] This circumstance is significant if, as I aim to demonstrate, even these strong inevitabilists permit a reasonably high degree of potential contingency.[30]

Hacking's focus is on the contingency or inevitability of scientific *results*. Hacking asks:

> If the results R of a scientific investigation are correct, would any investigation of roughly the same subject matter, if successful, at least implicitly contain or imply the same results?[31]

He responds:

> The answer to a clear question about some aspect of the world is determined by how the world is. When a question is a live one, and there is a context in which there are ways of addressing the question… then aspects of the world determine what the answer is, even though only people in a scientific society find out the answer. That is a difference… between metaphysics, what the answer is, and epistemology, whether we find the answer. The content of possible knowledge — the answers to live questions, once… asked — [is] not affected by [contingent factors]. The form of possible knowledge, the questions that in the course of research made sense, [is] affected.[32]

What does this mean in the case of quantum mechanics? Answering this question relies crucially upon what Hacking means by results, which remains ambiguous. Interestingly, Hacking appears to amalgamate 'results' and 'answers to scientific questions,' in that he treats the identification of a determinant for answers to scientific questions (the natural world) as evidence that scientific results are inevitable. But if we define Hacking's 'results' as 'answers to scientific questions' and apply this definition to the case of SQM and BQM, it does not yield a consistent account. For example, if we take the empirical claims of SQM and BQM, we might say that they produce the same results in that they make the same predictions. But

[28] The view that successful science necessarily progresses towards a universally accurate theoretical account of how the world is.
[29] Kinzel, K. (2015). 'State of the Field: Are the Results of Science Contingent or Inevitable?' *Studies in History and Philosophy of Science*. Vol. 52, p, 59.
[30] I do not take this 'scale of contingency' particularly seriously, since numerically ranking contingency/inevitability claims ignores the significant complexity of the positions that constitute the debate (see Martin 2013). I include it only to situate the supposed strength of Hacking and Weinberg's claims within the broader debate, so that the ramifications of my argument can be made clear.
[31] Hacking, I. (2000). 'How Inevitable are the Results of Successful Science?' *Philosophy of Science.* Vol. 67, p. S61.
[32] Ibid., p. S69-70.

ISSN: 2653-3146

when asking which physical processes *cause* these empirical results, SQM and BQM are in substantial disagreement. A claim to the effect that the natural world specifically determines these substantively contradictory descriptive claims seems difficult to defend, and so it seems that 'results' and 'answers to scientific questions' cannot be synonymous, at least in this case.

Attending to Hacking's form/content distinction may assist us here. This is partially elaborated when Hacking addresses Andrew Pickering's contingency thesis regarding a 'quarky' particle physics.[33] Hacking argues that Pickering's claim is not that quarks, the object, are contingent, but that "the *idea* of quarks... might be [contingent]."[34] "It was a highly contingent fact that human beings would ever form an idea of [quarks], but the existence of [quarks] is not contingent on any human thought or action."[35] Being informed by Hacking's entity realism, this perspective might be criticised on a number of fronts, but it provides some insight into how the form/content distinction functions for Hacking. Hacking acknowledges "that the 'forms' of scientific knowledge could have been different, yet still, we would be recognisably exploring the same aspects of nature."[36] On this basis, we can infer that since the 'ideas' of quarks might be different, and quarks are actual entities, Hacking understands that a non-quarky particle physics would be a different *form* of particle physics, but that the *content* of that physics would inevitably refer to quarks.[37]

How then does this clarify what Hacking means by 'results,' and hence how his account operates in relation to quantum mechanics? As noted, SQM and BQM substantively disagree about the physical processes that give rise to their empirical claims. But for Hacking, both are inevitably exploring the same aspects of nature — quantum systems. Their disagreement regarding physical processes makes SQM and BQM different *forms* of possible knowledge regarding quantum mechanics, in that they present different *ideas* about quantum systems. The descriptive claims of SQM and BQM therefore cannot be 'results' in Hacking's sense, since he associates results with inevitable content rather than contingent form.

Understanding this enables us to resolve the inconsistency that I indicated above. The descriptive claims of SQM and BQM, which are in substantive disagreement, are different forms of quantum mechanics. But we can understand their empirical claims, since SQM and BQM make the same predictions, to be the results of quantum mechanics: I propose that Hacking's 'results' ought to be interpreted in this fashion, as meaning empirical claims. So, for Hacking, it is the shared empirical claims of SQM and BQM that are inevitable. This then allows the SQM/BQM divergence to

[33] Pickering, A. (1984). *Constructing Quarks: A Sociological History of Particle Physics.*
[34] Hacking 2000, 'How Inevitable are the Results of Successful Science?,' p. S61, emphasis in original.
[35] Ibid., p. S62.
[36] Ibid., p. S71.
[37] Hacking does not actually suggest that a non-quarky particle physics is possible.

serve as evidence for Hacking's inevitabilism. Would any investigation of quantum systems produce the same results? In that SQM and BQM produce the same empirical claims, and these are understood as the results that Hacking refers to, Hacking's affirmative answer seems reasonable.

With this understanding in place, we can now ask what implications Hacking's inevitabilism has for the contingency theses recounted above. Cushing's claim is that the choice between SQM and BQM might be contingent on social and psychological factors.[38] It is not obvious that Hacking's claim is in any way inconsistent with such a statement, provided that we interpret SQM and BQM as different scientific forms in Hacking's sense. Hacking specifically admits that forms might be contingent, while Cushing specifically excludes the formalism of quantum mechanics — through which the empirical claims of SQM and BQM are derived — from his contingency thesis. For Hacking empirical claims are inevitable, and through this, we can infer that Hacking allows that scientific ontologies can be contingent. If this is accepted then the upshot is that, at least in the case of quantum mechanics, Hacking's inevitability permits Cushing's contingency.

Because Soler does not limit her contingency thesis as Cushing does, her contingentism may appear less compatible with Hacking's inevitabilism. Soler frames her long-term contingency as universally applicable to scientific courses of events.[39] In contrast, Hacking holds that no matter how scientific results are formulated (e.g. as SQM or BQM) they cannot be "incompatible," since scientific content is determined by aspects of nature.[40] Hacking is therefore suggesting that any contingency is limited by the determined character of scientific results, and so contingency cannot be as radical as Soler's thesis suggests. But, as Hacking notes, determining the compatibility or incompatibility of diversely formulated scientific results is not straightforward.[41] When diffuse results produced under different standards or epistemic assumptions are called compatible, this is typically dependent upon 'translating' one set of results into alignment with the standards under which the other set of results was produced. But this process "does not leave everything the same."[42] This is because 'results' are typically entangled with broader theory. This becomes evident if we take what might seem to be a clear case of relative compatibility, SQM and BQM's shared results. If BQM's results are held to be *compatible* with SQM, this will be because SQM can account for BQM's results under SQM's ontology. But SQM's results can be compatible with BQM on the same basis. Compatibility of results in no way resolves the SQM/BQM dilemma. As such,

[38] Cushing 1994, *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*, p. 100.
[39] Soler 2008, 'Revealing the Analytical Structure and Some Intrinsic Major Difficulties of the Contingentist/Inevitabilist Issue,' p. 232.
[40] Hacking 2000, 'How Inevitable are the Results of Successful Science?,' p. S71.
[41] Ibid p. S67.
[42] Ibid p. S67.

compatibility of results in itself does not tell us much about the aspects of nature that are determining these results. What Hacking's criterion of compatibility tells us is that since results inevitably refer to aspects of nature, they are likely to admit translation between standards. But it cannot tell us which standard is more accurate. It does nothing to tell us whether SQM or BQM is descriptively closer to actual quantum systems.

In short, if we amend Soler's argument so that long-term contingency can occur, but that empirical claims are inevitably determined by aspects of nature, then Hacking's inevitabilism tacitly permits Soler's contingency as well. Regarding SQM and BQM, the claim becomes that both may prove long-term viable alternatives, but that the results of either model will inevitably refer to actual quantum systems. This is a relatively minor amendment regarding *what can be contingent* on Soler's account. Empirical claims about the world will inevitably refer to the "natural world, indifferent to human beings."[43] But this does not then refute Soler's specific claim that SQM and BQM are potentially long-term viable alternatives, since they disagree *not* on their empirical claims (their results) but on their account of the physical processes that cause such, i.e. their ontologies. Although the empirical claims of SQM and BQM appear compatible in the sense that I have outlined, this does not itself nullify the substantive contingency that Soler postulates. As has already been shown, Hacking's inevitabilism permits contingency of ontology. Therefore, Hacking's inevitabilism also permits Soler's long-term contingentism.

As noted, Weinberg is a particularly strong inevitabilist. But although he claims that science follows the "pull of reality,"[44] one of my goals here is to show that Weinberg's arguments actually exclude definitive convergent realism. Weinberg writes:

> The achievements of science [are] permanent. What changes is our understanding of why theories are true, and also our understanding of the scope of their validity. For instance, at one time we thought that there was an exact symmetry in nature between left and right, but then it was discovered that this is only true in certain contexts and to a certain degree of approximation. But the theory… was not a simple mistake. Within its scope of validity, this symmetry has become a permanent part of science.

He continues:

> There is a 'hard' part of modern physical theories that usually consists of the equations themselves, together with some understandings about what the symbols mean operationally and about the sorts of phenomena to which they apply. Then there is a 'soft' part; it is the vision of reality that we use to explain to ourselves why the equation works. The soft part does change… but after our theories reach their

[43] Ibid., p. S70.
[44] Weinberg 2001a, *Facing Up: Science and Its Cultural Adversaries*, p. 103.

> mature forms, their hard parts represent permanent accomplishments. I think that physical theories are like fixed points… Our starting points may be culturally determined, our paths may be affected by our personal philosophy, but the fixed point is there nonetheless. It is something toward which any physical theory moves.[45]

Weinberg clarifies that this 'hard' part of physical theories can be partially characterised by predictive capacity.[46] Elsewhere, Weinberg writes that the 'permanence' of these accomplishments is twofold. Firstly, a means of reliable prediction will always 'function,' will always produce the same predictions regardless of how these predictions are interpreted.[47] Secondly, assuming that research continues, the ability to reliably predict phenomena will always feature as something that ought to be explained, despite any theory change that might occur.[48] This seems plausible enough — a key goal of many scientific enterprises is to provide explanations of empirical data and consistencies.

How do Weinberg's arguments function in the case of SQM and BQM? It appears that the formalism of quantum mechanics — Schrödinger's equation, from which both SQM and BQM are derived — ought to be accounted a 'hard' part of quantum mechanics, since it is used to derive reliable predictions. Leaving aside operational factors, the other hard aspect of quantum mechanics is the understanding that this formalism applies to quantum behaviour. If 'soft' parts of physical theories are visions of reality that explain why equations work predictively, then the divergent ontologies of SQM and BQM fit nicely into this category. As I have previously noted, these ontologies constitute very different conceptions of the physical processes which produce the empirical claims of quantum mechanics. This is consistent with Weinberg's anticipation that the descriptive content of SQM is likely to change in response to problems with quantum cosmology.[49] Although this is framed fairly vaguely, and Weinberg never substantively addresses BQM, he allows that the descriptive claims of SQM are *impermanent* or changeable, and this indicates that they must be soft on his account.

There are some obvious parallels between Weinberg's soft and hard parts of physical theories, and Hacking's form and content distinction. In the case of SQM and BQM, what Weinberg allows as 'soft' directly aligns with what Hacking regards as the potentially contingent forms of quantum mechanics — both refer to the descriptive or ontological accounts that seek to make sense of the empirical claims of quantum mechanics in general. Regarding the 'hard' parts of quantum mechanics, and Hacking's content, there is a (minor) difference between the two. Weinberg

[45] Ibid., p. 126.
[46] Ibid., p. 125.
[47] Weinberg 2001a, *Facing Up: Science and Its Cultural Adversaries*, pp. 118, 199.
[48] Weinberg, S. (1994). *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*, pp. 127-8.
[49] Ibid., p. 67.

ISSN: 2653-3146

emphasises mathematical devices, like Schrödinger's equation, that allow for reliable predictions. Hacking's argument emphasises actual empirical claims and predictions. But these views are readily reconcilable in this case, since a) the formalism is what allows for the empirical claims of quantum mechanics to be produced, and b) the empirical claims of quantum mechanics are what validate the formalism.

In this sense, Weinberg's inevitability argument as applied to SQM and BQM is actually fairly similar to Hacking's. The 'fixed point' of a physical theory of quantum mechanics is presumably related to the actual nature of quantum systems. Since Schrödinger's equation can make reliable predictions regarding quantum behaviour, it must refer to this fixed point. Therefore, if investigation into quantum behaviour occurs, it will — assuming success — inevitably strike upon this consistency of quantum systems, which makes reliable predictions possible through a formalism. All this occurs independently of how the physical processes that cause such are understood. Another minor difference is that Weinberg's account suggests that quantum formalism as a means of reliable prediction will inevitably feature as something to be explained by future theory. This seems plausible and does not directly contradict Hacking's claims. A more significant difference is that Weinberg's arguments suggest that successful scientific investigation into quantum behaviour will *inevitably produce increasingly accurate approximations* of quantum systems. Hacking only claims that scientific results will inevitably refer to actual quantum systems — not that this will result in more accurate accounts. Weinberg's argument is that, in the long-term, accounts of quantum systems will inevitably become increasingly accurate representations of actual quantum systems. The implication is that any contingency that does occur during scientific investigation is ultimately insignificant in the long-term.

But does Weinberg's long-term inevitabilism actually follow from his argument? I think it is fairly obvious that it does not. Let us see how much contingency remains possible under Weinberg's schema of inevitability. As with Hacking, it is not clear that Weinberg is in any substantive disagreement with Cushing. If we understand SQM and BQM's differences as a difference of 'soft' scientific parts, then Weinberg is willing to admit that these can be affected by social and psychological factors, at least temporarily. Cushing does not argue that the formalism of quantum mechanics — its 'hard' part — is contingent, but only that historical contingency has affected the unfolding history of quantum mechanics to date.

At face value, Weinberg's long-term inevitability claim *does* seem to come into significant conflict with Soler's views. Whilst Soler argues that alternative physical theories are possible in the long-term, Weinberg argues that scientific investigation leads inevitably to specific theoretical accounts. Soler argues that SQM and BQM could remain viable alternatives in the long run, whereas Weinberg's claim is that

any differences will inevitably be resolved. But Weinberg's argument, essentially presupposing convergent realism, does not stand up to scrutiny on his own account. For Weinberg, it is *invariantly predictive mathematical structures* — that is, nomological structures — that are inevitable. As it currently stands, for Weinberg it is the shared formalism of SQM and BQM that is inevitable. It may be that investigation into quantum behaviour will inevitably lead to more accurate means of prediction in this sense. But it is evident that ontology — the soft 'vision of reality' by which we explain the predictions made through nomological structures — cannot be derived from such. Weinberg's "pull of reality"[50] appears, by his own account, to be capable of producing *only* soft ontologies that remain susceptible to change based on social and psychological factors. The substantive disagreements between SQM and BQM cannot be decisively resolved via the hard parts of quantum mechanics to which Weinberg appeals.

The upshot is that Weinberg's account permits most of Soler's long-term contingency, although this does require another amendment to Soler's argument. In light of Hacking's argument, I allowed that in the long-term SQM and BQM would find their results inevitably determined by aspects of nature, although I suggested that the implications of this inevitability are necessarily ambiguous. Now, I propose that we ought to amend Soler's argument so that the nomological structure of quantum mechanics might be inevitable. This might appear to be a major concession to the inevitabilist, but it proves otherwise. To plausibly suggest that the nomological structure of quantum mechanics is contingent would require proposing a counterfactual alternative, and in this case, an alternative is nearly inconceivable. Moreover, as has been shown, an inevitable nomological structure is not related to the evidence of contingency found in the case of SQM and BQM — it cannot lead to a reconciling or definitive quantum ontology. So, although Weinberg's arguments regarding the inevitability of the 'hard' parts of quantum mechanics appear plausible, this does not resolve the substantive disagreements between SQM and BQM which most strongly suggest contingency. Weinberg (who, recall, may be as boldly inevitabilist as it gets) thus permits both Cushing's historical contingency, and a slightly amended form of Soler's long-term contingency, while specifically allowing for radically contingent scientific ontologies.

4. Concluding Remarks

This paper has aimed to demonstrate that, at least in regard to quantum mechanics, the contingency/inevitability debate is more or less lacking in legitimate philosophical tension. This is largely the result of a failure of conceptual clarity, wherein participants do not specify the targets of their arguments. Once this clarity

[50] Weinberg 2001a, *Facing Up: Science and Its Cultural Adversaries*, p. 103.

is achieved, it becomes fairly clear that the nominally competing claims regarding SQM and BQM are mutually sustainable, in that contingentists and inevitabilists target different aspects of quantum mechanics generally. It is significant that under scrutiny it becomes clear that not even the strongest inevitabilist is willing to argue that scientific ontology regarding unobservables is definitively determined by epistemic factors. The significance of social and psychological determinants in forming scientific ontology appears to be generally, if on the part of inevitabilists characteristically tacitly, accepted. We might reconsider whether arguments permitting such a large degree of historical contingency are deserving of the label 'inevitabilism.'

Finally, note that although quantum mechanics provides an excellent setting for the discussion of contingency/inevitability regarding science, the scope of my claims is limited by my focus on SQM and BQM. If more reliable general claims regarding the debate are to be formed, my approach will need to be applied to a broader set of cases.

ISSN: 2653-3146

References

Bohm, D. (1952). 'Reply to a Criticism of a Causal Reinterpretation of the Quantum Theory.' *Physical Review.* Vol. 87, pp. 389-390.

Bohm, D. (1987). 'Hidden Variables and the Implicate Order.' *Quantum Implications*, pp. 33-45. London: Routledge & Kegan Paul.

Bricmont, J. (2016). *Making Sense of Quantum Mechanics*. Cham: Springer.

Cushing, J. (1994). *Quantum Mechanics: Historical Contingency and the Copenhagen Hegemony*. Chicago: The University of Chicago Press.

Dirac, P. (1928). 'The Quantum Theory of the Electron.' *Proceedings of the Royal Society of London A.* Vol. 117, no. 778, pp. 610-624.

Durr, D, Goldstein, S, Norsen, T, Struyve, W, & Zanghi, N. (2014). 'Can Bohmian Mechanics be Made Relativistic?' *Proceedings of the Royal Society A*. Vol. 470, no. 2162.

Falciano, F, Pinto-Neto, N, & Struyve, W. (2015). 'Wheeler-DeWitt Quantization and Singularities.' *Physical Review D*. Vol. 91, no. 4, 043524.

French, S. (2008). 'Genuine Possibilities in the Scientific Past and How to Spot Them.' *Isis*. Vol. 99, no. 3, pp. 568-575.

Hacking, I. (1999). *The Social Construction of What?* Cambridge, Mass.: Harvard University Press.

Hacking, I. (2000). 'How Inevitable are the Results of Successful Science?' *Philosophy of Science*. Vol. 67, pp. S58-S71.

Heisenberg, W. (1955). 'The Development of the Interpretation of the Quantum Theory.' *Niels Bohr and the Development of Physics: Essays Dedicated to Niels Bohr on the Occasion of his Seventieth Birthday*, pp. 12-29. New York: McGraw-Hill.

Kinzel, K. (2015). 'State of the Field: Are the Results of Science Contingent or Inevitable?' *Studies in History and Philosophy of Science*. Vol. 52, pp. 55-66.

Legget, A. (2002). 'Testing the Limits of Quantum Mechanics: Motivation, State of Play, Prospects.' *Journal of Physics: Condensed Matter*. Vol. 14, no. 15, pp. R415-451.

Leggett, A. (2005). 'The Quantum Measurement Problem.' *Science*. Vol. 307, no. 5711, pp. 871-872.

ISSN: 2653-3146

Levy-Leblond, J. (2015). 'On the Plurality of (Theoretical) Worlds.' *Science as It Could Have Been: Discussing the Contingency/Inevitability Problem*. Pp. 335-358. Pittsburgh: Pittsburgh University Press.

Martin, J. (2013). 'Is the Contingentist/Inevitabilist Debate a Matter of Degrees?' *Philosophy of Science.* Vol. 80, no. 5, pp. 919-930.

Pauli, W. (1927). 'Zur Quantenmechanik des Magnetischen Elektrons.' *Zeitschrift fur Physik A Hadrons and Nuclei*. Vol. 43, pp. 601-623.

Pickering, A. (1984). *Constructing Quarks: A Sociological History of Particle Physics.* Chicago: The University of Chicago Press.

Soler, L. (2008). 'Revealing the Analytical Structure and Some Intrinsic Major Difficulties of the Contingentist/Inevitabilist Issue.' *Studies in History and Philosophy of Science.* Vol. 39, no. 2, pp. 230-241.

Soler, L. (2015). 'Introduction: The Contingency/Inevitability Debate' and 'Contingentists Should Not Care About the "Put-Up-Or-Shut-Up" Demand.' *Science as It Could Have Been: Discussing the Contingency/Inevitability Problem*. Pp. 1-42, 45-98. Pittsburgh: University of Pittsburgh Press.

Weinberg, S. (1994). *Dreams of a Final Theory: The Scientist's Search for the Ultimate Laws of Nature*. New York: Vintage.

Weinberg, S. (2001a). *Facing Up: Science and Its Cultural Adversaries.* Cambridge, Mass.: Harvard University Press.

Weinberg, S. (2001b). 'Physics and History.' *The One Culture? A Conversation About Science.* Pp. 116-127. Chicago: University of Chicago Press.

ISSN: 2653-3146

Undergraduate Philosophy Journal of Australasia www.upja.online
Vol. 3, December 2021

Actions of Trust and their Cognitive Motivation

Christian Carbonell[1]

University of Barcelona

Abstract

In this paper I offer a systematic account of actions of trust and inquire into their cognitive motivation. I first develop the distinction and relationship between attitudes and actions of trust, and then assess Paul Faulkner's thesis that the Humean model cannot explain the cognitive motivation of some actions of trust under circumstances of uncertainty. While I will accept his diagnosis, I will contend that a weaker version of the Humean model could provide this explanation. My proposal will be an attempt to show why some doxastic characteristics of trust would allow for this analysis. In particular, I will show how the nature of the reliance relation, which constitutes actions of trust, requires that trustors believe in the possibility of accomplishing their intentions by means of the trusted party's collaboration. I will argue that this means-end belief can cognitively motivate trust even in situations where the trustor is uncertain as to whether the trusted party will prove trustworthy.

1. Introduction

On the Humean account, a motivating reason for acting is a pair formed by a desire and a means-end belief.[2] If an agent has a motivating reason to ψ, this means that she desires X and that she has a belief representing the action of ψ-ing as a means to the attainment of X. Thus, her action could be analysed in terms of a desire-belief pair and we could conclude that, other things being equal, she acted as she did partly

[1] *Christian Carbonell is pursuing a Master's degree in analytic philosophy at the University of Barcelona, Spain. He is interested in human behaviour at large, with a special interest in its ethical and epistemic dimensions, and is currently studying the effects that different kinds of ignorance have on intentional and voluntary action.*

[2] See Smith 1987. Also, see Davidson 1963 and his notion of "primary reason".

ISSN: 2653-3146

because she was cognitively motivated to so act – it was her means-end belief that, in view of her desire, ultimately determined her to ψ.[3]

Now, the question may be raised as to whether all actions are necessarily motivated in these terms. For instance, could there not be a case where an agent is motivated to ψ although she does not believe that ψ-ing will make her satisfy her desire? According to Paul Faulkner[4], this is at least possible when it comes to some actions of trust, as trusting can sometimes involve uncertainty about outcome - a leap of faith as it were. In some specific situations where the trustor *A* desires something following from the trusted party's behaviour, it can happen that *A* does not believe that the trusted party *S* will behave as expected, and so *A* might not believe that her action of trusting *S* is a means to the satisfaction of her desire. In short, it can happen that *A* lacks a belief of the sort 'if I trust *S* to ϕ, I will get X' and yet that *A* is motivated to trust *S* to ϕ anyway. It then follows that trustors are not invariably required to have a particular motivating reason *R* formed by a desire for X and a means-end belief of the type just seen.

In this paper, I shall present and assess Faulkner's proposal that some actions of trust need not be motivated by *R,* and inquiry into whether there is an alternative motivating reason that could cognitively motivate trust even under circumstances of uncertainty. I will agree with Faulkner that the Humean model cannot always account for the motivation of trust, but will contend that a weaker version of the model could do the job instead in those situations. In particular, I will argue that the actions of trust Faulkner has in mind can be motivated under such circumstances by a weaker motivating reason R_w partly constituted by a weak means-end belief of the sort 'if I trust *S* to ϕ, then it is possible that I get X'. My point being that when *A* does not need *R* she still needs R_w, since the belief that one's action of trust is a means that makes possible the satisfaction of one's desire is needed to trust even under conditions of uncertainty. I will base this contention on the relationship between trust, competence, and intentionality.

2. What is it to trust?

2.1 Trust can be an attitude or an action

Faulkner's thesis concerns the cognitive motivation of actions of trust. More concretely, his argument against the Humean concerns *actions of affective trust*, as he calls them. Before presenting his thesis in Section 2, it would be helpful to provide a

[3] Recall Hume's famous sentence "reason is, and ought only to be the slave of the passions" (Hume 1739/1960, p. 415) and the fragment where he explains "according as our reasoning varies, our actions receive a subsequent variation" (Ibid., p.414).
[4] Faulkner 2014.

definition for this type of action. To this end, I will first differentiate actions of trust from attitudes of trust.

Trust can refer to an attitude or it can refer to a certain type of action that is done with that attitude. Thus, if trust as an attitude is defined as a state of mind where the trustor has positive expectations towards the trusted party's behaviour (e.g. Fred expects that George will pick him up at the airport, as opposed to the negative expectation that George will *not* pick him up), then an action of trust consists in an action that the trustor performs with those expectations.[5] In particular, *A*'s action of trust is an action that in being performed *A* relies on *S*'s ϕ-ing and positively expects something of *S*'s behaviour (e.g., Fred's waiting for George or Fred's not taking of a taxi can be actions of trust, for in so doing Fred relies on George's picking him up at the airport and expects that George will do so). Remove the positive expectation and you have mere reliance; remove reliance and you have simply the attitude of trust.

I will provide a definition of reliance shortly, but the point I want to stress now is that, contrary to the attitude of trust, the defining feature of action of trust is that they involve a relation of reliance. For the attitude alone does not imply reliance but at most the *disposition to rely* on someone.

Consider a case where you trust your friend would not tell your romantic secrets to anyone, but you have not even told her yet who your crush is. In this situation, you would have just an attitude of trust towards your friend. You would simply expect that if you were to tell her any secret she would not disclose it, and although this expectation would dispose you to rely on her, you would not be relying on her yet. Now, imagine another situation where you had an urge to confess your love for someone *and* decided to confide the secret to her. As I see it, only in this latter case would you be relying on her behaviour.

Roughly speaking, reliance is as a relation where *A*'s action depends on *S*'s ϕ-ing where the nature of this dependence is intention-specific. Therefore, it could be said that only if you told her the secret you would rely on your friend not disclosing it, in the sense that an intention of your telling of the secret (e.g., the intention of putting your mind at ease by privately confessing your love) would depend on your friend's behaviour (i.e., on her keeping the secret) to be accomplished. It is because her keeping the secret would help you accomplish the circumstantial intention of your action that you would be relying on her. So if in this situation you also put an expectation on her, you would be then performing an action of trust. The telling of

[5] As we will see, there can be two relevant kinds of expectations when it comes to trust, namely, predictive and normative. This example here is an example of a positive predictive expectation (which consists in the belief that *S* will ϕ), but I could have equally used an example of a positive normative expectation (which consists in the belief that *S* should ϕ, which negation would consists in the belief that *S* should *not* ϕ). So a positive expectation is either the belief that something will be the case or the belief that something should be the case.

your secret would qualify as an action of trust because in doing so you would be relying on your friend keeping the secret and you would expect that she would keep it.

This is why I consider actions of trust, contrary to mere attitudes, to involve an intention-specific dependence between the trustor and the trusted. Attitudes of trust do not involve this dependence in themselves precisely because when all you have is the attitude, you do not yet rely on someone. When all you have is just the attitude, you have not yet set out to accomplish any intention of yours by means of the trusted party's collaboration, so you do not yet depend on her doing anything. On the other hand, actions of trust do involve this intention-specific dependence in which reliance consists.[6]

I would like to make one last point to avoid confusion. As I said, trust can refer to an attitude or an action. And the problem is that this twofold conceptualization does not match our ordinary usage of the term 'trust'. Indeed, on most occasions we use 'trust' to refer to the attitude alone, as when we say things like 'I trust my colleagues –they would never let me down'. Other times we use it to refer to a situation of mere reliance, as when we explain 'I had to trust the old bridge, for I had no choice but to cross the river'. So, since we are interested in actions of trust, following our daily trust talk might be philosophically misleading, because we use the term 'trust' in situations where there is no reliance (like in the former case: you had not asked your colleagues any favour yet, and so you do not rely on their doing something), as well as in situations where there is no positive expectation (like in the latter case: you did not positively expect the bridge would resist your weight, and probably you may have expected otherwise).

2.2 Trust can be predictive or affective

Once the distinction between attitudes and actions is clear, it is possible now to define two kinds of actions of trust, namely, actions of predictive trust and actions of affective trust. To this end, I shall compare two scenarios inspired by Annette Baier's passing reflection on trust and Kant's behaviour.[7]

---

[6] Perhaps there are actions that *are* actions of trust in themselves, presumably because they essentially involve reliance and positive expectations. In any case, the fact I want to stress here is that most actions of trust can be non-trusting actions as well. For instance, Fred's not taking of a taxi can be a non-trusting not taking of a taxi if the intention with which he does it does not depend on George's behaviour to be accomplished (i.e., if Fred does not rely on George). Fred could simply not take a taxi and do so with the intention of not spending more money, instead of doing it with the intention of being picked up by George. Since only this latter intention depends on George's picking Fred up to be accomplished (as Fred could save money by taking a walk home, say), the not taking of a taxi with the intention of not spending more money would not qualify as an action of trust. This shows that the relational state of the intention goes somewhat to define the trusting nature of the action.

[7] Baier 1986, p. 235.

Scenario 1 (S1)

Kant was very regular with his habits. Every day, exactly at the same time, he would go out for a walk. His neighbours from old Koenigsberg would regularly see him passing by the same places at more or less the same time. Because Kant was that regular in his exercise, it is not implausible to imagine a scenario where, say, the nurses of a hospital used him as a sort of reminder: every time he would pass by the hospital at a certain time, the nurses would give their patients their beverages against typhus.

Scenario 2 (S2)

Imagine that, after a few days without news from Kant, a preoccupied nurse reaches out to him. She lets him know about how she and her colleagues have been counting on his regular walks to successfully do their job. She also tells him about some catastrophic incidents that followed his unexpected not-going-out-for-a-walk. And she goes on to beg him to please pass by the hospital each day, as the lives of her patients are at risk. The philosopher therefore dutifully accepts her demand and promises to pass by every day at the same time he usually passes by.

What are the similarities and differences between both scenarios? The first relevant similarity is that the nurses stand in a relation of dependence with Kant's doings, where this dependence consists in relying on his punctual walks. By depending on Kant's punctuality, the nurses rely on his behaviour to accomplish the intention-specific task of giving the medicines to their patients on time. And another similarity, as I will show shortly, is that in both scenarios the nurses put a positive expectation on Kant's passing by at a certain time. Therefore, in each scenario, the nurses perform an action of trust, namely, the action of giving the medicines to their patients right when they see Kant passing by the hospital. For when they give the medicines to their patients under these circumstances the intention with which they do so depends on Kant's punctually passing by.[8]

As for the differences, the first divergence between both scenarios can be drawn in terms of the reactive attitudes that may be prompted. Whereas in S1 the nurses might have felt "disappointed with [Kant] if he slept in one day, but not let down by him, let alone had their trust betrayed"[9], in S2 they would have been entitled to present those sorts of attitudes (feelings of betrayal) towards him, after he had promised to punctually pass by everyday. And the second and most important

[8] It may seem wrong to say that the giving of a medicine (or the waiting at the airport, the not taking of a taxi, the telling of a secret, and so on) can qualify as an action of trust. As I see it, this should not seem wrong as long as we agree with the Anscombe-Davidson thesis that actions are single events that admit several descriptions. My point is then that the nurses' giving of a medicine when they see Kant passing by and the nurses' action of trust are the same action, the same event, albeit described in different ways. See Davidson 1969, and Anscombe 1979 for further discussion.
[9] Baier 1986, p. 235.

distinction, which explains the first, has to do with the nature of the positive expectation the nurses put on Kant. I borrow concepts from Martin Hollis[10] in saying that, whereas in S1 the nurses held a predictive expectation on the philosopher, in S2 the expectation is rather normative.

When predictive expectations are held, the one who expects has a certain positive expectation about the expectee's future behaviour, often based on her regularity. The nurses from S1 could have thought that Kant would perform again as he regularly does (and then use him as a reminder), and may have believed something like 'Kant will pass by at *t*'. Indeed they may have reasoned 'we can trust Kant will pass by at *t*, because he always does'. Basically, they took his walks to be a regular thing for him to do, and so predicted that he would do it again in the future at t and relied on his walks at *t* because of this. This explains why they would have not been entitled to feel betrayed: deceived predictions do not justify that reaction, just like unexpected weather variations do not justifiably count as betrayal. Now, when it comes to normative expectations, this is rather the other way round.

When expectations of this sort are held, the expector rather assumes the existence of some normativity governing the expectee's future behaviour. In particular, the former can just expect the latter to act as expected due to a certain sort of normativity she assumes to exist, regardless of any regularity she may be aware of. Thus the nurses of S2 at *t* relied on Kant's passing by, and they did also expect him to be passing by at *t*, precisely because they assumed the existence of certain normativity in his acceptance of the demand. Arguably, they may have thought promises bound the promisor to act as promised, and so they may have grounded their normative expectation on Kant's behaviour for this reason. They may have believed something like 'Kant should pass by at *t* because he has so promised', and they may have reasoned 'we can trust Kant will punctually pass by, because he has committed himself to'.[11]

The point to be made here is that these differences in expectation account for two ways in which trust can be understood, and consequently identify each scenario with a particular kind of action of trust. Following Faulkner's terminology[12], S1 concerns *actions of predictive trust* because they involve predictive expectations,

[10] Hollis 1998, p. 11.
[11] Katherine Hawley (2014) has argued that the source of this normativity as applied to trust is expressed commitment. Promises or agreements, as well as social norms or contracts, would therefore justify having normative expectations. In this sense, Hawley proposes that to trust *S* to do something is "to believe that she has a commitment to doing it and to rely upon her to meet that commitment." (Ibid., p. 10.)
[12] Faulkner 2007; Faulkner 2014.

ISSN: 2653-3146

whilst S2 concerns *actions of affective trust* because they involve normative expectations.[13] I turn now to define each kind of action.

As the name makes it explicit, the expectation characteristic of actions of predictive trust is a sort of positive prediction. In this case, Faulkner remarks, "the expectation that the trusted will do something is just the prediction or belief that they will do so".[14] Trusting in this sense is therefore "a matter of a belief about the future, an inductive inference".[15] And this is precisely what happens in S1. The nurses expected Kant to pass by given his regularity in so doing: they predicted that he would pass by again at a certain time as the result of inductive reasoning. Then, actions of predictive trust may be defined thus: *A*'s action is an action of predictive trust when in so acting *A* relies on *S*'s ϕ-ing and predictively expects *S* to ϕ. An action of predictive trust may be the act of using a thermometer to find out one's exact temperature, or the act of not bringing water to a meeting because there is one colleague who always brings enough for everyone, and so on.

But actions of trust can also be affective when the constitutive expectation is normative, that is, when the expectation consists in assuming that "the trusted party has certain [normative] motivations and act on the basis of these."[16] This is why the distinctive thought of the nurses in S2 is not 'Kant *will* pass by', but 'Kant *should* pass by': their expectation appeals to Kant's normative motivations to behave trustworthily rather than to his mere regular behaviour. No prediction is essentially involved there, but just a sort of demand concerning Kant's behaviour. So here is a definition for actions of affective trust: *A*'s action is an action of affective trusts when in so acting *A* relies on *S*'s ϕ-ing and normatively expects *S* to ϕ. An action of affective trust may be the act of handing money to a shopkeeper to get an apple in return, or the act of giving credence to a witness because he has sworn to tell nothing but the truth, and so on.

In conclusion, then, actions of affective trust share with actions of predictive trust the fact that the trustor relies on the trusted party's behaviour, and the fact that the former positively expects something of the latter. Where the two differ from each other is that in the first case, the trustor's expectation is essentially normative, and so she is entitled to have reactive attitudes towards the trusted when the trusted does not satisfy her normative expectations: we can easily imagine the nurses complaining to Kant 'we trusted you – and you failed us!'. In cases of predictive

[13] Faulkner dubbed this kind of trust «affective» because of its association with the kind of reactive attitudes that occur in affective, interpersonal relationships. See Faulkner 2014, p. 1978. We will see how this reactive attitudes are characteristic of affective trust shortly.
[14] Faulkner 2014, p. 1978.
[15] Faulkner 2007, p. 880.
[16] Faulkner 2014, p. 1978.

trust, however, these reactive attitudes neither tend to arise nor do they justifiably arise.[17]

3. A non-doxastic view of trust

3.1 Trust and the Humean account of motivation

Faulkner holds a non-doxastic view of trust, which means that he endorses the thesis that trust needn't involve nor entail the belief that *S* will ϕ. Specifically, he says so with regards to *actions of affective trust* under circumstances of uncertainty: "when it comes to acts of affective trust, the problem is that *A* can trust *S* to ϕ without... the belief that *S* will ϕ."[18] This is the reason why, according to him, the Humean analysis cannot sometimes account for the cognitive motivation of actions of affective trust. In a situation where *A* believes that *S*'s ϕ-ing is a means to the attainment of X and where *A* is trusting *S* to ϕ (and trivially believes so)[19], it is clear that lacking the belief '*S* will ϕ' would impede *A* from having a means-end belief of the sort 'if I trust *S* to ϕ, I will get X', which is precisely the sort of means-end belief that the Humean analysis would require that *A* have to be motivated to trust.[20]

In effect, the Humean analysis states that a motivating reason for *A* to ψ consists in a pair formed by a desire for an outcome X and a means-end belief representing her ψ-ing as a means to the attainment of X. Michael Smith proposed the following definition:

> R at t constitutes a motivating reason of agent A to ψ iff there is some X such that R at t consists of a desire of A to X and a belief that were he to ψ he would X[21]

---

[17] It may be helpful to note that "pure" actions of predictive trust (i.e., where the expectations had are just predictive) very often involve objects rather than people. As Hollis points out, "we trust one another to behave predictably in a sense that applies equally to the natural world at large. I trust my apple tree to bear apples, not oranges." (Hollis 1998, p. 10). This would partly explain why reactive attitudes do not justifiably arise when only our predictive expectations are deceived –objects are out of the realm of (social) normativity, and so their behaviour cannot be the object of justified resentment. Basically, we cannot put a normative expectation on objects or animals. Normative expectations are restricted to interpersonal relationships, and so it is affective trust.

[18] Faulkner 2014, p. 1979.

[19] Strictly speaking, *A*'s trivial belief is not 'I am trusting *S* to ϕ', but the belief 'I am ψ-ing' where the ψ-ing *is* an action of trust. Thus, for instance, the nurses in S2 would trivially believe that they are giving the medicines to their patients right when they see Kant passing by the hospital. I will henceforth use the verb 'to trust' and derivatives to refer to an action of trust unless it is made explicit that the trusting in question refers to the attitude.

[20] The rationale is the following. If *A* believes 'I am trusting *S* to ϕ' then she cannot believe 'if I trust *S* to ϕ, *S* will ϕ' unless she also believes '*S* will ϕ'. As a consequence, if *A* believes 'if *S* ϕ-s, I will get X' then she cannot believe 'if I trust *S* to ϕ, I will get X' unless she also believes 'if I trust *S* to ϕ, *S* will ϕ'.

[21] Smith 1987, p. 36. Notice I changed the original symbols of the quote to make them consistent with both Faulkner's and my symbolization. This change does not alter the original meaning of the quote, and may even make it clearer.

Thus, according to this analysis, why were the nurses motivated to affectively trust Kant? In other words, why did they give the medicines to their patients when they saw Kant passing by the hospital? Simply because they had a motivating reason of the sort *R*. That is, because they desired to give the medicines to their patients on time and believed something like 'if we give the medicines to our patients when we see Kant passing by the hospital, we will give the medicines on time'. However, Faulkner points out, this could perfectly be otherwise.

Faulkner would say that the nurses' action could have not been motivated by a desire-belief pair of that type, because they might have lacked the kind of means-end belief which constitutes *R*. In particular, they might have lacked a means-end belief of the sort 'if we give the medicines to our patients when we see Kant passing by the hospital, we will give the medicines on time' because they could have been uncertain as to whether Kant would punctually pass by and, accordingly, they could have not believed that Kant would eventually punctually pass by.[22]

So it seems that the Humean analysis cannot account for the cognitive motivation of actions of affective trust under circumstances of uncertainty. According to this analysis, as we have just seen,[23] the acquisition of a means-end belief of the sort 'if I trust *S* to ϕ, then I will get X' in the specific situation considered imposes the following conditions:

(C1) That *A* believes '*S* will ϕ'
(C2) That *A* believes 'if I trust *S* to ϕ, then *S* will ϕ'

The problem is, however, that affective trust can be motivated even in the absence of the belief '*S* will ϕ' and the belief 'if I trust *S* to ϕ, then *S* will ϕ ' due to its particular, cognitive characteristics. I will now present Faulkner's explanation of these characteristics, and in Section 3 I will show how a weaker version of the Humean analysis could accommodate them.

3.2 The cognitive characteristics of affective trust

Faulkner offers an illustration of a waiting husband to explain in detail the cognitive characteristics of affective trust:

[22] It is important to notice the difference between 'not believing that *p*' and 'believing that not *p*'. While the former means that one does not have a belief as of *p* (e.g., I don't believe it's raining), the latter means that one has a belief as of not-*p* (e.g., I believe it's not raining). Also, while the former is not sufficient for the latter (someone may not believe that it's raining and yet not believe that it's not raining –say, because she is uncertain about it because she has not yet collected sufficient evidence), it seems the latter is in fact sufficient for the former (if someone believes it is not raining, then one does not believe that it is raining). As we will see, this difference will play a crucial role in Faulkner's proposal and should be kept in mind for the remainder of the paper.
[23] See footnote 19.

> We are in the process of a messy divorce, and you were nothing but unreliable in the plans you made with me in the past, so I have no confidence you will show up today. This is not to suggest I believe that you will *not* show up; I might have no confidence you will show but if I believed that you will not show it would be wrong to describe my waiting as an act of trusting… [The husband] can decide to trust *S* to ϕ *just because* trust need not involve the belief that *S* will ϕ.[24]

The relevant point being made in this illustration is that in affectively trusting his wife to show up the husband needn't believe that she will actually show up to be motivated to wait for her.[25] But why so? Why can the husband be motivated to affectively trust his wife to show up even though he does not believe that she will eventually show? Mainly due to two factors: first, because in view of counter-evidence one can have a normative expectation on someone's behaviour without having a predictive expectation; second, because affective trust can be cognitively motivated solely by the presumption that the trusted will prove trustworthy.

Contrary to the expectation of predictive trust, which consists in the belief that *S* will ϕ, the husband's normative expectation needn't be grounded on sufficient evidence for the belief that the wife will show. Whereas counter-evidence against the belief that *S* will ϕ would make *A* withdraw a predictive expectation, the past unreliability of the wife is not (to a certain point) sufficient for making the husband withdraw the normative expectation he put on her. Simply, he can still think that the wife should show up because they have arranged a meeting even though he does not believe that she will actually show up. So having no confidence in his wife's future behaviour does not prevent him from putting a normative expectation on her, and so neither does it prevent him from affectively trusting her.[26] What is more, having no confidence does not prevent the husband from *presuming* that his wife will show up. And this is key to Faulkner's proposal.

According to him, the fact "that affective trust implies a presumption of trustworthiness is important, because presumption is not constrained in the same way as belief".[27] In particular, it is important because the presumption is not evidentially-constrained in the same fashion as belief is. Thus "A can continue, up to a point, to think well of S even in the absence of evidence, or even in the face of

---

[24] Ibid.

[25] The other point being that trust (both predictive and affective) requires not believing that *S* will not ϕ. Conversely, distrust does involve the belief that *S* will not ϕ –that is to say, a negative expectation towards *S*'s behaviour.

[26] The relation between *S*'s being unreliable and *A*'s affective trust on *S* is complex. I tried to capture this complexity with the qualification 'to a certain point' between brackets. I contend that unreliability can only make *A* withdraw her affective trust if it is sufficient to make her believe that S will *not* ϕ. As long as it permits just not believing that *S* will ϕ, *A* can still hold to her assumption that *S* should be normatively motivated to ϕ.

[27] Faulkner 2007, p. 884.

ISSN: 2653-3146

counter-evidence"[28]. This is why Faulkner suggests that making this presumption about *S*'s future ϕ-ing is like accepting that *S* will ϕ, and not like believing that *S* will ϕ –namely, because (contrary to believing) one can accept that proposition *p* obtains even in the face of counter-evidence. As Cohen explains, "a person who does not fully believe that *p* can nevertheless justifiably accept that *p*"[29] since to accept that *p* is just "to have or adopt a policy of deeming, positing, or postulating that *p* –that is, of going along with that proposition".[30]

Therefore, someone who accepts that *p* can happily ignore the evidence available about the likelihood of not-*p* and act on the acceptance that *p*, because going along with *p* simply means to act as if *p* were true. And this is what the husband of the illustration does: "in trusting his wife the husband brackets his belief that his wife will in all likelihood not show, and brackets those beliefs that give him reason for thinking this, and gives his wife the benefit of the doubt."[31] Thus, the husband can think that she should show up and at the same time presume or accept that she will, even though the evidence makes him not believe that she will in fact show up.[32]

Then, Faulkner concludes, if the husband can be cognitively motivated to wait for her and rely on her showing up to accomplish a certain intention of his thereby, no matter how uncertain or wary he might feel about it, is precisely because he can just hold to the presumption that she will eventually show. As Faulkner remarks, "this background of acceptance then specifies a way of thinking about trusting S to ϕ, which provides A with a motivation for so doing".[33] Now, someone may ask whether this would "really count as trust". The reason being that if the husband does not believe that she will not let him down, then he is not really trusting his wife. My response to this question is that the husband does trust her wife, although to merely say that he trusts her can be philosophically misleading.

Our trust talk presents two competing criteria for when to say that someone is trusting. On the one hand, for example, we say of someone that trusts when she believes that the trusted party will not let her down. That is to say, when *A* believes that *S* will ϕ, she has the confidence that *S* will ϕ. Is in this sense that we usually say that our friends are the ones in which we can trust. On the other hand, however, we also say of someone that trusts when she lacks certainty about whether the trusted party will let her down or not. When *A* does not believe that *S* will ϕ, for instance,

[28] Faulkner 2020, p. 338.
[29] Ibid., p. 369.
[30] Cohen 1989, p. 368.
[31] Faulkner 2014, p. 1986.
[32] Notice that the issue here addressed is deep down related to Michael Smith's moral problem. The moral problem evidences how having a normative reason to ϕ does not necessarily motivate one to ϕ. Thus, from the belief that '*S* should ϕ' one cannot derive the belief '*S* will ϕ' because *S* might not be even motivated to ϕ. One can justifiably believe the former and yet lack grounds to believe the latter. See Smith 1994.
[33] Ibid., p. 1987.

we say that all she can do is to trust that *S* will eventually ϕ. And this is actually something that can happen in contexts where *S*'s ϕ-ing is normatively constrained. As Faulkner puts it, in such contexts *A* might not believe that *S* will ϕ and yet believe that *S* should ϕ, which allows us to say of *A* that she trusts *S* to ϕ –in particular, that *A affectively trusts S* to ϕ. So my reply is that our trust talk sometimes does not allow us to see how trust can be grounded solely on normative expectations. If we only attend to one of the criteria we may overlook the other one and conclude that lack of confidence entails absence of trust, as if there was just one kind of trust.

4. Trust, competence, and intentionality

4.1 Two types of means-end beliefs

The notion of a means-end belief as stated in Smith's schema may be ambiguous. It strikes me that it can refer to two different types of beliefs that express different degrees of certainty about outcome. I will now show why Faulkner's proposal is only correct as long as it concerns means-end beliefs of the stronger type. And in the next subsection I will argue for the claim that the cognitive motivation for actions of affective trust in the kind of situation considered could be provided by a means-end beliefs of the weaker type.

On a stronger conception, the belief would be 'if *A* ψ-s, then *A* will get X'. When this belief is applied to the illustration of the husband, it could be rendered as, say, 'if I wait for my wife to show up, I will get the divorce papers signed'. And, on a weaker conception, the belief can be stated as 'if *A* ψ-s, then it is possible that *A* gets X', which is roughly rendered as 'if I wait for my wife to show up, then it is possible that I get the divorce papers signed'. The difference between these two types of beliefs should be understood thus: while the former represents the action of trust as ensuring the attainment of X, the latter represents it as merely making it possible. One who had this latter belief would therefore be more uncertain about the outcome of her action than someone who had the former.

Having established this distinction, it is easy to see how Faulkner's proposal does not apply to means-end beliefs of the weak type and, more importantly for my purposes, why it would not rule out the application of a weaker Humean analysis. Borrowing from the last section, the acquisition of the belief 'if I trust *S* to ϕ, then it is possible that I get X' imposes the following conditions:

(C1') That *A* believes 'it is possible that *S* ϕ-s'
(C2') That *A* believes 'if I trust *S* to ϕ, then it is possible that *S* ϕ-s'

And, as I see it, not believing that *S* will ϕ does not impede one from satisfying C1' and C2'. Basically, lacking the belief '*S* will ϕ' is compatible with believing that it is

possible for *S* to ϕ. So even in a situation where the husband lacks confidence he could perfectly believe that it is at least possible for his wife to show up, and so he could also believe that it is possible that his wife will show up if he waits for her. In turn, these beliefs would allow him to believe in the possibility of getting the divorce papers signed if he waits for her.

So, put simply, it seems plausible that even if the husband is so uncertain that he does not believe 'my wife will show up', he could still have a weak means-end belief of the sort 'if I wait for my wife, then it is possible that I get the divorce papers signed' just in virtue of satisfying C1′ and C2′.

4.2 Doxastic conditions for trust

If I am right, then, it seems that trustors can have this weak means-end belief even under circumstances of uncertainty as long as they can also believe in the aforementioned possibilities. But I want to contend something else, namely, that trustors in this context *need* in fact to have these three beliefs. Consequently, I want to defend that a weaker version of the Humean analysis, according to which having a motivating reason R_w to act is to have a desire and a weak means-end belief, could account for the cognitive motivation of actions of affective trust performed in these situations.

So to justify these claims I will need to explain why in affectively trusting *S* to ϕ, *A* needs to have the weak means-end belief in question, which in turn will require that I explain why *A* needs to satisfy C1′ and C2′. That the latter holds can be seen in the fact that trust involves reliance on others' competence, and that the former is the case can be seen in the fact that belief in the possibility of the intended is necessary for intending. Let's consider each fact in turn.

I proposed that when trust is practical, and not merely attitudinal, the trustor relies on the trusted party doing something. So here I want to follow Baier[34] in claiming that this reliance is reliance on another's competence. Since relying on *S*'s ϕ-ing is relying on her capabilities to behave in a particular way; and since 'being competent to ϕ' means being capable to behave in the relevant ways as to ϕ, I take Baier's approach to be right. So, for instance, when the husband waits for his wife to show up with the intention of having the divorce papers signed, he is relying on her capabilities to properly behave so as to show up. In order to affectively trust her, the husband needs to believe that she is capable of showing up (thus satisfying C1′) such that, since he trivially believes that he is waiting for her, he also needs to believe that if he waits for her it is at least possible that she shows up (thus satisfying C2′).

[34] Baier 1986, p. 259.

The first reason that I take to favour these requirements that follow from the considerations around competence and trust is that not satisfying C1' would imply that the husband is not trusting his wife, but distrusting her. In effect, if *A* were to believe that *S* cannot ϕ, *A* would consequently believe that *S* will not ϕ and so *A*'s attitude would be rather described as distrust. Instead of a positive expectation, *A* would have a negative expectation and her action could not qualify as an action of trust at all. And the second reason is that these requirements seeman to accommodate some doxastic implications concerning normative expectations as involved in actions of trust. In cases of affective trust, where the attitude consists in the belief that *S* should ϕ, it seems that *A* needs to satisfy C1' because of the intuitive principle that 'ought' implies 'can'. Believing that *S* should ϕ, that *S* is somehow obligated to ϕ, presupposes believing that it is possible for *S* to ϕ because a true normative demand seems to requires so.

I also proposed that the distinguishing feature of trust as an action is that it involves an intention-specific relation of reliance, as the trustor relies on the trusted party to accomplish an intention of hers. Since this intentionality is characteristic of actions of trust, I claim that they are doxastically constrained in a particular way because intending requires the satisfaction of a doxastic condition of some kind.[35] This condition, I suggest with philosophers like Moya or Davidson[36], is that for an agent to intend to get X by means of ψ-ing it is necessary (and sometimes sufficient) to believe that it is at least possible for her to get X by means of ψ-ing. Thus, for instance, when the husband waits for his wife to show up with the intention of having the divorce papers signed, he needs to believe that it is possible for him to get them signed if he waits for his wife. He would therefore need to believe 'if I wait for her to show up, it is possible that I get the divorce papers signed'.

If this conclusion that follows from the relationship between intentionality and trust seems plausible to me is because it sounds reasonable that failing to satisfy the doxastic condition suggested would be enough for the husband to decide not to affectively trust his wife. In effect, just like a rational agent would not intend to win the Christmas lottery if she deemed it literally impossible to win (say, because she knew it was rigged), it seems that the husband would not trust his wife with the intention of getting the papers signed if he considered it literally impossible to have them signed thereby (say, because he knew that she suffers from an extreme psychological condition that prevents her from meeting any normative demand whatsoever).

Moreover, the conclusion also seems plausible because it is able to account for the trustors' behaviour in cases where they have been let down. For instance, it can explain why someone in the husband's shoes would insist on meeting the wife

[35] This doxastic condition has taken several forms in the literature. See Mele 2010 for an overview.
[36] See Moya 1990 and Davidson 2002.

despite her past unreliability. Since he believes that it is possible for him to get to see her if he trusts her, even if he finds the chances quite low and does not believe that she will eventually turn up, he can still be motivated to engage in whatever compensatory behaviour is needed to accomplish his intention. He could, for example, be willing to rearrange a meeting and try his luck one more time. After all, as pointed earlier in brackets, on some occasions belief in the possibility of the intended can be sufficient to intend it, and on this occasion I would say that this belief is not only necessary but also sufficient for the husband to intend to have the papers signed by means of engaging in trusting activity with his wife.[37]

At this point, then, I can suggest the following weak Humean analysis of the husband's action of affective trust. Why did the husband wait for his wife to show up even though he did not believe that she would show up? Because he had a motivating reason of the sort R_w. That is, because he desired to have the divorce papers signed and believed something like 'if I wait for my wife to show up, then it is possible that I get them signed'. In view of his desire, this weak means-end belief could have cognitively motivated him to affectively trust his wife, or so I hope to have shown.

5. Concluding Remarks

In this paper, I have primarily done two things. First, I have developed the difference between trust as an attitude and trust as an action, and have identified in what exact sense reliance is related to trust. I have also provided a systematic account of actions of trust, emphasizing the role of intentions and attitudes in defining the nature of the action. And, second, I have presented and assessed Faulkner's position about the practical rationality of trust. I have followed him in that trust needn't involve the belief that S will ϕ, and offered the more positive proposal that it does need to involve the belief that it is possible that S ϕ-s. I concluded on this basis that, in order to perform the actions of trust he had in mind under circumstances of uncertainty, it is at least necessary that the trustor ascribes a weak instrumental value to her trusting.

While the husband's trust can be motivated even if he is uncertain as to whether his wife will eventually show up and prove herself trustworthy (even if he does not believe that), it cannot be cognitively motivated if he is uncertain about the

[37] Moya offers an enlightening illustration to explain this point. He explains that one can intend to hit a very far target by shooting an arrow even though one might not believe that one will hit it. Thus, he remarks, "in having an intention I commit myself either to make (if I think I will be able) or to try to make (if I do not think I will be able) its content true" (Moya 1990, p. 138). The point being that even a small prospect of success (which implies belief in the possibility of succeeding, and not in its probability) can, at least sometimes, be sufficient for intending. Now, it is worth noting that this is a contentious claim and some would say that the doxastic condition proposed is too weak if considered sufficient. For instance, see Grice 1971. For further discussion see Davidson 2002.

possibilities of her showing up and his getting the papers signed if he trusts her (if he does not believe that). Otherwise put, while he can bracket evidence against her future showing up or about the unlikelihood of his having the papers signed, he cannot likewise obviate evidence against these possibilities. The fact that actions of trust involve reliance on the trusted party's competence and the fact that this reliance is intention-specific seems to me to have this doxastic impact on trustors. Having these beliefs, I claim, is a necessary and sometimes sufficient condition for cognitively motivating actions of trust under the circumstances of uncertainty that Faulkner considers.

And it is important to underscore the 'sometimes' in this conclusion. For the question may be raised as to whether having this set of beliefs is a sufficient condition for trust in any context. For instance, someone could claim that one would not be motivated to entrust her daughter to a suspicious babysitter only in virtue of having these beliefs. It is certainly plausible that in a situation where one senses ill will on the part of the trusted party one would not be motivated to trust, such that perhaps believing that the trusted party holds goodwill towards us must be also a necessary condition for trusting in situations of this kind.[38] Now, however interesting this issue may be, I have to reply that I did not purport to provide a set of individually necessary and conjointly sufficient doxastic conditions for every context of trust. I just focused on one particular situation characterised by the uncertainty of the trustor in relation to the trusted party's behaviour, and so I deliberately set aside any further question about necessity and sufficiency.[39]

[38] This is in fact Baier's proposal, which I do not consider thoroughly promising. For, as I see it, it is not clear whether every action of trust requires assuming goodwill on the part of the trusted. For example, a very influential political prisoner can trust his captors will release him if he discloses some important information about his country's tactics. The point being that he can do this even if he assumes ill will on the part of his captors. See Baier 1986.

[39] This paper is based on a fragment of my Bachelor Thesis, which I defended at the University of València in June 2021. Consequently, there are many people to whom I am indebted. Special thanks to Sergi Rosell for supervising the thesis, and to Paul Faulkner and Tobies Grimaltos for commenting on a previous draft and discussing some of the points addressed here. I am also grateful to Marta Cabrera and Eric Olson for helpful discussion, and to the members of the thesis tribunal Chon Tejedor and Jordi Valor. Some parts of this paper were presented at the 3rd UPJA Virtual Conference for Undergraduate Philosophy, to whose audience I thank their valuable questions. This paper has also been improved with the suggestions of the anonymous referees.

References

Anscombe, G. E. M., (1979). 'Under a description'. *Noûs*, *13*(2), pp. 219–233.

Baier, A., (1986). 'Trust and antitrust'. *Ethics*, *96*, pp. 231–260.

Cohen, J., (1989). 'Acceptance and belief'. *Mind*, *98*(391), pp. 367–389.

Davidson, D., (1963). 'Actions, reasons, and causes'. *The journal of philosophy*, *60*(23), pp. 685–700.

Davidson, D., (1969). 'Individuation of events'. In *Essays in honor of Carl G. Hempel*, edited by N. Rescher, pp. 216–234. Dordrecht: D. Reidel.

Davidson, D., (2002). 'Intending'. In *Essays on actions and events.* Oxford: Claredon Press.

Faulkner, P., (2007). 'On telling and trusting'. *Mind*, *116*(464), pp. 875–902.

Faulkner, P., (2014). 'The practical rationality of trust'. *Synthese*, *191*, 1975–1989.

Faulkner, P., (2020). 'Trust and testimony'. In *The Routledge handbook of trust and philosophy*, edited by J. Simon, pp. 329–340. New York: Routledge.

Grice, H. P., (1971). 'Intention and uncertainty'. *Proceedings of the British Academy*, *57*, pp. 263–279.

Hawley, K., (2014). 'Trust, distrust and commitment'. *Nôus*, *48*(1), pp. 1–20.

Hollis, M. (1998). *Trust within reason*. Cambridge: Cambridge University Press.

Hume, D., (1739/1960). *A treatise of human nature*, edited by L. A. Selby-Bigge. Oxford: Claredon Press.

Mele, A., (2010). 'Intention'. In *A companion to the philosophy of action*, edited by T. O'Connor and C. Sandis, pp. 108–113. Oxford: Blackwell.

Moya, C., (1990). *The philosophy of action: an introduction.* Cambridge: Polity Press & Blackwell Inc.

Smith, M., (1987). 'The Humean theory of motivation'. *Mind*, *96*(381), pp. 36–61.

Smith, M., (1994). *The moral problem.* Oxford: Blackwell.

ISSN: 2653-3146

Undergraduate Philosophy Journal of Australasia www.upja.online
Vol. 3, December 2021

The Method of Wide Reflective Equilibrium and Self-evidence

Tiago Carneiro da Silva[1]

Federal University of Rio de Janeiro

Abstract

In this essay, my ultimate aim is to show that the method of wide reflective equilibrium (MWRE) can be improved in a way that allows us to detect self-evident propositions in a reasonably effective way. In order to do this, I first argue that appealing to self-evidence does not have to be considered a dogmatic approach in the search for moral justification. I do this while describing characteristics of self-evidence that are worth considering in devising a moral methodology. This allows us to see how the search for self-evident propositions may be compatible with the MWRE. I then defend that the method is not as radically opposed to the appeal to self-evidence as it has commonly assumed. When doing this, I argue that the MWRE is more effective in leading us to find self-evident beliefs than one might initially expect. Finally, based on some features self-evident beliefs have, I propose that, in addition to following the steps that the MWRE requires us to follow, we should meet two further requirements in order to detect self-evident propositions in a more effective way. Furthermore, the resulting methodological proposal, I argue, can be desirable even if there happens to be no self-evident propositions.

1. Introduction

Is there a correct, or best, method for us to acquire justified moral beliefs? If so, what method would that be? A very popular candidate is called "the method of wide reflective equilibrium" (henceforth, "the MWRE"), as it was named by John Rawls in his *The Independence of Moral Theory*.[2] This method essentially tells us to reflect on our beliefs about issues of a given area of enquiry while considering relevant theories

[1] *Tiago Carneiro has recently completed his undergraduate degree in philosophy at the Federal University of Rio de Janeiro, Brazil. He is currently looking to pursue a master's degree and carry out research on moral intuitions.*
[2] Rawls 1974.

ISSN: 2653-3146

and arguments in a way that leads us to form a coherent set of beliefs.[3] This end point of reflection, achieved after revising our beliefs in light of these relevant considerations, is what is called "wide reflective equilibrium".[4]

The method is so popular that some of the defenders of the method even hold that there is no other reasonable alternative to it.[5] They argue that one of the advantages of the method in relation to other possible alternatives is that it is not dogmatic. This is because the method takes every belief to be revisable. In this sense, its proponents generally hold that it is opposed to the idea that we can acquire justified moral beliefs by appeal to self-evidence apart from the application of the method. Appealing to the method commits one to the idea that there are unrevisable beliefs — and this, they claim, is a dogmatic approach to morality.[6] According to them, the only way we could appeal to self-evidence non-dogmatically is if we reach this conclusion *after* applying the MWRE, as it takes every belief to be revisable (even the belief that there are self-evident propositions), and those that accept this possibility believe that this is unlikely to happen.[7]

Rejection of self-evidence seems to come at a cost, though. First, to reject it is to abandon a firm ground for moral justification. It does not seem, for example, that one can be justified in having the belief that it is right to enslave black people even if this belief were part of a completely coherent set of beliefs or if it were acquired after reflecting on all of one's beliefs and alternatives to them. Self-evidence, on the other hand, seems to provide such a firm ground. Second, denying self-evidence requires one to explain why some specific beliefs seem to be justified independently of coherence or of what merely seem true to us. Taking an example from Cuneo and Shafer-Landau, it seems that one who adequately understands the sentence that says "for beings like us, in worlds like ours, the interests of others can be morally weightier than our own" and believes it on the basis of this understanding would be justified in believing it.[8] This belief does not have to be part of a coherent set of

[3] Here, I shall focus on methodological issues in the moral domain.
[4] The "wide" in "wide reflective equilibrium" indicates that a wide range of types of proposition is taken into consideration when one applies this version of the method.
[5] Smith 1994, pp. 40– 41; Scanlon 2003, p. 149; DePaul 2006, p. 618; and Walden 2013, p. 254. DePaul, for example, argues that we are not destined to have firmer grounds for "very much of what we believe " than the intuitions we have after having carefully taken all relevant views into consideration, which is what the MWRE tells us to do. Therefore, we cannot reasonably appeal to, say, self-evidence independently of the application of the method. Indeed, we should do our best to reflect rigorously about morality, but the most we can do is to follow the steps the method requires us to follow. In addition, other less rigorous alternatives would be too permissive.
[6] Daniels 1979, p. 264–267; Brink 1989, p. 8–9; Ebertz 1993, p. 213; Rawls 1999, p. 19.
[7] Rawls 1974; Daniels 2016.
[8] Cuneo & Shafer-Landau 2014.

ISSN: 2653-3146

beliefs, nor is it justified only because it seems true to us. Rather, this seems to be a special type of belief because its content is probably self-evident. [9]

Having these *prima facie* reasons to accept the existence of self-evident propositions in mind, we have at least some good reasons to attempt to devise a method that effectively detects them. However, although I do think that applying the MWRE is not the best way to detect them, I believe the method is a good starting point for us to devise a better procedure in that respect. To show why this is so, I will argue in favour of three main theses:

(i) The appeal to self-evidence is not intrinsically dogmatic. This, if correct, would show that dogmatism does not have to be a problem for those who hold that self-evident propositions exist in ethics. Additionally, this will show applying the MWRE and appealing to self-evidence are compatible in an important respect, as both approaches would be shown not to be dogmatic.

(ii) There are good reasons for preserving the steps the MWRE requires us to follow so that we can acquire self-evident moral beliefs and be self-evidentially justified, which is important for my third thesis.

(iii) The likelihood for us to acquire self-evident moral beliefs is increased if we, in addition to following the steps required for us to apply the MWRE, meet further requirements that are based on the conditions for being self-evidentially justified. Doing this, I will argue, is better than only applying the MWRE as it is traditionally conceived.

If these three theses are correct, the ultimate goal of this essay can be established: we can improve the MWRE as a method for acquiring justified beliefs (and moral knowledge). If there are self-evident propositions (as we have at least some reasons to think so) and we apply this resulting method, we will more effectively detect moral truths and gain moral knowledge. Increasing our chances of doing this is a

[9] There are other arguments against self-evidence other than the one that says that appealing to it is to commit oneself to some sort of dogmatism. One such argument is that of Brink, which assumes that being justified in believing a proposition *p* requires one to have a higher-order belief that expresses the proposition that *p* is a type of belief *T* and beliefs of type *T* are justified (Brink 1989, pp. 116-122). If this is so, it would not be possible to be non-inferentially justified, as we would be if we were self-evidentially justified. However, as Shafer-Landau later persuasively replied, this assumption is problematic for at least two reasons (Shafer-Landau 2003, pp. 252-253). First, there does seem to be beliefs that do not require higher order beliefs to be justified, such as simple perceptual beliefs. Second, it threatens to start a vicious infinite regress, as being justified in having a second-order belief would require one to have a third-order belief, and so on. Since we do not have an infinite number of meta-beliefs of the required kind, none of our beliefs would be justified, including the belief that Brink's assumption is correct.

result we can expect from a better method, since increasing our chances of acquiring moral knowledge is desirable for a method of moral justification. However, even if there happens to be no self-evident moral propositions, the proposal I offer for enhancing the MWRE based on features of self-evidence may, I will argue, still be worth accepting.

2. Self-evidence and Dogmatism

Before dealing with the compatibility between the MWRE and self-evidence, it is useful to show that appealing to self-evidence does not have the disadvantage that proponents of the method generally hold that such appeal has — namely, that of being dogmatic. This will pave the way for the discussion of how the MWRE can be applied in order to lead us to be self-evidentially justified. This shall be clear soon enough.

I take Robert Audi's account of self-evidence to provide a very plausible epistemology.[10] It captures the essence that is thought to be present in classical examples of propositions that are candidates for self-evidence — namely, that they are true propositions which can be known *a priori* and non-inferentially (a knowledge that, as Ross says, requires a "certain mental maturity" to be achieved). Also, it provides a detailed explanation of the role that reason has in apprehending those truths: that of leading us to understand them adequately.[11] This way, he proposes that self-evident propositions would be those true propositions that (i) if adequately understood by *S*, *S* is justified in believing them on the basis of that understanding, and (ii) if believed in such a way, S knows them. Based on this, three common misconceptions about self-evidence may be dispelled. First, notice that these features of self-evidence do not entail that *S will believe* a self-evident proposition *p* once *S* adequately understands it. If *S* understands *p*, even if adequately, this does not mean that *S* will immediately see that *p* is true.[12] Second, adequately understanding *p* does not mean that *S* will necessarily take it to be self-evident (*S* may not even know what self-evidence is or may even deny that self-evident propositions exist). Third, this account of self-evidence does not entail that there is no way of inferring *p* from premises.

Contrary to classical intuitionist philosophers such as H.A. Prichard, W. D. Ross and G.E. Moore, this account of self-evidence allows us to defend that self-evident

[10] Audi 2015, pp.65-66; Audi 2018, pp. 1-3.
[11] This explains why disbelieving a strong candidate for self-evidence may very well be a product of inadequate understanding. Take, for example, the oft-cited proposition that all bachelors are unmarried. If someone denies it, we have strong reasons to think that the person in question either does not understand at least one of the concepts in the proposition or does not understand the conceptual relation between them.
[12] This may happen, but not necessarily so.

propositions "need not be unprovable, need not be obvious, and need not be rationally beyond dispute".[13] This provides a reply to criticisms against classical intuitionists that charge them for being dogmatic due to their defence of self-evidence. Consider Audi's example that 'if *p* entails *q*, and *q* entails *r*, and *r* entails *s*, and *s* is false, then *p* is false'. This may be obvious to some people (maybe to those who have experience in dealing with logic), but it is certainly not obvious to

most people. To see that such a proposition is true might require one to draw an inference from '*p* entails *q*, and *q* entails *r*, and *r* entails *s*' to the conclusion that '*p* entails *s*', which may lead one to see more clearly that if *s* is false, then *p* is false.[14] Nevertheless, although this inference might be a way to adequately understand it, it is *this understanding* that grounds the self-evidential justification. Also, these steps might be a way to show to another person who does not adequately understand it that the long proposition is true. However, if one is able to show these inferential steps to another person, one may still justifiably believe the long proposition solely on the basis of adequately understanding it. [15]

In fact, self-evident propositions might even sound counterintuitive. Another example of Audi's is that "a child can be born by its grandmother". At first sight, at least, it might seem to *S* that such a proposition is not true. It may take *S* to think of a specific case in which a mother has a child with her son for *S* to understand it appropriately and see its truth. One who knows the story of Oedepus Rex, who had children with his mother unknowingly, may see this more easily than a person who does not. In any case, it is possible that, when one is not thinking about such cases, one does not find that proposition intuitive even after accepting its truth.

Given these considerations, some possible explanations for why people may deny self-evident propositions may be offered. People tend not to accept counterintuitive propositions. This way, people may be less prone to accept self-evident propositions that sound counterintuitive at first. Self-evident propositions that involve complex relations among the concepts that figure in them do not tend to be obvious and may require imagination or application of inferences for one to be able to understand them adequately. Moral propositions are arguably more likely to be harder to be adequately apprehended, as they involve more complex concepts and relations

[13] Audi 2015, p. 66.
[14] Notice that this is an *internal* inference; that is, this inference has a premise that is supplied by the very proposition that is the conclusion of the inference (the long self-evident proposition). That is why this inference is only one way of adequately understanding the long proposition, rather than a conclusion of an argument that has as premises propositions other than those that are supplied by it. In any case, it is possible to infer self-evident propositions from "external premises". However, one is not required to draw *this kind* of inference in order to be justified in having self-evident beliefs.
[15] For one to adequately understand it, one needs to understand it to a degree that is possible to see that it is true, and being able to explain a self-evident proposition is evidence that that person has this degree of understanding. Hence, it is plausible to defend that adequately understanding a proposition involves being able to explicate one's understanding of it.

among them than the examples I gave.[16] Self-evident propositions are truths that can be apprehended *a priori*, but that does not mean that they have to be analytical.[17] They are propositions we can see to be true by reflection on the relations among their constituents alone, even though we may need sense experience to learn the concepts involved.[18] What we do before we "see" that a self-evident proposition is true is to correctly apprehend the relations among the concepts that are part of it. If these relations are complex, as they seem to be in moral propositions, the harder it is to understand them adequately — and the less likely it is for them to be obvious.[19] A candidate for self-evidence which might be an example of a complex moral truth is the proposition expressed by Sidgwick's principle of justice:

> It cannot be right for *A* to treat *B* in a manner in which it would be wrong for *B* to treat *A*, merely on the ground that they are two different individuals, and without there being any difference between the natures or circumstances of the two which can be stated as a reasonable ground for difference of treatment.[20]

It may be hard to understand many propositions adequately, as it involves more than mere semantic comprehension. A bilingual person may be able to translate the sentence 'a child can be born by its grandmother' correctly after comprehending it

---

[16] The examples I gave might seem to indicate that self-evident propositions are only analytical or logical truths. However, they are only straightforward ways to demonstrate how self-evident propositions may be provable or counterintuitive. If these kinds of self-evident propositions may be so, there is good reason to think that self-evident moral propositions may also be (that is, if there really are self-evident propositions — be they moral or non-moral).

[17] One example of a proposition that is probably self-evident and not analytical is the proposition that nothing can be red and green all over at the same time. "Not green" is not plausibly part of the *meaning* of "red". Therefore, this proposition would not be true in virtue of its meaning. The same would be true for at least most moral propositions.

[18] It is important to note that the dependence on experience is limited to the acquisition of concepts. The truth of the proposition that faulty electrical plug sockets tend to cause fires, for example, can only be justifiably believed empirically. If this proposition is what is meant by the sentence "faulty electrical plug sockets can cause fires", then this sentence is not self-evident. If, on the other hand, what is meant by it is that it is conceptually possible that faulty electrical plug sockets can cause fires, then this might very well be self-evident. However, this does not seem to be what we mean when we say such a sentence, as when we say it we have the intention of warning against having faulty electrical plug sockets, as they are *likely* to cause fires. The example that a child can be born by its grandmother, on the other hand, does not express tendencies that can only be empirically established; it is only a proposition that contradicts something that seems impossible (that is why it is probably a self-evident proposition which is more interesting than the one that expresses that it is conceptually possible that faulty electrical plug sockets can cause fires).I thank an anonymous referee for pressing me on this point.

[19] Hence, a method that allows us to identify self-evident propositions accurately needs to take into account how adequately we understand them. This is what my proposal in section 4 does.

[20] Sidgwick 1907, p. 380. Notice that this is a *candidate* for self-evidence, and it may even be a false proposition. However, given that this seems to be another way of stating that moral properties supervene non-moral ones, which has been defended by philosophers of several different backgrounds, this seems to be a good candidate. If it is indeed self-evident, denial of this proposition might be explained by

semantically. But for this person to understand it adequately, more reflection will be necessary. In fact, there may be many "comprehensional variables", which may account for at least part of the denial of self-evident propositions.[21] These "variables" are the factors that determine how well a person understands a concept or proposition. One such variable concerns the ability to translate a sentence from one language to another. Another concerns the ability to identify instances of propositions, such as identifying that the wrongness of John killing Mary for fun instantiates the proposition that torturing people for fun is wrong.[22]

Having these factors that may lead us to deny self-evident propositions in mind, it should be clear by now that we are not infallible in detecting self-evident beliefs. Moreover, these considerations are only part of the story: not only can we deny self-evident propositions, but we can also accept as self-evident what is *not* self-evident. We may believe that a proposition is self-evident due to cultural, or even genetic, influences.[23] As self-evident propositions are objectively true, neither our culture nor our genes determine what propositions are self-evident.[24] Apart from this, cognitive biases also affect our moral judgments and are possibly a cause for our acceptance (or rejection) of some moral propositions which contradict self-evident propositions.[25] Thus, it seems to be desirable to find ways of diminishing these influences in our moral thinking.

It is important to note that, although we may not believe a self-evident proposition *p* even after appropriately understanding it, understanding *p* adequately does give a rational person a disposition to believe it. Thus, we should expect rational people to have such disposition after we conclude that they have appropriately understood *p*. This does not mean, though, that people who adequately understand *p* will necessarily be thinking irrationally if they deny *p*. One can rationally be committed to moral scepticism or to a theory that denies a self-evident proposition. Such a scepticism or theory may provide *some* justification for denying it. To the extent that

[21] For a more detailed account of what "adequate understanding" amounts to, see Audi's 2018 article.
[22] I will later mention some other comprehensional variables which might be useful to take into consideration when trying to detect self-evident beliefs in a more effective way.
[23] There are arguments that appeal to evolution to criticise moral realism, which indirectly attack the classical appeal to self-evidence (see, for example, Street 2006 and Joyce 2006). It is possible, though, to limit the range of the target of that kind of argument to only a certain type of moral belief. One example of how this can be done can be seen in Lazari-Radek & Singer 2014, pp. 174-199. In particular, they target only those intuitions that merely increased the likelihood of reproductive success, which, they argue, are intuitions that are not impartial (egoistic intuitions and those that place more importance on our kin or group for their own sake). Analogously, debunking arguments of such a limited range may be based on considerations regarding cultural differences.
[24] This way, self-evident propositions cannot clash or contradict each other. Hence the importance of appealing to a method that tell us to aim for coherence, as does the MWRE.
[25] See, for example, studies which show evidence of bias in moral thinking, such as Tversky & Kahneman 1981, p. 453; Petrinovich & O'Neill 1996; and Haidt & Baron 1996. Although the first of these studies is not specifically about bias in moral thinking, it does show evidence of that type of bias.

they provide such a degree of justification, those who deny it based on such reasons are not irrationally denying it.

Accepting that one can rationally deny self-evident beliefs and that one is not infallible in detecting self-evident propositions is a stance that can hardly be considered dogmatic. If one accepts that there are self-evident propositions, one can hold such a view and does not have to take oneself to be infallible in detecting them. Moreover, having the factors that may lead one to fail to detect self-evident moral propositions in mind will help us to try to get around the problems we might face in the search for self-evidence. In what follows, I will explore how the well-known MWRE may help us to detect self-evident propositions and how its basic structure can lead us to devise a better method.

3. The Method of Wide Reflective Equilibrium and Self-evidence

The MWRE tends to be presented as a method that conflicts with the idea that there are self-evident propositions. Some say that, after applying the method, what does (or most likely does) the justificatory work in the final system of beliefs is coherence. [26] Other philosophers say that only other forms of "weaker" foundations have justificatory power in it.[27] Both these positions oppose the idea that applying the method would lead us to find strong non-inferential foundations for moral justification or knowledge, such as self-evident propositions. [28]

In contrast to these views, I believe that the MWRE can not only be compatible with self-evidence (which would mean that coherence is not the only factor that does the justificatory work) but also be more effective in detecting self-evident propositions than one might expect. To argue in favour of this, I will first present a general characterisation of the method, as it is to be applied in moral philosophy, and then attempt to show how it could lead us to be self-evidentially justified.[29]

---

[26] See, for example, Brink 1989, Rawls 1974 and Daniels 2016.

[27] For example, although Ebertz and McMahan defend that what does the justificatory work in systems of moral beliefs is not coherence, they do think that the method should be understood as a "modest foundationalist model" of justification which does not appeal to self-evidence (Ebertz 1993; McMahan 2000).

[28] Recall what I said in the introduction regarding Rawls and Daniels's positions on the improbability of us acquiring beliefs that serve as self-evident foundations for all other beliefs of the relevant domain. This would be improbable because when we apply the method, we reflect on our beliefs in light of all other relevant beliefs. We also revise them in light of them. This, they argue, increases the likelihood that we will only acquire inferential beliefs.

[29] The method can be applied in many domains, such as epistemology, logic and normativity in general. Indeed, one of the first characterizations of the method was described by Goodman, who wanted to apply it in the domain of logic (Goodman 1953). After Goodman, Rawls offered an influential characterization of the method, defending that we should employ it in order to theorize about justice (Rawls 1971). Daniels, then, proposed an even more detailed description of the method and is still recognized for having offered the most complete account of the method (Daniels 1979).

The MWRE can be divided into four or five stages, depending on what account of the method we are talking about.[30] Below, I describe it as having five stages:

(i) We first identify our initial moral beliefs – the ones we have when starting to apply the method. These beliefs may be either particular or general. That is, they may be moral beliefs about particular cases (e.g., that it was wrong for John to lie) or very general abstract beliefs (e.g., that we should aim at the good on the whole).

(ii) Then we select from that set of initial moral beliefs some beliefs which are simultaneously (a) those in which we are confident; (b) those that are not likely to be influenced by egoism or strong emotions (e.g., desires regarding only one's own welfare, fear or anger); (c) those that are likely to be stable in the face of scrutiny; and, at least if we follow Rawls's remarks about considered judgements in a 1951 paper,[31] (d) those that have not been derived consciously from the application of principles. Those beliefs are called "considered moral judgements".

(iii) After having selected the considered moral judgements, we try to find principles that account for, or systematize, those judgements. These principles should be ones which would lead the person who is applying the method to accept, by only using these principles as premises in conjunction with propositions that describe relevant situations, the same considered moral judgements previously selected. [32]

(iv) Instead of just making adjustments in our system of beliefs in order to match principles with considered judgements (which, if successful, would lead to a coherent system of beliefs called "narrow reflective equilibrium"), we also take into account moral and non-moral theories (as well as arguments in favour and against them) that might be relevant to the moral domain.

(v) By reflecting on everything that has been taken into consideration so far, we revise our initial considered judgements, principles, and moral and morally

[30] The second of the steps described is omitted by some philosophers. There are at least two: Goodman 1953 and Lewis 1983.
[31] Rawls 1951.
[32] For example, for 'it is wrong to eat meat' (call it 'P') to be a principle of this kind, P in conjunction with sentences describing relevant facts such as 'Peter ate meat' should lead to a conclusion that is itself a considered judgement, such as 'it was wrong for Peter to eat meat'. Notice that this is a condition for a statement to be considered the type of principle that *figures in this step of the method*. Remember that it is possible for a considered judgement to be just as general as, or even more general than, the principle that I presented as an example.

> relevant non-moral theories in order to achieve a final coherent system of beliefs. This endpoint of reflection is called "wide reflective equilibrium".

The MWRE is the version of the method that is considered to lead us to have justified moral beliefs. On the other hand, the method of *narrow* reflective equilibrium (*i.e.*, the one which does not tell us to take into consideration things such as moral theories, arguments in the literature, etc.) is not taken to be relevant to the debate of moral justification[33]. Indeed, this simpler method just tells us to systematise beliefs we had before applying it — and this does not seem to be a very promising way of acquiring systems of beliefs that are systematically justified. This is why in this essay I am focusing on the method of wide reflective equilibrium.

To point out that the method *itself* does not deny that there are self-evident propositions, it is important to notice that the description of the method does not tell us what makes beliefs justified. Therefore, it does not tell us whether there are propositions such that one would be justified in believing them solely on the basis of adequately understanding them. However, I do not only think that the method is neutral as to the existence of self-evident propositions; I also believe that we can acquire *self-evidential* justification when applying the method — that is, that we can acquire self-evident beliefs on the basis of adequately understanding them. Nevertheless, one could counter this view with the following argument:

> *P1* – We are only self-evidentially justified if we believe a self-evident proposition based on our appropriate understanding of it.
> *P2* - The method of reflective equilibrium only leads us to have beliefs in a purely inferential way — and not based on our understanding of them.
> *C* – The method does not lead us to be self-evidentially justified. [34]

I do not think that *P1* is problematic.[35] The problem with the argument, though, is

[33] Rawls 2001; Lazari-Radek & Singer 2014; Daniels 2016. Kamm defends that a version of the method of narrow reflective equilibrium reveals some underlying psychological structure of our moral thinking (Kamm 1993, p. 8). It might seem at first that Kamm is an exception to this trend. However, as Kamm herself recognizes, attempting to reveal this structure is a *descriptive* task; not a *justificatory* one.

[34] Some clarifications might be appropriate. First, notice that it allows that it is possible to be self-evidentially justified *simpliciter*. It allows that one could be self-evidentially justified by other means. For example, one could believe a self-evident proposition on the basis of adequately understanding it when one is applying the method of narrow reflective equilibrium. Second, it concedes that one can acquire self-evident beliefs when applying the method. That is because one can have a self-evident belief p even if one does not believe p based on one's understanding of p. For example, one can believe it solely on the basis of argumentation from other premises present in the system of belief. Being self-evidentially justified is different from having self-evident beliefs.

[35] Indeed, it is part of the very definition of self-evidential justification that one is self-evidentially justified only if one believes a self-evident proposition based on an adequate understanding of it.

P2. Although the MWRE is a dialectical procedure, this does not mean that every belief in reflective equilibrium is acquired in a purely inferential way. First, as I mentioned before, one may draw inferences in order to *adequately understand* a proposition, and this does not mean that one accepts it on a purely inferential basis. It is the *understanding* that grounds justification; not the inference that helped one to understand the proposition in question. Second, it is possible for a belief to be a *product* of an inference without it being an *inferential* belief, even if such an inference was not necessary for an adequate understanding of it. Consider the case in which I do not believe the self-evident proposition *p*, even after obtaining an adequate understanding of it, because I believe the following inference:[36]

P1 - If *r*, then ¬*p*
P2 – *r*
C – ¬*p*

This way, I do not believe *p* even though *p* is self-evident. However, my reason for believing ¬*p* may be undermined if afterwards I believe ¬*r* based on an inference such as:

P1 – If *w*, then ¬*r*
P2 – *w*
C - ¬*r*

As *p* is self-evident, I tend to believe *p* based on my adequate understanding of it. Furthermore, I do not see a reason to doubt it anymore, as I no longer believe *r*.[37] I then come to believe *p* without *p* being a conclusion of an argument. All that has happened is that I do not believe that there is a defeater for *p* anymore, as the last inference made me change my mind about the truth of *r*. Therefore, my belief that *p* was a product of an inference; however, such a belief is not inferential. This kind of reasoning may occur quite regularly when one applies the MWRE. Especially when we consider philosophical arguments, such as the ones we may encounter when applying the method, we may come to believe certain sceptical theses which lead us to deny propositions that are self-evident. However, these sceptical theses may be undermined as well, which increases the likelihood that we will believe the self-evident propositions these theses denied.

[36] Recall that adequately understanding a proposition does not mean that one *will* believe it. There is only a *tendency* to believe it, which may be overridden by some other opposing belief that happens to be more appealing to the person in question.
[37] Recall what I said about self-evidential justification: if one adequately understands a self-evident proposition, one tends to believe it.

Given what has been argued above, I believe we can conclude that the MWRE not only is compatible with the search for self-evident propositions, but also compatible with self-evidential justification. It may be argued, though, that it is unreasonable to endorse the method in order to acquire self-evidential justification, as the MWRE is not effective in leading us to detect self-evidence, nor in leading us to adequately understand self-evident propositions.

Regarding this, I do concede that it is not the best method for acquiring self-evident beliefs and for being self-evidentially justified.[38] Nevertheless, I do think that it is more effective in leading us to find self-evident propositions than one might expect, and we could make use of the features of the method in order to detect them effectively. The first reason for this is that it is plausible that at least some of our considered judgements are self-evident or have as content propositions that are similar to some self-evident propositions.[39] Thus, as the MWRE tells us to aim for coherence, applying the MWRE will probably lead us to filter out the judgements that most obviously do not cohere with self-evident propositions.[40] If so, this could at least diminish the probability that we will hold too many beliefs that more clearly deny self-evident propositions, such as beliefs that state that torturing for fun is morally good.[41]

Another reason to think that the MWRE's ability to lead us to find self-evident propositions is not so low is related to the mistaken assumption that they must be very general. This assumption seems to be based on the idea that self-evident propositions are unprovable. However, self-evident beliefs *can* be provable. This means that there may be general propositions (some of which may also be self-evident) that can serve as premises to prove them. Consider some examples offered by Terence Cuneo and Russ Shafer-Landau that might be good candidates for self-evidence:[42]

a) For beings like us, in worlds like ours, it is *pro tanto* wrong to engage in the recreational slaughter of a fellow person.
b) For beings like us, in worlds like ours, it is *pro tanto* wrong to humiliate others simply for pleasure.

[38] That is exactly why I am proposing that we add further requirements than the ones the MWRE tells us to meet.
[39] Recall the characteristics of considered judgements from page 10. As we tend to believe self-evident propositions after having adequately understood them, they may very well be selected as considered judgements.
[40] Rawls denies that considered judgements are self-evident, and I agree that, as such they are not (Rawls 1999, p. 507). That does not mean, though, that some of them may be.
[41] To be sure, even if this is how it happens to be, it would certainly still be possible for one to hold unreasonable beliefs.
[42] Cuneo & Shafer-Landau 2014, p.7.

c) For beings like us, in worlds like ours, it is *pro tanto* wrong to torture others just because they have inconvenienced you.

Part of ethical theorizing might be to find out how propositions such as these can be proved. This, however, allows that one can be justified in believing them on the basis of adequately understanding them (in case they are indeed self-evident). Maybe one of the premises that may be used to prove them can be something like Sidgwick's proposal that 'the good of any one person is no more important from the point of view (if I may put it like this) of the universe than the good of any other' (call it "the principle of benevolence").[43] If this is the case, it can be said that there are more self-evident propositions than several philosophers believe there are, as they think that the view that defends self-evidence must assume that self-evident propositions are only unprovable axiomatic propositions.[44] As there would be more self-evident propositions to be discovered, it is more likely that we will discover them by employing the MWRE than it would be if there were only axiomatic self-evident propositions.

To be sure, as it is likely that there are more self-evident propositions than it was assumed in the past, this makes virtually *every* method more likely to detect them. What reasons do we have to pick the MWRE instead of any other method to try to detect them? One reason for choosing the method is that it tells us to consider many moral views that have been accepted by many reflective people. As I mentioned before, these reflective people probably thought carefully about the propositions they accepted, and thus were more likely to adequately understand them and, consequently, find them intuitive and defend them. A second reason is that it tells us to consider many particular intuitions, which may help us to adequately understand self-evident propositions that we may encounter, and maybe even serve as evidence in favour or against certain self-evident principles.[45] Third, it tells us to consider arguments which may count in favour or against the reliability of certain intuitions. Fourth, although it not merely tells us to form coherent systems of beliefs, the fact

[43] To be sure, maybe even none of them is self-evident — these are only some plausible examples that I am using to show how there may be less general self-evident propositions that can be proved by more general ones. In any case, both this proposition and the less general ones seem to be propositions that one would be justified in believing solely on the basis of adequately understanding them. To be sure, understanding them in such a way does not mean that one *will* believe them; however, *if* one believes them on such a basis, it is plausible to say that one will be justified. That is why they are candidates for self-evidence.

[44] Influential intuitionists who have explicitly defended such a view are Moore, H.A. Pritchard and Ross (Moore 1903, p. x; H.A. Prichard 1912, p. 29-30; and Ross 1930, p. 29). Rawls and Daniels seem to have defended it as well, and this may be a reason why they thought it was so unlikely that one would find beliefs that could be self-evident when applying the MWRE (Rawls 1974, p. 8; and Daniels 2016). Brink, a defender of the MWRE, holds more explicitly that self-evident propositions are unprovable (Brink 1989 p.112-113).

[45] For example, if these more particular intuitions tend to be unconscious applications of self-evident principles.

that it tells us to pursue coherence is an important feature which may make it more likely for us to find self-evident propositions.[46]

For these reasons, I do not think that we should abandon the general steps of the procedure. However, for us to increase its effectiveness in leading us to find self-evident beliefs, I do think that there ought to be additional steps. In the next section, I will make a general outline of what kind of steps we can add.

4. A Better Method for Detecting Self-evident Beliefs

The procedure I will propose in this section preserves all the steps one has to take in order to apply the MWRE. However, I believe that a better method should have more requirements than the MWRE (as it is traditionally defended) imposes. One of them is related to adequately understanding candidates for self-evidence. After having considered and reflected on many moral judgements when one applies the MWRE, one is likely to have found reasonable candidates for self-evidence and may be even self-evidentially justified regarding some propositions. If some of these candidates are self-evident, and one adequately understands them, one will tend to believe them.[47] This gives us reason to try to identify the propositions we think we and others adequately understand and tend to believe them on the basis of that understanding. These will be candidates for self-evidence. Once these candidates for self-evidence have been identified, one could apply an additional procedure that filters out candidates that do not satisfy the characteristics we expect self-evident propositions to have. The procedure I have in mind is based on the following reasoning:

> *P1* – If *S* adequately understands proposition *p* and *p* is self-evident, *S* tends to believe it on the basis of adequately understanding it.

Now, imagine we empirically confirm the following:

> *P2* - *S* adequately understands *p* and does *not* tend to believe it on the basis of adequately understanding it.

In that case, we could conclude, then, that *p* is not self-evident. This reasoning allows us to exclude some candidates for self-evidence, and based on this I propose the first

[46] If we tend to believe self-evident propositions on the basis of understanding them, and if we try to make our system of beliefs more coherent, coherence may lead us to find other self-evident beliefs that are coherent with the ones we hold.
[47] And it will be more likely for us to be self-evidentially justified.

additional methodological requirement to be followed after following the steps the MWRE tells us to follow:

Additional methodological requirement 1- One should either empirically test or look for studies that empirically test whether people adequately understand candidates for self-evidence and believe them on the basis of that understanding.[48] The test may yield different results and we should adopt different positions accordingly. It may be found that:

(i) Subjects who adequately understand a candidate for self-evidence satisfy either (a) or (b) below:

(a) They do *not* tend to believe the candidate for self-evidence on the basis of adequately understanding it. In that case, we should not consider it self-evident (or at least consider it a bad candidate, as the empirical experiment may not be perfect).
(b) They tend to believe it on the basis of adequately understanding it. In that case, other things being equal, the proposition in question should remain a candidate for self-evidence and our confidence that it is self-evident may be increased for having passed the test.

It may also be found that:

(i) Subjects who do not adequately understand a candidate for self-evidence satisfy either (a) or (b) below:

(a) They do *not* tend to believe a candidate for self-evidence on the basis of adequately understanding it. In that case, we have a reason to think why they deny it even though it may be self-evident (*i.e.*, they may deny it because they do not adequately understand it).
(b) They tend to believe it. In that case, we have reason to think that it may be a proposition that is not self-evident even though people do tend to believe it.

It could be tested, based on this requirement, for example, whether people who adequately understand Sidgwick's principle of benevolence tend to believe it on the

[48] What I propose is essentially an application of experimental philosophy that helps us to discard candidates for self-evidence that we would not otherwise discard. It might also provide reasons for accepting candidates for self-evidence that have passed this test. This way, my position fits within both the so-called "negative" and the "positive" programmes, as this type of empirical studies may have as effect the rejection of some candidates for self-evidence as well as the vindication of other candidates.

basis of adequately understanding it.[49] If they do not, this indicates that it is probably not self-evident.[50] If they do, other things being equal, we may still take it to be a candidate for self-evidence. If no one, or if no significant number of people, adequately understand it, more studies designed to make subjects better understand the principle of benevolence and/or more studies with a higher number of subjects should be conducted.[51]

Notice that we cannot ascertain without a doubt that such a candidate for self-evidence is *indeed* self-evident. It is possible that it is not true (which would imply that it is *not* self-evident) even if it passes the test. However, if the test has worked and a proposition has not passed the test, it can be safely concluded that the candidate for self-evidence is *not* self-evident.[52] Discarding candidates for self-evidence this way makes it more likely that we will hold self-evident beliefs, as there will be less wrong candidates for us to accidentally hold.

Admittedly, this is a very complex procedure. And this is not even the whole story: there are many factors we must consider in order to check whether someone adequately understands a candidate for self-evidence. Audi, for example, argues that for one to understand a proposition adequately, one must satisfy *nine* "comprehensional variables". Recall that adequately understanding a proposition does not merely involve semantic understanding. It also involves, for instance, being able to respond adequately to questions related to what follows or does not follow from the proposition in question, being able to explicate one's understanding of it, being able to identify instances of what the proposition states *etc.* All these factors, even though they are many, should be taken into account if we want to identify self-evident beliefs in a more effective way.

The fact that it is too complex may lead one to reject such a procedure. It might be argued, for instance, that people are justified in having moral beliefs without having to apply such a procedure. This, the objection goes, is sufficient for us to just dismiss the procedure. I think this would be a bad objection, though. The point is not that people will *only* be justified if they follow this procedure.[53] Indeed, if we accept that there are self-evident propositions, one *is* justified in believing them solely on the basis of adequately understanding them. There are even other ways one can be

[49] Recall that this principle states that the good of any one person is no more important "from the point of view of the universe" than the good of any other.
[50] That is, if it were self-evident, people *would* tend to believe it.
[51] The relevant result is seen only when subjects adequately understand the candidates for self-evidence.
[52] If a candidate for self-evidence passes this test, though, this may be a further reason for accepting it as self-evident, because it has not been discarded as self-evident *and* because it is still a candidate for self-evidence on reflection.
[53] Here, I am trying a different approach than philosophers such as Rawls, Brink and Daniels seem to have tried, in that the method I am proposing is not necessary for justification, whereas they seemed to defend that the MWRE *is* necessary.

justified in having a belief. However, if we can diminish the likelihood that we will take certain propositions to be self-evident when they are not or diminish the likelihood that we will *not* take self-evident propositions to be self-evident, I believe it is reasonable to do so.

Science, for example, has developed very complex methods to find out things about the world. Yet, we should not dismiss them because they are complex. Accepting the methods of science also does not commit us to not accepting the conclusions we arrive via less complex methods. I am justified in believing there is a computer in front of me even though such a belief is grounded merely on my seeing it. I do not have to wait for science to tell me whether I really am seeing it in order to be *prima facie* justified.[54] Like science, though, my proposal may lead us to increase the degree of justification of our systems of moral beliefs.

A further question that may arise is the following: why use an empirical procedure to find out whether people tend to believe a proposition instead of trying to discover our tendencies to believe it *a priori*? The answer is that the statistical apparatus that can be used in empirical studies are more powerful than our introspection when it comes to detecting patterns. The former would be more sensitive to detect the relevant factors that lead us to believe certain propositions than the latter. By appealing to the type of experimental data I am defending we should pursue, we could detect tendencies of belief or disbelief that are influenced by factors other than an appropriate understanding of propositions and select candidates for self-evidence more reliably.

This last point leads us to another consideration: cognitive biases may affect people's judgement and, consequently, lead them to either tend to believe propositions that are *not* self-evident non-inferentially or tend *not* to believe propositions that *are* self-evident even when one understands them adequately. This way, the studies we should appeal to in order to satisfy this additional requirement I am proposing must control for cognitive biases. They should consider whether, for example, the instances of self-evident principles subjects identify when they are being tested for adequacy of understanding are triggering biases. Controlling for, say, order effects would be desirable in such studies.[55]

Controlling for biases such as these in types of studies that are different from the ones I have mentioned may also be useful for the search of self-evident propositions. A subject may deny a self-evident proposition because he or she thinks that an

[54] Science may increase my degree of justification by confirming, say, that I am not hallucinating, but my being justified is not determined only by appealing to science.
[55] "Order effects" occur when the mere difference in the order of sentences subjects consider lead to differences in the acceptance of such propositions.

intuition he believes speaks against it, and this intuition may be biased. This allows us to propose another additional requirement.

Additional requirement 2: One should either test or look for studies that test whether people are biased when they have a particular intuition that contradicts a particular candidate for self-evidence. The test may yield different results and we should adopt different positions accordingly:

(i) If subjects are biased, our confidence that the candidate for self-evidence is self-evident should be maintained.

(ii) If subjects are not biased, our confidence that the candidate for self-evidence is likely to be self-evident should, other things being equal, diminish.

Consider Sidgwick's principle of benevolence again, which may be denied because of, say, the intuition that the good of our kin and of other people from the group we are part of is intrinsically more important than the good of other people. In that case, we should test or look for studies that test whether this intuition (or some other intuition) that leads people to deny Sidgwick's principle is a product of bias.[56] If it is biased, then, other things being equal, the principle should remain a candidate for self-evidence.

To be sure, we should reflect on the findings and conclusions drawn from each kind of study and aim at coherence.[57] In any case, even if there are no self-evident propositions, a method that meets these two requirements I proposed still seems better than the MWRE as traditionally defended. A method that requires us to empirically see whether the moral beliefs we hold non-inferentially tend to be accepted once they are adequately understood may help us to see whether we are accepting moral beliefs without understanding them properly. The way the MWRE is traditionally conceived does not require this and is, therefore, less likely to yield this desirable result. And a method that requires us to see whether we tend to accept certain moral beliefs for reasons we do not find morally relevant may lead us, as individuals, to be more cautious in accepting them. This also seems to be an advantage in and of itself, regardless of whether you are a moral realist.

[56] Maybe we could find out that this intuition is a product of evolutionary forces that do not tend to select beings for their ability to apprehend *a priori* truths. However, this is only one way such intuition could be biased. If there are other possibilities of bias, empirical studies should try to detect them. If all reasonable possibilities of bias have been investigated, our confidence that Sidgwick's candidate for self-evidence is indeed self-evident should, other things being equal, be diminished. This is indeed a difficult task, as detecting bias is difficult — let alone different types of bias. However, if this can be, it should.

[57] Studies that meet the two requirements at once may even be more effective in detecting self-evident propositions and, ultimately, leading us to have moral knowledge.

5. Conclusion

In this essay, I first tried to show that appealing to self-evidence is not a dogmatic move. This opens one door to taking self-evidence more seriously and considering the MWRE as a way to find self-evident propositions. Then, I argued that the method is more effective in leading us to acquire self-evidential justification than we might initially expect.

Being able to acquire self-evidential justification by applying the steps required by the MWRE, as traditionally conceived, makes us capable of selecting candidates for self-evidence. I argued, then, that we can filter out candidates for self-evidence by meeting the additional requirements I proposed. This could help us to decide which propositions to favour when we reflect on our considered judgements, principles, and moral and morally relevant non-moral theories.[58]

In the end, even if self-evident propositions do not exist, these additional requirements still seem to be desirable in the search for moral justification. We do not have to be moral realists to defend this. Not accepting these additional requirements might, then, not only run the risk of being further away from possible foundations of morality, but also lead us away from having more justified moral beliefs even according to a non-realist perspective. [59]

[58] Admittedly, many more details could be added to these requirements. For example, details of what kind of tests could be applied in the empirical studies in order to identify whether subjects adequately understand the propositions in question.

[59] I would like to thank, for their invaluable comments, the two referees and the member of UPJA's editorial team who have reviewed my essay. Also, thanks to my supervisor, Antonio Saturnino, for all the discussions we had during the time I did my research on the method of reflective equilibrium and intuitionism; and to my colleague, Pedro Freire, for the very helpful comments and conversations on an earlier version of this essay.

References

Audi, R. (2015). 'Intuition and Its Place in Ethics.' *Journal of the American Philosophical Association.* Vol. 1. pp. 57–77. Cambridge University Press.

Audi, R. (2018). 'Understanding, Self-evidence and Justification.' Philosophy and Phenomenological Research, vol. 99, Issue 2, pp. 358-381. United States of America: Wiley-Blackwell.

Brink, D.O. (1989). *Moral Realism and the Foundations of Ethics*. Cambridge: Cambridge University Press.

Cuneo and R. Shafer-Landau (2014) 'The Moral Fixed Points: New Directions for. Moral Nonnaturalism.' *Philosophical Studies.* Vol. 171(3). pp. 399-443

Daniels, N. (1979). 'Wide Reflective Equilibrium and Theory Acceptance in Ethics.' *The Journal of Philosophy*, vol. 76, no. 5, pp. 256-282. New York: Journal of Philosophy, Inc.

Daniels, N. (2016). 'Reflective Equilibrium.' *The Stanford Encyclopedia of Philosophy.* Retrieved February 28, 2021 from https://plato.stanford.edu/entries/reflective-equilibrium/

De Lazari-Radek, K.; Singer, P. (2014). *The Point of View of the Universe: Sidgwick and Contemporary Ethics.* Oxford: Oxford University Press

DePaul, Michael R. (2006). "Intuitions in Moral Inquiry." *The Oxford Handbook of Ethical Theory*, edited by David Copp, 595–623. Oxford: Oxford University Press.

Ebertz, R. (1993). 'Is Reflective Equilibrium a Coherentist Model?' *Canadian Journal of Philosophy*, vol. 23, no. 2, pp. 193-214, Cambridge University Press.

Goodman, N. (1955). *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press.

Haidt, J., & Baron, J. (1996). Social roles and the moral judgement of acts and omissions. *European Journal of Social Psychology*, Vol. 26(2). pp. 201–218

Joyce, R. (2006). *The Evolution of Morality.* Cambridge: MIT Press.

Kamm, F. M. (1993). Morality, Mortality, Vol. I: Death and Whom to Save From It. New York: Oxford University Press.

Lewis, D. (1983). *Philosophical Papers*, Vol. I. Oxford: Oxford University Press.

McMahan, J. (2000). "Moral Intuition." *The Blackwell Guide to Ethical Theory.* Edited by LaFollette Hugh, 92–110. Oxford: Blackwell.

McMahan, Jeff. 2000. "Moral Intuition." *The Blackwell Guide to Ethical Theory*, edited by LaFollette Hugh, pp. 92–110. Oxford: Blackwell.

Moore, G.E. (1903). *Principia Ethica*. London: Cambridge University Press.

Petrinovich, L., & O'Neill, P. (1996). Influence of wording and framing effects on moral intuitions. *Ethology & Sociobiology*, Vol. 17(3). pp. 145–171.

Prichard, H.A. (1912). 'Does moral philosophy rest on a mistake?'. *Moral Obligation*. Oxford: Oxford University Press.

Rawls, J. (1974-5). 'The Independence of Moral Theory'. *Proceedings and Addresses of the American Philosophical Association*, vol. 48, pp. 5–22. American Philosophical Association.

Rawls, J. (2001). *Justice as Fairness: A Restatement*. Cambridge, Mass.: Belknap Press of Harvard University Press

Rawls, J., (1999). *A Theory of Justice*, 2nd edn. Cambridge, Mass.: Belknap Press of Harvard University Press.

Ross W.D. (1930). *The Right and the Good.* Oxford: Oxford University Press.

Scanlon, T. M. (2003). 'Rawls on Justification.' In The Cambridge Companion to Rawls, edited by S. Freeman (ed.), pp. 139-167. Cambridge: Cambridge University Press.

Sidgwick, H. (1907). *The Methods of Ethics*, 7th edn. London: Macmillan.

Smith, M. (1994). *The Moral Problem*. Oxford: Blackwell Publishers.

Street, S. (2006). 'A Darwinian Dilemma for Realist Theories of Value.' *An International Journal for Philosophy in the Analytic Tradition.* Vol. 127, No. 1. pp. 109-116. Springer.

Tversky, A.; Kahneman, D. (1981). 'The Framing of Decisions and the Psychology of Choice'. *Science*, Vol. 211. pp. 453–8.

Walden, Kenneth. (2013). 'In Defense of Reflective Equilibrium.' Philosophical Studies, vol. 166, no. 2, pp. 243–256.

ISSN: 2653-3146

Undergraduate Philosophy Journal of Australasia www.upja.online
Vol. 3, December 2021

Beyond the Soul and Virtue: Benefit in Stoic Ethics

Yunlong Cao[1]

Johns Hopkins University

Abstract

Readers of Stoic ethics may find 'benefit' (*ōpheleia*) an essential but enigmatic concept. It directly connects to some critical terms of Stoic ethics, such as 'good' and 'virtue,' but there is no extant discussion of a definition. Beyond the superficial connections, what makes 'benefit' beneficial? Why is benefit a good thing? I argue that these essential questions remain unanswerable for a good reason: Stobaeus committed to a specious claim about benefit in his Anthology, which has misguided later commentaries. Either the Stoics themselves made a stronger contrast between sages and inferior people at the cost of coherence, or Stobaeus simply mischaracterized the Stoics' ideas in his descriptions. This paper aims to clarify Stobaeus's inaccurate description and reconstruct a coherent and comprehensible interpretation of benefit in the Stoic spirit, with the help of Stoic cosmology. To benefit is to further nature's agreement. Given the available evidence, I argue that Stoics seem to, or should, be committed to my interpretation. This paper is structured as follows. Section 1 offers a quick background of Stoic ethics. Section 2 discusses two important characteristics of benefit. Section 3 discusses Stobaeus's description of benefit and inferior people. Section 4 attempts an interpretation of benefit. Finally, Section 5 discusses Inwood and Gerson's interpretation and argues that it is inadequate.

1. Background: Virtue, Sage, and Appropriate Action

The Stoic value system divides things into three categories: the good, the bad, and the indifferents. The virtues (*aretē*) are good, including "prudence, justice, courage, moderation, and the rest".[2] The vices (*kakia*) are bad, including "foolishness,

[1] *Yunlong Cao is pursuing a combined graduate degree in philosophy at Johns Hopkins University. His main philosophical interests are epistemology, metaphysics, and the history of philosophy.*
[2] Diogenes Laertius, 7.101–3, in *The Hellenistic Philosophers*, 58A.

ISSN: 2653-3146

injustice, and the rest".[3] The indifferents (*adiaphora*) include everything else, such as "life, health, pleasure, beauty, strength, wealth, reputation, noble birth, and their opposites, death, disease, pain, ugliness, weakness, poverty, low repute, ignoble birth and the like…".[4] Though we may often prefer some indifferents over their opposites—e.g., life over death—the preferred indifferents are far less important than virtues.

The Stoics also divide people into two categories. The sages (*sophos*) are virtuous, wise, and happy all at once.[5] Sages are always happy because "happiness consists in virtue since virtue is a soul which has been fashioned to achieve consistency in the whole of life".[6] In addition, sages possess all virtues. "Menedemus of Eretria eliminated the plurality and the differentiations of the virtues, holding that there is a single one, called by many names; for it is the same thing that is called moderation and courage and justice, like 'mortal' and 'man'".[7] Furthermore, sages never make mistakes, not because they know everything in the world, but because they know when to withhold judgement if a mistake is possible. According to Stobaeus, "They [the Stoics] say that the wise man [a sage] never makes a false supposition, and that he does not assent at all to anything incognitive, owing to his not opining and his being ignorant of nothing".[8] In brief, the Stoics set the bar to be a sage so high that only few in humanity have ever achieved virtue.

The rest of us are all inferior people (*phaulos*).[9] For the Stoics, if one is not virtuous and wise, then he is vicious, silly, inferior, and base. One cannot be partly virtuous and partly inferior. Diogenes offered an illustration — "a stick must be either straight or crooked, so a man must be either just or unjust, but not either more just or more unjust, and likewise with the other virtues".[10] Oddly, the Stoics would call the vast majority of humanity inferior and vicious; however, this does not mean that nobody can do anything right. With some help from luck or the advice of sages, we non-sages are still capable of appropriate actions (*kathēkonta*), "which [have] an affinity with arrangements that are according to nature."[11] Appropriate actions are the best actions in a given circumstance. Even better than appropriate actions are the

[3] Ibid.
[4] Ibid.
[5] I follow Prof. Richard Bett in using the gender-neutral translations, 'sages' and 'inferior people', instead of the traditional gendered translations, "wise men" and "base men". For an argument for the gender-neutral translations, see Bett 2012, p. 531. For an argument against the gender-neutral translations, see Inwood and Gerson, p. ix.
[6] Diogenes, 7.89 in *The Hellenistic Philosophers*, 61A.
[7] Plutarch, "On Moral Virtue", 440E–441D, in *The Hellenistic Philosophers*, 61B. For an explanation on the inter-entailment of virtues, see Bett 2012, p. 533.
[8] Stobaeus, 2.111,18–112,8, in *The Hellenistic Philosophers*, 41G.
[9] 'Inferior' could be considered harsh by the general public, but it appropriately captures the negative connotations of the Greek term *phaulos*. Admittedly, calling almost everyone vicious or inferior is one of the strange things about Stoic ethics.
[10] Diogenes, 7.127, in *The Hellenistic Philosophers*, 61I.
[11] Diogenes, 7.108, in *The Stoic Reader*, 101.108.

ISSN: 2653-3146

right actions (*katorthōmata*), a special kind of appropriate actions done by the sages according to their virtues.

The Stoics take these two theses to be compatible: 1) Inferior people can make progress towards virtue; 2) All inferior people are equally inferior. Plutarch quoted an illustration: "just as in the sea the man an arm's length from the surface is drowning no less than the one who has sunk five hundred fathoms, so even those who are getting close to virtue are no less in a state of vice than those who are far from it".[12] Inferior people closer to virtue may perform more appropriate actions, but their soul is in the same inferior state with other inferior people. According to Chrysippus, "The man who progresses to the furthest point performs all proper functions [i.e., appropriate actions] without exception and omits none".[13] Though inferior people are all equally vicious, one can, in some sense, make progress towards virtue by performing more and more appropriate actions. Furthermore, "virtue is teachable… as is evident from the fact that inferior men become good [i.e., virtuous people]".[14] Though the Stoics set a high bar for virtue, they are still optimistic that inferior people can become sages.

2. Two Characteristics of the Stoic Conception of Benefit

This section aims to locate 'benefit' in the Stoic conceptual framework. There are two main characteristics of benefit agreed upon and emphasized by the Stoics. First, benefit is defined by virtue. Diogenes wrote that "to benefit is to change (*kinein*) or maintain (*ischein*) something in accordance with virtue, whereas to harm is to change or maintain something in accordance with vice."[15] If an action is done per one's virtue, someone receives some benefit from this action. Moreover, benefit helps define good. Sextus Empiricus wrote that "the Stoics, sticking fast to the common conceptions so to speak, define the good as follows: "Good is benefit or not other than benefit, meaning by 'benefit' virtue and virtuous action… For virtue, which is a disposition of the commanding-faculty, and virtuous action, which is an activity in accordance with virtue, are benefit directly."[16] Benefit is virtue or virtuous action, and it is the only genuinely favourable action. The Stoics emphasized that benefit is related to virtue and virtue alone. Preferred indifferents, such as health and wealth, might serve as means of virtuous actions. For example, a sage can give to charity according to her virtue. However, what really benefits is the sage's virtuous action, not the money she gave.

[12] Plutarch, "On Common Conceptions", 1063A–B, in *The Hellenistic Philosophers*, 61T.
[13] Stobaeus, 5.906,18–907,5, in *The Hellenistic Philosophers*, 59I.
[14] Diogenes, 7.91, in *The Hellenistic Philosophers*, 61K. The Stoic definition of 'good' includes virtuous people. See *The Hellenistic Philosophers*, 60G.
[15] Diogenes, 7.102, in *The Stoics Reader*, 101.102.
[16] Sextus Empiricus, "Against the professors", 11.22–6, in *The Hellenistic Philosophers*, 60G.

ISSN: 2653-3146

Second, the Stoics also emphasized that benefit is common (*koinos*). Stobaeus wrote that "all goods are common to the virtuous, and all that is bad to the inferior. Therefore a man who benefits someone also benefits himself, and one who does harm also harms himself. All virtuous men benefit one another … but the foolish are in the opposite situation."[17] Benefit is shared by all virtuous people. Thus, a virtuous action does not only benefit an individual or a small group, but every virtuous person simultaneously. I will address two alternative interpretations of Stobaeus's quote.

One may explain the commonness of benefit as a result of the social practices of the virtuous. Contingently, sages only benefit but never harm each other. If sage Alex benefits sage Bailey today, Bailey will benefit Alex back tomorrow; sages always choose to benefit each other. Although this interpretation captures sages in interaction, it does not explain Stobaeus's description that: "all virtuous men benefit each other, even though they are not in all cases friends of each other or well disposed [to each other] … because they do not have a [cognitive] grasp of each other and do not live in the same place."[18] Sages can benefit each other without even knowing each other. If we interpret common benefit as merely reciprocal, then sages have to pick out specific benefit receivers for each virtuous action, which requires interaction. However, Stobaeus argued that interaction is not required for benefit between sages. Therefore, interpreting common benefit as reciprocal social practices is inadequate.

Another explanation of the commonness of benefit appeals to the structure of souls and actions: if sages benefit someone, they also benefit their own souls. For example, if sage Catherine benefits sage Dive with a virtuous action, she also benefits herself by making her soul more virtuous. Although this interpretation captures the Stoic emphasis on the soul, it leads to a dilemma: nobody can benefit and improve their soul at the same time. Sages can benefit, but they already have perfect souls for they are perfectly virtuous. How can they further improve their soul beyond perfection? Inferior people can make progress, but they neither possess virtue nor can perform virtuous actions. How can they benefit anyone? I will further argue against this interpretation in Section 5.

3. A Tension in the Descriptions: Can Inferior People Receive Benefit?

Stobaeus did not think inferior people can receive benefit. He explicitly wrote that "no base man [i.e., an inferior person] either benefits or is benefited."[19] By definition, inferior people are not virtuous; thus, they cannot benefit because they cannot act

[17] Stobaeus, 2.101,21–102,3, in *The Hellenistic Philosophers*, 60P.
[18] Stobaeus, "Anthology, 2.11", in *The Stoics Reader*, 102.11i.
[19] Stobaeus, "Anthology, 2.11', in *The Stoics Reader*, 102.11d.

per their nonexistent virtue. However, I argue that Stobaeus's commitment that inferior people cannot *be benefited* is mistaken and untenable for three reasons. Once this commitment is corrected, we will be in a position to understand benefit.

3.1 A Sage-Inferior-Person Interaction

First, Stobaeus cannot explain a sage-inferior person (hereafter, SIP) interaction. "All virtuous men benefit one another … but the foolish are in the opposite situation."[20] When two sages interact, they benefit each other; when two inferior people interact, they harm each other. This characterization does delineate a clear contrast between sages and inferior people. However, what happens when a sage interacts with an inferior person? They cannot benefit each other: the inferior person cannot be benefited; they cannot harm each other: the sage cannot be harmed. However, benefit and harm are supposed to be common. How can Stobaeus make sense of a SIP interaction? Is such interaction possible?

One possible response is acknowledging the possibility of such interaction but denying that benefit and harm are reciprocal between sages and inferior people. The sage receives benefit, but the inferior person receives harm. However, this interpretation suggests that it is virtuous for sages to 'take advantage' of the inferior people. If a sage knows their interaction can only harm the inferior, will it still be virtuous for her to interact with the inferior person? Though the sage is benefited, would the Stoics be willing to accept such selfish actions as ethical and virtuous? Harm, like benefit, is not about indifferents but about souls. Therefore, the Stoics cannot simply dismiss this harm as trivial, rendering this response problematic.

The other possible response is to deny the possibility of a SIP interaction. One possible argument is that sages and inferior people cannot interact because sages are rare. If there were no sages alive while Stobaeus was writing, they could not interact with inferior people. Another possible argument is that virtuous people tend to live together and perhaps in one city. In such cases, the separation between sages and inferior people dismisses the theoretical tension within SIP interactions. Neither inferior people are benefited, nor are sages harmed, due to the lack of interaction.

However, I argue that the appeal to the separation between sages and inferior people is not only inadequate to resolve this tension, but also goes against the Stoic spirit. This separation lacks the modal strength to provide a theoretical solution to the tension. It could at most describe what is happening in the world: most inferior people do not know many sages, so they are not often benefited. However, this statistical generality does not suggest that inferior people *cannot* be benefited. Furthermore, this separation conflicts with the Stoic view of the community. Cicero

[20] Stobaeus, 2.101,21-102,3, in *The Hellenistic Philosophers*, 60P.

wrote that "we are driven by nature to desire to benefit as many people as possible…".[21] Separating sages and inferior people into two different worlds prevents sages from benefiting as many people as possible. Hierocles argued that one should care about the entire human race with his illustration of concentric circles. He wrote that "the outermost and largest circle, which encompasses all the rest, is that of the whole human race…it is the task of a well tempered man, in his proper treatment of each group, to draw the circles together somehow towards the centre…".[22] The Stoics teach people to concern themselves with the whole human race, and thus separating sages and inferior people is against the Stoic spirit.

SIP interactions are necessary for Stoicism. If we allow that inferior people can be benefited, sages can interact with inferior people and hence benefit them in accordance with their virtues.

3.2 The Logic of Stobaeus's Description

Stobaeus might not have meant to argue that inferior people cannot be benefited, because the supporting sentence does not entail the claim. He wrote that "…no base man [i.e., an inferior person] either benefits or is benefited. For benefiting is <to change> or maintain something in accordance with virtue, and being benefited is to be changed in accordance with virtue".[23] The conjunctive "for," suggests that the latter sentence supports the former. However, this support falls short of an entailment. I reconstruct his argument in standard form:

> (P1) *Benefiting* is <to change> or maintain something in accordance with virtue.
> (P2) Inferior people cannot *change or maintain* something in accordance with virtue.
> (C3) Thus, inferior people cannot *benefit*.

This argument is sound because the implicit assumption (P2) is true. Inferior people do not possess virtue; therefore, they cannot act in accordance with virtue. Consider the other argument:

> (P4) *Being benefited* is to be changed in accordance with virtue.
> (P5) Inferior people cannot *be changed* in accordance with virtue.
> (C6) Thus, inferior people cannot *be benefited*.

---

[21] Cicero, "On Ends, 3.62-8", in *The Hellenistic Philosophers*, 57F.
[22] Hierocles, qtd. in Stobaeus 4.671,7–673,11, in *The Hellenistic Philosophers*, 57G.
[23] Ibid. Please note that "angle brackets < > represent material supplied by the editors [i.e., Inwood and Gerson] of the original Greek and Latin texts to repair a gap in the ancient text caused by damage or omission by scribes in the course of transmission" ("Abbreviations and Conventions", in *The Stoics Reader*).

ISSN: 2653-3146

This argument is unsound because the implicit assumption (P5) is false. Nothing intrinsically prevents an inferior person from being changed in accordance with virtue. Whether an action is in accordance with virtue depends on the subject instead of the object. If the subject is virtuous, the action is guaranteed to be virtuous. The vice of an object cannot override the subject's virtue.

"To change or maintain'' can have different interpretations. A stronger interpretation requires a causal role of the action. For example, a sage gives some advice to an inferior person. If the inferior people would act otherwise without the sage's advice, then the sage changed or maintained something. A weaker interpretation requires an action to change something in the modal strength. In this example, if the inferior people would perform an appropriate action without the sage's advice, but the sage's advice made the appropriate action more likely, then it changed or maintained something. Since the Stoics emphasized the modal strength of one's reasoning behnnd actions, as evidenced by their distinction between appropriate and right actions, I suggest the weaker interpretation. However, the sage can benefit the inferior people on both interpretations. They just do so slightly less often when we adopt the stronger interpretation.

3.3 Clement's Description of God's Benefit

Stobaeus's commitment conflicts with an argument made by another Stoic — Saint Clement of Alexandria — that God benefits everyone. He argued that "God is agreed to be good; therefore god benefits. But the good in so far as it is good does nothing but benefit; therefore god benefits everything… therefore god cares for and attends to man".[24] God benefits everyone, virtuous and inferior alike. If benefit is only available among the virtuous, how can God benefit the inferior people?

One possible response is that God's benefit and humans' benefit are two different notions. Humans can only benefit sages, but gods can benefit everyone. However, this is incorrect because the distinction conflicts with the idea that god(s) and virtuous humans benefit each other equally. Plutarch wrote that "according to Chrysippus… Zeus does not surpass Dion [i.e., a random name of a sage] in virtue, and Zeus and Dion are benefited equally by each other when one meets with a motion of the other since they are [both] wise".[25] God(s) and virtuous humans benefit each other equally because they do not differ in virtue. Thus, there is only one kind of benefit in a god-sage interaction. Furthermore, the direct connection between virtue and gods' benefit suggests that there is only one concept of benefit. The Stoics were not equivocating two notions of benefit.

[24] Clement, "The Teacher", 1.8.63.1–2, in *The Hellenistic Philosophers*, 60I.
[25] Plutarch, "On Common Conceptions", 1076A, in *The Stoics Reader*, 32.

A worry is that Saint Clement, a Christian theologian, could have appropriated the Stoic term for his own purpose. Stoics could have argued that inferior people could not receive benefit, which is against the Christian spirit that God benefits everyone. Thus, Clement altered the meaning of benefit slightly to be compatible with Christian ideas. Therefore, we should not regard Clement's description as evidence of this particular interpretation of Stoic ideas. I think that we should interpret Clement's argument with caution, possibly as "applied Stoicism," because of his Christian influence. However, we should also not completely disregard this evidence, because Clement's argument's conceptual relation between 'good' and 'benefit' matches the one in other Stoics' descriptions. As I quoted before, Clement argued that "God is agreed to be good; therefore God benefits".[26] There is a direct connection between the concepts of good and benefit, which matches the Stoic definition of good as "benefit or not other than benefit".[27] This agreement suggests that relations between concepts in Clement's description are isomorphic to relations between concepts in descriptions by other Stoics: if Clement appropriated the Stoic concept of benefit, he must also have appropriated good and virtue. Thus, it still supports that inferior people could be benefited. When these two Stoics disagree, I do not argue that Clement is a more trustworthy source than Stobaeus. However, the existence of Clement's passage offers another piece of evidence that Stobaeus's commitment could have been a mistake.

In brief, I argue that allowing that inferior people could be benefited offers a more consistent picture of Stoicism. Although Stobaeus explicitly wrote that inferior people could not be benefited, he either simply mischaracterized the Stoics' ideas or the Stoics themselves made a stronger contrast between sages and inferior people at the cost of coherence. I argue that inferior people could be benefited for three reasons. First, the evidence does not entail the claim that they cannot be. Second, Stobaeus cannot explain a SIP interaction. Third, Stobaeus is in conflict with Clement. From the last two sections, I have gathered three features of benefit from the Stoics' descriptions: (1) benefit is about virtue rather than indifferents; (2) benefit is common; (3) inferior people and sages alike can receive benefit. In the next section, I propose an interpretation in accordance with the above three descriptions and explain why benefit is beneficial.

4. Understanding Benefit: Bridging Stoic Ethics and Cosmology

I argue that benefit furthers the agreement of nature. A virtuous action brings nature into a state of agreement more than its alternatives could. Conversely, an inferior action takes away others' opportunity to perform a virtuous action, advancing nature's agreement less than its virtuous alternative could. Specifically, virtuous

[26] Clement, "The Teacher", 1.8.63.1–2, in *The Hellenistic Philosophers*, 60I.
[27] Sextus Empiricus, "Against the professors", 11.22–6, in *The Hellenistic Philosophers*, 60G.

actions make people agree in everyday activities. Generally, virtuous actions help inferior people progress towards virtue, thus furthering the universe's agreement towards the ultimate conflagration.[28] I will first explain my interpretation, then support it with two arguments.

The ultimate agreement of nature consists in everything being virtuous. The Stoics believed that nature would return periodically into a state of virtuous fire. Plutarch wrote that "Whenever they [the Stoics] subject the world to conflagration, no evil at all remains, but the whole is then prudent and wise".[29] At that time, everything agrees because everything is virtuous. However, this is not the case at the time this paper is being written: at least part of the universe, namely the author himself, is an inferior person who lacks virtues. With this difference, I propose a way to explain the degree and progress of nature's agreement, similar to how the Stoics explained an individual's progress towards virtue. Nature is said to be in more agreement, or closer to the state of virtuous fire, if more members, such as individual human beings, are virtuous.

This cosmological agreement *of* nature is compatible with Stoicism's ethical end (*telos*)—living in agreement *with* nature (*homologoumenōs tēi phusei zēn*). Diogenes summarized Chrysippus's explanation, "living in agreement with nature comes to be the end, which is in accordance with the nature of *oneself* and that of *the whole*, engaging in no activity wont to be forbidden by the universal law, which is the right reason pervading everything and identical to Zeus, who is this director of the administration of existing things".[30] Our ethical goal is to live according to nature's plan. If we can do so and achieve virtue, we further nature's agreement by making more things virtuous. The benefit of virtuous actions can further be understood in these terms.

The specific agreement of nature can manifest itself as concord among people. I wish to emphasize that "[sages] are in agreement about matters concerned with life".[31] Virtuous actions are disjunctive when conflicts shall be avoided but concerted when cooperation is preferred. Here is an illustration: Three different jobs (farmer, blacksmith, and fisherman) are available for three sages (Eric, Frank, and Grace). According to their nature, each sage should have a best option, and their options shall not conflict. Since they are virtuous and wise, it will be easy for the sages to sort out each other's talents and preferences. Therefore, they act in agreement with nature's ideal plan. The order of action does not matter; when nature plans for Eric to be a farmer, this plan is commonly known to the group. Frank and Grace will not

[28] The Stoics believe that our world will eventually return to a state of virtuous and intelligent designing fire. See Section 3 "Physical Theory" of SEP entry *Stoicism* and Chapter 52 "Everlasting Recurrence" of *The Hellenistic Philosophers*.
[29] Plutarch, "On Common Conceptions", 1067A, in *The Hellenistic Philosophers*, 43N.
[30] Diogenes, 7.87–9, in *The Hellenisitic Philosophers*, 63C.
[31] Ibid.

choose to be a farmer even if they can choose before Eric. No matter what happens, sages agree with each other, therefore they are in the best agreement with nature. Thus, sages benefit each other by actualizing nature's best plan.

Similarly, nature not being in agreement can manifest itself as discord among people. Suppose the same task is assigned to three inferior people: Harry, Isabella, and John. Nature's ideal plan is inaccessible to all participants. Suppose Harry chooses first, and if he was a sage, Harry should choose to be a fisherman. However, due to his ignorance, Harry only has a one-in-three chance to pick the appropriate action. Thus, it is quite likely that Harry chooses either Isabella's or John's best option. The inferior people cannot manifest nature's ideal plan without significant luck. They can then become hostile: e.g., Isabella and John both want to be the blacksmith, so they fight for it. However, the sages will never be in such a situation. Inferior people harm each other by not furthering nature's agreement, while also making nature agree less when the inferior action is not an appropriate action. The illustration only considers three inferior people, so some luck seems tolerable; however, many more agents are involved in each real-world action. Therefore, it is extremely hard for inferior people not to harm other people.

However, the agreement of nature is more abstract than in my illustrations; benefit and harm have a deeper meaning than a superficial agreement. What ultimately benefits everyone is the ideal world where everyone is virtuous, and nature is not in complete agreement until then.[32] When more people are virtuous, nature agrees better. Furthering nature's agreement has a twofold meaning. Specifically, a sage actualizes nature's ideal plan at any given moment; generally, a sage helps people progress in their virtue, ultimately bringing nature into a state of agreement. "Furthermore we are driven by nature to desire to benefit as many people as possible, and especially by giving instruction and handing on the principles of prudence".[33] Cicero's quote supports the more abstract notion of nature's agreement. Sages wish to benefit people with the progress of their souls, which ultimately makes nature in total agreement.

With virtue, a sage's action is guaranteed to agree with nature. If an inferior person acts in accordance with her vices, then very likely, nature's ideal plan is "disrupted." The inferior person harms others with mismatches and conflicts. One potential worry is that given the Stoic providentialist view of the universe, how can an action benefit more? If non-virtuous actions are also part of nature's plan, how can they still harm? Would not all actions carry out nature's ideal plan, and thus, are virtuous?

[32] Stoics also believed that we are already in the best possible world with God's providence. (See Bett 2009, vol. 6, 392.) Please note that this belief is compatible with nature's incompleteness, which is an *internal* comparison among temporal stages within our world. However, the Stoic best possible world belief concerns an *external* comparison with other possible worlds.
[33] Cicero, "On Ends", 3.62-8, in *The Hellenistic Philosophers*, 57F.

According to Long and Sedley, "On the Stoic view determinism and moral responsibility are not merely compatible, they actually presuppose each other".[34] Thus, I assume that human responsibility can be preserved, though it is possible that the Stoics themselves were not consistent. With a potential solution to resolve the tension between human responsibility and fate, a solution for the tension between benefit and fate should be similar. Thus, for the purpose of this paper, I wish to bracket the tension between benefit and determinism.

Since there is no extant definition of benefit in the primary sources, I cannot back my interpretation with direct evidence. I propose my interpretation with two arguments. First, the textual position suggests that benefit and concord are related. Second, interpreting benefit with nature's agreement satisfies the three characteristics discussed in Sections 2 and 3. I do not argue that my interpretation is the only possibility. It is possible to have another consistent interpretation that can satisfy all three characteristics of benefit. In that case, I do not have any argument for my interpretation beyond the textual position of Stobaeus's discussion of concord. However, no such interpretation yet exists in a well-known and acknowledged format, and the logical space for benefit in the Stoic conceptual framework is not very broad given the three confining characteristics. Therefore, I propose my interpretation as a pioneering approach for the topic.

First, the textual position suggests that benefit and concord are related concepts. While I do not argue they are identical, concord and benefit also share a tight connection similar to the relations among good, virtue, and benefit. Stobaeus wrote that:

> [The Stoics] say that all good things belong <in common> to the virtuous, in that he who benefits one of his neighbors also benefits himself. Concord (*homonoia*) is a knowledge of common goods, and that is why all virtuous men are in concord with each other, because they are in agreement about matters concerned with life. The base [i.e., inferior] are enemies and do harm to each other and are hostile, because they are in discord (*diaphōnein*) with each other.[35]

Stobaeus mentioned concord and discord between the discussions of benefits and harms. The structure suggests that the concept of concord and discord is connected to benefit and harm.

In addition, my interpretation satisfies all three characteristics of benefit. 1) This account of benefit is related to virtue because only sages know for sure how to act in agreement with nature. Right action is guaranteed to agree with nature. Indifferents

[34] Long and Sedley 1987, Commentary on Ch. 62.
[35] Stobaeus, "Anthology", 2.11, in *The Stoics Reader*, 102.11b.

cannot intrinsically benefit because indifferents cannot further nature's agreement. Sometimes preferred indifferents can be selected according to nature's plan, but at other times they do not. 2) Benefit and harm are common to everyone because there is only one agreement of nature: advancing or disrupting its agreement affects everyone simultaneously. Two sages can benefit without knowing each other because they advance the agreement of nature, which is shared by the other. 3) Sages and inferior people alike can be benefited because we share the same nature. In the next section, I will compare and contrast my interpretation with an alternative interpretation of benefit.

5. An Alternative Interpretation and Its Inadequacy

In the Glossary of the Stoics Reader, Inwood and Gerson wrote that "in Stoicism, benefit is narrowly defined in terms of the attainment or preservation of a virtuous state of one's soul".[36] This interpretation follows Stobaeus's description that inferior people cannot be benefited since they do not have a virtuous state of soul to be attained or preserved. I have argued against this description in Section 3, showing Inwood and Gerson's interpretation is at least incomplete in that respect. In addition, although they point out a crucial feature of benefit—it is narrowly about virtue and souls, instead of the everyday "benefit" of indifferents—their definition is inadequate because it cannot distinguish virtuous actions from the inferior ones.

Stoics did not only use benefit as a noun (*ōpheleia*) but also as a verb (*ōphelein*). Therefore, benefit is a relation between people. Benefit is common: when one benefits, each receiver has to be benefited. As a noun, benefit can be understood as the attainment or preservation of virtue. One receives benefit if one's virtue is attained or preserved. As a verb, it is more obscure. Can a person preserve the virtuous state of another person's soul? I argue that one cannot. Imagine two sages, Kiely and Laiken. Kiely performs a virtuous action, e.g., walking around prudently, which should benefit all sages, including Laiken.[37] According to Inwood and Gerson's definition, Kiely helps Laiken either attain or preserve her virtue. Trivially, Kiely cannot help Laiken attain virtue because Laiken was presupposed to be virtuous. Therefore, Kiely must help Laiken preserve her virtue. I argue this is impossible for two reasons. First, preservation assumes that virtues can be lost; if not, all actions preserve virtue, making virtue preservation a meaningless tautology. This assumption is contentious yet still possible. According to Diogenes, "Chrysippus holds that virtue can be lost, on account of intoxication or depression,

[36] Inwood and Gerson 2008, p. 206.
[37] I used walking prudently as an example for the ease of illustration. If a sage does an action, it is guaranteed to be a right action. In addition, walking is one of the classical illustrations in the Stoic discussions. For an example, see Stobaeus, in *The Hellenistic Philosophers*, 59M.

but Cleanthes takes it to be irremovable owing to secure cognitions."[38] Though there is no extant elaboration on what does "intoxication or depression" mean, we could reasonably assume that most actions done by inferior people cannot intoxicate or depress a sage; otherwise, sages have quite volatile souls, which is against the Stoic spirit that virtues come with fixity.[39] Therefore, inferior actions cannot often endanger the virtuous state of a sage's soul.

Second, can Kiely's action preserve Laiken's virtue any more than a random inferior action? Imagine a counterfactual world in which an inferior person, Max, was walking imprudently. Compared to the actual world, the state of Laiken's soul is not affected at all: whether their walk is prudent or not, Laiken can choose the appropriate action (or the right action since she is a sage), i.e., the best available walking route, for herself. Once a sage achieves virtue, she cannot easily lose the virtue for its fixity. To contrast, consider an inferior person who performs many appropriate actions, but does not have virtues yet. The inferior person's state of soul can be changed by others because she is not yet wise. The preservation of Laiken's virtue is irrelevant to Kiely's or Max's walking around because a sage's soul is not easily disturbed. If Max's inferior action and Kiely's virtuous action do not alter Laiken's virtuous soul, how can Kiely benefit Laiken while Max does not, under Inwood and Gerson's interpretation? Even if Chrysippus is correct, a sage can only lose virtues in unusual circumstances, e.g., Max imprudently trips Laiken, and Laiken's head falls on a stone which happens to cause a severe concussion, and so forth. If this ad hoc possibility is in a close possible world, Kiely's prudent walk will preserve Laiken's virtue. However, we cannot expect a rare counterfactual case to happen for each virtuous action. Therefore, a sage's virtue does not have to be constantly preserved.

If we interpret benefit as furthering the agreement of nature, then virtuous actions can be distinguished from inferior actions. Suppose nature has an ideal plan for each person to walk on a specific route. Kiely benefits Laiken by walking on her own route instead of Laiken's. Therefore, they achieve concord and are in agreement with nature. If Nat, trying to study virtue, joins the sages walking around, it is easier for him to choose his ideal route since the sages already rule out two possibilities for him. Therefore, ultimately, Laiken and Kiely help Nat progress towards virtue. If the inferior person Max joins their walking and imprudently takes Laiken's route, Laiken has to choose her second-best route. She can remain virtuous since a sage should not be disturbed by another person's walking. Laiken was not benefited because of the lack of agreement. However, she was not harmed either because ultimately, the lack of agreement and inferior population only remain: Max was not virtuous in the first place.

---

[38] Diogenes, 7.127, in *The Hellenistic Philosophers*, 61I.
[39] See *The Hellenistic Philosophers*, 59I.

In general, understanding benefit as preservation of virtue is inadequate because virtue is stable. Sages do not need benefit to actively preserve their virtues at all times. Thus, it is challenging to distinguish benefit from inferior actions. However, agreement of nature is a common task requiring cooperation. Since a sage cannot do everything by herself, my interpretation leaves space for the sages to receive benefits from other sages. The agreement of nature is a more suitable candidate for benefit than the attainment and preservation of virtue. I do not argue that one's attainment and preservation of virtue is excluded from the furthering of nature's agreement. The relation between the attainment of virtue and nature's agreement is similar to the relation between right actions and indifferents: a sage can prudently select some preferred indifferents as a right action. Though indifferents are part of this action, what makes this action virtuous is the sage's virtue. Similarly, sometimes, sages can benefit by helping others achieve virtues; but they do not have to. Benefit really consists of advancing nature's agreement, which is a guaranteed result from right actions.

6. Conclusion

Although we do not have a passage explicitly defining benefit and explaining why it is beneficial, we can reconstruct the concept with the Stoics' descriptions. Benefit is beneficial because it furthers nature's agreement. My interpretation of benefit suggests a strong connection between the virtuous state of the soul and the virtuous state of nature, tying the Stoic ethical and cosmological theories together. Though later Stoics appear to focus their philosophical interests solely on ethics, the earlier Stoics treat other areas of philosophy, such as cosmology, on a par with ethics.[40] The connection between cosmology and ethics in early Stoicism explains the unavoidable inadequacy of interpreting benefit merely as the virtuous state of the soul. One advantage of my interpretation is that nature's agreement can be understood at different levels: specifically, it is our everyday concord within the human community; generally, it is the virtuous state of nature. Hence my interpretation retains the Stoic emphasis on virtue over indifferents and captures the Stoic comprehensive view of nature simultaneously.

[40] Bett 2012, p. 530.

References

Cicero. "On Ends, 3.62–8." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 57F. Cambridge: Cambridge University Press. Apple Books Edition.

Clement. "The Teacher, 1.8.63.1–2." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 60I. Cambridge: Cambridge University Press. Apple Books Edition.

Diogenes Laërtius. "7.102,8." In The Stoics Reader: Selected Writings and Testimonia, edited by Brad Inwood and Lloyd P. Gerson, Text 101. Indianapolis: Hackett Publishing Company, Inc. Kindle Edition.

Diogenes Laërtius. "7.87–9;7.89; 7.91; 7.101–3; 7.127." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 63C, 61A, 61K, 58A, 61I. Cambridge: Cambridge University Press. Apple Books Edition.

Plutarch. "On Common Conceptions, 1063A–B; 1067A." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 61T, 43N. Cambridge: Cambridge University Press. Apple Books Edition.

Plutarch. "On Common Conceptions, 1076A." In The Stoics Reader: Selected Writings and Testimonia, edited by Brad Inwood and Lloyd P. Gerson, Text 32. Indianapolis: Hackett Publishing Company, Inc. Kindle Edition.

Plutarch. "On Moral Virtue, 440E–441D." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 61B. Cambridge: Cambridge University Press. Apple Books Edition.

Sextus Empiricus. "Against the Professors, 11.22–6." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 60G. Cambridge: Cambridge University Press. Apple Books Edition.

Stobaeus. "Anthology, 2.101,21–102,3; 2.111,18–112,8; 4.671,7–673,11; 5.906,18–907,5." In The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary, edited by A. A. Long, and D. N. Sedley, 60P, 41G, 57G, 59I. Cambridge: Cambridge University Press. Apple Books Edition.

ISSN: 2653-3146

Stobaeus. "Anthology, 2.11." In The Stoics Reader: Selected Writings and Testimonia, edited by Brad Inwood and Lloyd P. Gerson, Text 102.11b,d,i. Indianapolis: Hackett Publishing Company, Inc. Kindle Edition.

Baltzly, D. (2019). "Stoicism." The Stanford Encyclopedia of Philosophy, edited by Edward N. Zalta. https://plato.stanford.edu/archives/spr2019/entries/stoicism/.

Bett, R. (2009). "Stoicism." In Oxford Encyclopedia of Ancient Greece and Rome, edited by Michael Gagarin, vol. 6, pp. 389-95. Oxford: Oxford University Press.

Bett, R. (2012). "Stoic Ethics." In A Companion to Ancient Philosophy, edited by Mary Louise Gill and Pierre Pellegrin, pp. 530–48. Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781444305845.ch27.

Brennan, T. (2006). "Goods and Indifferents." In The Stoic Life: Emotions, Duties, and Fate. Oxford Scholarship Online. DOI: 10.1093/0199256268.001.0001.

Inwood, B, and L. P. Gerson. (2008). The Stoics Reader: Selected Writings and Testimonia. Indianapolis: Hackett Publishing Company, Inc. Kindle Edition.

Long, A. A., and D. N. Sedley. (1987). The Hellenistic Philosophers: Volume 1, Translations of the Principal Sources with Philosophical Commentary. Cambridge: Cambridge University Press. Apple Books Edition.

ISSN: 2653-3146

Undergraduate Philosophy Journal of Australasia www.upja.online
Vol. 3, December 2021

Practical Animal Reasoning

Martin Walter Niederl[1]

University of Vienna

Abstract

In his latest paper on animal agency, Glock (2019) presents a series of arguments to the extent that non-linguistic animals are capable of acting rationally and for reasons. This notwithstanding, he still denies them the ability to conceptualise reasons *as* reasons. I will argue that, in using Glock's account, one *can* in fact claim that non- linguistic animals are capable of conceptualising reasons as reasons. For this, I will apply Glock's own criteria for concept-possession to the concepts of a *reason* and of *intention*. My argument will thus be twofold. First, I will directly argue for the idea that animals can conceptualise reasons as reasons. Second, I will refer to empirical research suggesting that animals attribute intentions to others. If the ability to conceptualise intentions really is necessary for conceptualising reasons, then this research should provide further plausibility to the claim that animals can conceptualise reasons as reasons. I thus submit that my arguments will further improve upon Glock's account by (1) showing that animals can *conceptualise reasons as reason*, (2) lending further support to the idea that non-human animals can *act rationally*, and (3) providing some initial foundation for the claim that they can *reason*.

1. Introduction

Glock[2] argues (against many traditional views)[3] that non-human, non-linguistic animals are capable of acting rationally, for reasons, and in light of reasons. However, he stops short of arguing that animals can also reflect upon these reasons. This is because it seems that one has to be able to conceptualise them *as* reasons in order to reflect upon them. As such, he holds that the ability to reflect upon reasons depends on one's familiarity with the *concept* of a reason. The possession of this

[1] *Martin W. Niederl is a philosophy undergraduate at the University of Vienna, Austria. His research interests converge on the nature and normativity of practical reasons, as well as their connection to agency and moral responsibility.*
[2] Glock 2019.
[3] Cf. Alvarez 2010; Brandom 2010; Davidson 1982, 2001; Frankfurt 2004, pp. 18-19; Hacker 2007; Hampshire 1959; Marcus 2012; McDowell 1996; Stoecker 2009.

ISSN: 2653-3146

concept, in turn, requires a certain linguistic repertoire and thus excludes animals to the extent that they lack linguistic competences. I will argue, however, that, in using Glock's[4] own criteria for concept-possession, one *can* actually claim that animals can conceptualise reasons *as* reasons. My arguments will thus further improve upon Glock's current account in three ways: (1) by showing that animals can conceptualise reasons as reasons; (2) by lending further support to the idea that non-human animals can act rationally; and (3) by providing some initial grounds for the claim that animals can reason.

The paper will proceed as follows. Section 2 will be concerned with presenting Glock's[5] arguments for animals as rational agents, capable of acting for reasons and in light of reasons. In Section 3, I will present an objection, anticipated by Glock, which claims that language is necessary in order to act in light of reasons. Based upon this objection, I will then argue that non-linguistic animals can conceptualise reasons as reasons in Section 4. My argument for this idea will be twofold. First, I will directly argue that animals can conceptualise reasons as reasons according to Glock's account of concept-possession. One might object, however, that one can understand reasons as reasons only if one understands intentions as such. Thus, second, I will refer to empirical research suggesting that animals attribute intentions to others. If the ability to conceptualise intentions really is necessary for conceptualising reasons, then this research should provide further plausibility for the claim that animals can conceptualise reasons as reasons. Note that I will continue to talk of 'animals' mostly without qualification, however as the paper progresses it should become clear which non-linguistic animals I am referring to – namely those who satisfy Glock's conditions for agency. These would plausibly include chimpanzees, pigs, jays, and other animals that are similarly cognitively equipped.

2. Rational Animal Agency

According to Glock, animals are capable of not only acting, but of acting rationally, intentionally, for reasons, and in light of reasons.[6] This section will be dedicated to outlining Glock's account. Briefly, the idea is that animals pursue goals (or desires), while being guided by their descriptive representations of their environment (i.e. beliefs). So, animals act rationally insofar as they act upon personalised goals; intentionally because they act on belief-desire pairs; and they act for reasons because intentional action implies action for reasons. Finally, since facts and reasons are coextensional, animals act in light of reasons because they act in light of facts. Now, let me elaborate.

[4] Glock 2010, 2019.
[5] Glock 2019.
[6] Ibid.

ISSN: 2653-3146

Glock begins by arguing that animals can act. Start with the fact that animals exhibit behaviour.[7] An activity constitutes behaviour if and only if that activity is "aimed at satisfying needs guided by perception".[8] In order to count as action, such behaviour requires two more differentiations. First, the agent needs the ability to decouple stimulus from response – i.e. the same stimulus needs to allow for differing responses. This ensures flexibility and intelligence in behaviour by enabling the subject to learn from past experience and adjust its behaviour accordingly. As such, it is because of this stimulus/response decoupledness that dogs can learn to not attack strangers, or to only 'do their business' outside – same stimuli, differing responses. Second, there needs to be "a *differentiation* of *cognitive* states like belief and *conative* states like desire".[9] This allows the agent to use its descriptive representations of its environment in order to further its goals (or desires).

Were cognitive and conative states not decoupled, the would-be-agent would exhibit mere 'pushmi-pullyu representations'. These representations are in themselves insufficient for action in the relevant sense.[10] They are such that a given belief is necessarily linked to a given desire, and thus necessarily results in activity. Consider this often-cited example: Leopard frogs will (allegedly) snap their tongue at anything that (a) is in their visual field and (b) is a black spot. So, their representation of a black spot immediately results in the snapping of their tongues — this, irrespective of whether the black spot really is a fly, and irrespective of whether they are actually hungry. This is not real action. Action needs consideration, guidedness, and independent motivation.

So, if an agent exhibits flexible behaviour that is guided by its descriptive representations of its environment and directed towards individual goals, then this constitutes action. There certainly are animals that exhibit such behaviour.[11] Consider chimpanzees identifying and using different tools to satisfy their needs and desires (i.e. hunting for ants or termites), or pigs seeking a puddle of mud because (and only if) they want to 'bathe' in it. Hence, certain non-human animals act. Next, let's consider the claim that non-human animals can act *rationally.*

Glock differentiates between "four general conceptions of rationality"[12]:

> According to the first, it is the capacity to maximize satisfaction of one's interests or goals; according to the second, it is responsiveness to reasons; according to the third, it is the ability to reason – draw theoretical and//or practical inferences and to avoid

[7] Ibid., pp. 647-58.
[8] Ibid., p. 649.
[9] Ibid., p. 650.
[10] Cf. Millikan 1995; and Millikan 2005, chapter 9.
[11] See also Call and Tomasello 2008; Glock 2010, 2019; Held et al. 2001; Krupenye et al. 2016, 2017; Krupenye and Call 2019; Millikan 2005, 2006.
[12] Glock 2019, p. 656.

ISSN: 2653-3146

> inconsistencies; according to the fourth, it is the ability to justify one's actions and beliefs to others.[13]

He sets aside the first conception because in order to be subject to rationality's normative constraints, one has to first have the general capacity for rational action. In a slogan, capacity precedes normativity.[14] The third conception amounts to the capacity to reason which seems to presuppose certain linguistic capacities, and is thus bracketed as well. Lastly, Glock rules out the fourth conception since he believes that it begs the question in favour of lingualism because it again presupposes certain linguistic capacities and thus "excludes animals *ab initio*".[15]

It is the second notion (*reasons-responsiveness*) that goes in the right direction, by emphasising a feature that is common to all four conceptions – a feature that rationality shares with intelligence. An animal's intelligent actions are based on their stimulus/response decoupledness. This decoupledness ensures that they are capable of acting flexibly when confronted with familiar as well as novel situations and problems, learning from past solutions, and even transferring and adapting relevant information. These features and capacities also prominently figure in rationality. Glock adheres to an *internalist* interpretation of rationality. Internalism about rationality holds that what it is rational to do is what one *most wants* to do of all the personal goals one has adopted. Such goals need not be very sophisticated. A dog's goals might just involve *chasing the cat* and *chasing the tennis ball* – and she might in fact opt for the latter simply because she has learnt that this activity brings just as much joy while avoiding sanctions from humans. As such, she chooses her goals flexibly, adaptively, and based on past experience. Since some intelligent animals adopt personalised goals in light of their past experience, such animals count as rational given that they can act upon these goals.[16] Thus, "intelligent animals with goals of their own can act rationally".[17]

Notice that this already implies that animals can act *intentionally* and *for* reasons within a certain tradition. If an agent acts in order to achieve a goal of its own, then it acts "*with the intention* of achieving the goal".[18] Now, this can be understood in two ways.[19] In a minimal sense, ascribing an intention is just ascribing the pursuit of a goal. Intentional explanations are explanations that involve the pursuit of goals. In a broadly Davidsonian framework, on the other hand, 'rationalisations' are

---

[13] Ibid.
[14] Cf. Dretske 2006 for a similar argument to that extent. The idea is that I cannot be evaluated as *rational* or *irrational* if I am merely *a-rational*, *pre-rational*, or *non-rational* (and thus do not even have the general capacity for rationality).
[15] Glock 2019, p. 656.
[16] See also Dretske 2006 and Millikan 2006 for other (but similar) accounts of minimal rationality.
[17] Glock 2019, p. 658.
[18] Ibid.
[19] Note that Glock's position is compatible with both.

explanations of actions in terms of beliefs and desires, where beliefs and desires together *constitute* intentions.[20] Hence, such rationalisations are intentional explanations of action. It is this sense of 'intention' that I will from now on adopt. Additionally, rationalisations are taken to explicate the reason(s) for which the agent in question acted. As Glock puts it: "In a perfectly clear and established sense, a goal or intention with which an action is performed counts as a *reason* for which the action is performed".[21] In other words, acting with an intention is sufficient for acting for a reason. Now, recall that *action* itself already implied descriptive representations of one's environment (i.e. beliefs) and the pursuit of goals (or desires). As such, since beliefs and desires constitute intentions, if one acts upon one's goals (or desires), then one acts intentionally and thus also for a reason. As already mentioned, some non-human animals act upon personal goals.[22] Hence, they act intentionally and for a reason. Now, there are two immediate questions to consider next: What we should believe *reasons* to be, and what it means to *act in light of them*.

Glock adopts an *objectivist* position when it comes to reasons.[23] This is just to say that reasons are not mental states, but facts. My reason for taking an umbrella is not my belief that it is raining, but the *fact* that this is so. More specifically, it is neither my act of *believing*, nor the *content* of my belief that counts as a reason – rather, it is the *fact* that it rains that counts. Reasons and facts are coextensional. Hence, one does not have to be aware of one's mental states in order to be aware of the reasons one has for acting. An awareness of one's surroundings suffices – i.e. an awareness of the facts of one's environment. This does not commit one to externalism about rationality because objectivism centres around facts *as perceived by the agent* – it is still perspectival. As such, it does not matter whether it 'actually' rains – it just needs to seem to me that this is in fact so. This notwithstanding, it is not mental states that provide reasons for action, but rather "*what is believed* or *what is desired*," where this refers to facts.[24] As such, when one represents the world, one represents reasons. Animals have representations of their environment (i.e. of the facts surrounding them).[25] Hence, they have representations of reasons. It is in this sense that they also *have* reasons to act. And when they do act, they will act in light of the facts that obtain (or potentially obtain)[26] around them. Hence, animals can act in light of reasons.

[20] Cf. Davidson 1963.
[21] Glock 2019, p. 660.
[22] See footnote 10.
[23] Cf. Alvarez 2010.
[24] Glock 2019, p. 661.
[25] See footnote 10.
[26] Glock believes that one can also act on 'potential facts' in order to explain how one can act for reasons (i.e. facts) that really are not there (cf. Glock 2019, p. 664). Alternatively, one might refer to *apparent reasons* (cf. Alvarez 2010).

ISSN: 2653-3146

This section was dedicated to elucidating Glock's arguments for animals as rational agents capable of acting rationally, intentionally, for reasons, and in light of reasons. Action is behaviour that involves stimulus/response decoupledness, as well as cognitive/conative state decoupledness. Within a certain tradition, cognitive and conative states (i.e. beliefs and desires) together constitute intentions. Hence, the capacity for action implies the capacity for intentional action, which in turn implies the capacity to act for reasons. Finally, since reasons and facts are coextensional, acting in light of reasons is equivalent to acting in light of facts. Since animals certainly act in light of the facts obtaining in their environment (through their representations of these facts), they act in light of reasons. Now, notice that this is not yet to say that animals represent reasons (i.e. facts) *as reasons.* That is, their representations of the facts surrounding them does not yet have to take the form of a reason – it might just be that they represent reasons (i.e. facts) *as facts,* without that involving the concept of a reason. In the next section, I will consider this issue more closely by way of an objection that Glock anticipates – the *Lingualist Objection.*

3. The Lingualist Objection

Glock[27] concludes his paper by considering a possible objection to animals' ability to act in light of reasons. The objection holds that the ability to conceptualise reasons *as reasons* is necessary for acting in light of reasons, which in turn implies certain linguistic capacities. This section will be dedicated to laying out what this *Lingualist Objection* amounts to exactly. Further, I will briefly indicate which premise Glock himself rejects. As I will argue in the next section, however, he could resist the argument on even further grounds. This will be the subject of Section 4.

The *Lingualist Objection* proceeds as follows: One has to be able to *reflect* upon one's reasons in order to act in light of them. The ability to reflect upon reasons, however, requires the ability to think about them *as* reasons – to *conceptualise* them as reasons – where this entails a certain minimal linguistic repertoire. Animals do not possess this linguistic repertoire. Hence, such animals cannot reflect upon reasons *as* reason, which implies that they cannot act in light of them. Here's a roughly formalised version of this argument in order to clarify it:

(P1) $(InLight \rightarrow ReflR)$
(P2) $(ReflR \leftrightarrow ReflRasR)$
(P3) $(ReflRasR \rightarrow ConceptualiseR)$
(P4) $(ConceptualiseR \leftrightarrow LingComp)$
(P5) $(Animal \rightarrow \neg LingComp)$

[27] Glock 2019.

(C1) $(Animal \rightarrow \neg ReflRasR)$ [from P2 – P4]
(C2) $(Animal \rightarrow \neg InLight)$ [from P1 & C1]

Let me elaborate. First, some theorists claim that in order to act in light of reasons, one has to be able to reflect upon them.[28] Intuitively, I cannot be properly said to act in light of reasons if I cannot weigh them, consider them, prioritise some over others, in short, reflect upon them. Second, Glock (*contra* his opponents) holds that reflecting on reasons *as* reasons is both necessary and sufficient for reflecting on them at all. It is certainly sufficient, but it is also necessary since those views suggest that one cannot reflect upon reasons if one cannot recognise or understand them as such. Third, conceptualising reasons *as* reasons is necessary for reflecting upon them. Here's another way to think about this. The lingualist can concede that facts and reasons are *extensionally* equivalent. That is, in enumerating every possible fact, one simultaneously enumerates every possible reason. The problem is that in order for it to be true (*de re*) that animals reflect upon reasons, they have to reflect upon them *as* reasons. Although reasons are extensionally equivalent to facts, it would not be sufficient for them to merely reflect on facts *as facts* precisely because the way in which they conceptualise these facts makes all the difference as to whether it is true (*de re*) that they reflect upon *facts* or upon *reasons*. One would have to reflect on a reason *in the guise of a reason* in order for it to be true (*de re*) that one reflects upon reasons at all (and not just mere facts). In yet other words, one reflects upon reasons rather than facts only if one conceptualises (understands, regards) them *as* reasons and not as mere facts.[29] The lingualist holds, however, that animals which lack even basic linguistic competence fail to live up to this demand.

So, fourth, linguistic competence seems to be both necessary and sufficient for conceptualising reasons as reasons. Linguistic competence should here be taken as that kind of competence we would intuitively ascribe to linguistic animals like ourselves. So, it is *not* about mere communicative competence in the broader sense, but rather about language-related abilities in the strict sense. It is, after all, the *Lingualist* Objection. The reason why Glock holds that such linguistic competence is sufficient for conceptualising reasons as reasons is that small children can recognise reasons as reasons "by answering 'why?' questions by 'because …'." They are able to have a "partial grasp of the concept of a reason" in virtue of their linguistic capabilities.[30] So it is their ability to verbally justify themselves that explains their ability to conceptualise reasons as reasons. Now, why is language necessary for this? First, according to proponents of the Lingualist Objection, language just is necessary for the possession of concepts in the first place – especially those of such an abstract

[28] Cf. Brandom 2010; Davidson 1982; Frankfurt 2004, pp. 18-19; Hacker 2007; McDowell 1996.
[29] Plausibly, this is (at least in part) because the verb "reflecting" is a psychological verb and thus creates an intensional context (cf. Wild 2008, chapter IV).
[30] Glock 2019, p. 669.

nature like the concept of a *reason*.[31] In other words, according to the Lingualist Objection, acting in light of reasons implies the ability to reflect upon them. Reflecting on reasons implies reflecting on them *as* reasons, which requires the ability to conceptualise them. Concept-possession, however, is possible only through language. Hence, non-linguistic animals cannot conceptualise reasons. Second, Glock himself concedes that "reflecting on reasons as reasons comes in degrees, *depending on* the linguistic repertoire".[32] Given the aforementioned implication relations, language is both necessary and sufficient for conceptualising reasons *as* reasons.

In the end, Glock concedes that "*in so far as* reflecting on reasons requires reflecting on them as reasons and therefore mastery of the concept of a reason, it depends on some mastery of the idiom of intentional explanations".[33] He can concede this because he counters the Lingualist Objection by rejecting (P1) – he argues that reflecting on reasons as reasons is not necessary for acting in light of reasons. Briefly, "*A* couldn't develop a capacity to reflect on *A*'s reasons if *A* didn't *have* reasons [to act in light of in the first place]".[34] Demanding that the capacity to reflect upon reasons precede one's having reasons would just result in an infinite regress. Thus, he holds that animals can act in light of reasons despite their inability to conceptualise them as reasons. I would argue, however, that he could do more than this. First, he could also explicitly reject (P4) – the idea that linguistic competences are *necessary* for conceptualising reasons as reasons. As I will argue in the following section, non-linguistic animals can conceptualise reasons as reasons. Second, I believe that, given my arguments, he can also account for more types of rationality in animals than he himself suggests. Recall Glock's four conceptions of rationality from Section 2: (i) maximisation of one's interests, (ii) reasons-responsiveness, (iii) the capacity to reason, and (iv) the capacity to justify one's actions and beliefs. Recall further that Glock sets aside conceptions (i), (iii), and (iv). If what I will argue is correct, then Glock's account also provides some grounds to believe that animals can (iv) justify their actions – just in a non-linguistic manner. Moreover, my arguments will provide a first step towards rethinking whether non-linguistic animals might not actually be able to *reason* – i.e. conception (iii). I will discuss the implications of my additions to Glock's account in the conclusion. Let me now turn to my arguments for why non-linguistic animals can conceptualise reasons as reasons.

4. Conceptualising Reasons Without Language

Relying on Glock's account of concept-possession, I would like to suggest that it is possible for non-linguistic animals to conceptualise reasons *as* reasons.[35] In order to

[31] Cf. Brandom 2010; Davidson 1982; Frankfurt 2004, pp. 18-19; Hacker 2007; McDowell 1996.
[32] Glock 2019, p. 669 (emphasis mine).
[33] Ibid., p. 669.
[34] Ibid.
[35] Cf. Glock 2010. See also Schmidt 2015 for another ability-account of concept-possession.

argue for this, I need to establish two things. First, it needs to be possible that non-linguistic animals can possess concepts. Second, animals need to be capable of conceptualising something as abstract as a *reason*. I will argue for each of these claims in turn. The latter claim will be argued for in two different ways. The first is to present a direct argument for the plausibility of animals' abilities to conceptualise reasons. The second way amounts to an indirect argument for this idea by granting plausibility to a possible necessary condition for conceptualising reasons – namely, the ability to conceptualise intentions.

Non-linguistic animals can have concepts of concrete objects (e.g. tables, rocks, ants, etc.). Glock himself provides an account of how this is possible. He argues that concept-possession is an ability – namely, the ability to *classify* objects. Classification here means "[recognising] *x as* being *F* rather than non-*F* or *G*".[36] This, in turn, involves judgment, where judgment is nothing but "a deliberate response to a question".[37] But wouldn't that require language after all? Not necessarily. For Glock, "questions are in the first instance linguistic fallout from problems", and animals are certainly capable of "facing a problem and deciding in a flexible manner on which options to pursue".[38] As outlined in Section 1, their stimulus/response and belief/desire decoupledness secures exactly this. Hence, since animals can discriminate between options in a deliberate manner, they can judge things to be an *F* rather than a non-*F* or a *G*. This, in turn, is sufficient for their possessing a given concept. Consider the following example given by Glock:

> Take a chimpanzee that has learnt to use different tools in the pursuit of *dorylus* ants and *macrotermes* termites. It is plausible to maintain that it judges its prey to be of one kind rather than another, and similarly for its tools. […] In short, judgement arises out of a capacity for deliberate discrimination in the context of problem-solving, a capacity that we share with some animals that are highly sophisticated without possessing language.[39]

Since the chimpanzee – call her Ruth – can differentiate between termites and ants, it seems that she can recognise termites *as* termites rather than non-termites or ants. That is, Ruth can conceptualise termites *as* termites. Of course, Ruth might not conceptualise them *as termites* in our sense. But she does seem to conceptualise them *as something*. There is a way in which she represents and classifies them. Note that Glock himself seems to limit his argument to concepts about concrete objects. I believe he need not do this. In fact, he can claim that animals can conceptualise

[36] Ibid., p. 30.
[37] Ibid.
[38] Ibid.
[39] Ibid.

something as abstract as a reason.

4.1 The Direct Argument: Conceptualising Reasons as Reasons

Start with a minimal condition of what it takes to regard a fact *as a reason*. Following Scanlon, one could suggest that, minimally, regarding a fact as a reason is regarding it as "a consideration that counts in favour of" a given action.[40] Now, consider this example again: "My reason for taking an umbrella is that it is raining, *not* that I *believe* that it is raining; for it is the weather rather than my own mental state that makes taking an umbrella *good in my eyes*".[41] So, the weather (i.e. a given fact) is what makes a specific course of action appear *good in my eyes.* Call this making something *conatively salient.* Thus, I propose that, in the most minimal possible sense, regarding facts as reasons is regarding them as *sources of conative salience*. It is regarding them as that which makes certain courses of action, objects, states of affairs etc. good or desirable in the agent's eyes.[42] Regarding a fact as a reason, then, is regarding it as that in virtue of which something appears to me *in the guise of the good*.

Which brings me to the final question: Can animals conceptualise reasons as reasons? For this claim to hold, it would need to be true that our chimpanzee – Ruth – conceptualises the termites as a reason for choosing one tool rather than another – as a source of conative salience. Recall that 'conceptualisation' here amounts to deliberately discriminating between "*x as* being *F* rather than non-*F* or *G*".[43] So, Ruth would need to be able to deliberately discriminate between the fact that there are termites as being a source of conative salience rather than not. In other words, she would need to be able to identify the termites as that in virtue of which one specific tool stood out (i.e. became conatively salient) from the others, to the extent that she actually chose that tool. How would one determine whether that was the case? Paradigmatically, by *asking* whether that was the case. As already mentioned, Glock holds that language is sufficient for possessing the concept of a reason precisely because it allows one to communicate one's reasons and answer 'why'-questions.[44] But saying that language is necessary would arguably beg the question in favour of lingualism. We need to consider the possibility that a similar mechanism can take place without one's verbal articulation of it. Now, recall further that Glock holds that "questions are in the first instance linguistic fallout from problems".[45] Ruth faces a problem when she spots the termites – the problem which tool to use. Now, arguably, in picking one tool rather than another, Ruth provides us with an answer

[40] Scanlon 1998, p. 17.
[41] Glock 2019, p. 662 (last emphasis mine).
[42] Of course, this should be taken to mean '*pro tanto* good' and not '*all things considered* good' since there can still be conflicting reasons.
[43] Glock 2010, p. 30.
[44] Glock 2019, p. 669.
[45] Glock 2010, p. 30.

to that problem: she picks that specific tool *because* it is termites she wants for lunch. The termites are that in virtue of which a specific tool appeared conatively salient to Ruth – she regarded the fact that there are termites as a reason for choosing that specific tool. Were another chimpanzee to be confused about her picking this tool rather than another, Ruth could (and plausibly *would*) provide a perfectly useful answer simply by pointing to the termites. This behaviour would testify to her understanding of the termites as being the reason why she picked that specific tool. But what if she never actually exhibits this behaviour? Glock has an answer to this as well: "Concept-possession must belong to the category of potentiality. For unlike concept-exercise, concept-possession is enduring or static rather than episodic or occurrent".[46] Thus, even if Ruth never actually has to provide answers for such queries, we can plausibly ascribe to her the possession of the concept of a reason in virtue of her general capacity to answer such 'why'-problems with 'because'-behaviour.

Notice, incidentally, that this also provides some grounds for the claim that non-linguistic animals can act rationally in another of Glock's senses: (iv) justifying one's actions and beliefs to others. In a perfectly coherent sense, justifying one's action can consist in pointing towards the reason(s) one had for performing it. More specifically, motivating reasons imply explanatory reasons, which in turn imply minimally justificatory reasons. If the fact that it rains motivates me to take my umbrella, this fact also explains my action and minimally justifies it. Already Davidson noticed that motivating reasons also minimally justify one's action (even if not very well).[47] And again, justification need not be a verbal exercise. First, because it would again beg the question in favour of lingualism, and second because "questions are in the first instance linguistic fallout from problems".[48] So, in answering 'why'-problems with 'because'-behaviour, Ruth not only indicates what made this course of action appear good in her eyes, she also explains her action and minimally justifies it. That is, in pointing towards the termites, Ruth provides some justification for her taking the termite-tool. And again, capacity precedes normativity. The capacity to justify *at all* precedes the capacity to justify *well.* The fact that there are *termites* might be a bad justification for taking the *ant*-tool, but it is a justification nonetheless.[49] Hence, if non-linguistic animals exhibit such behaviour, they (minimally) justify their actions and thus at least partially satisfy Glock's fourth conception of rationality. More specifically, they satisfy the practical aspect of this rationality conception. It is, of course, an empirical question whether animals do act in that way. But I take it that findings on social learning in non-human animals

[46] Ibid., p. 27.
[47] Davidson 1963, pp. 690-91.
[48] Glock 2010, p. 30.
[49] Again, cf. Dretske 2006 for similar ideas concerning capacity preceding normativity.

should testify for the plausibility of this claim.[50] Thus, I submit that non-linguistic animals can (minimally) justify their actions.

There is a possible objection lurking, however. Recall that Glock believes that "*in so far as* reflecting on reasons requires reflecting on them as reasons and therefore mastery of the concept of a reason, it depends on some mastery of the idiom of intentional explanations".[51] The objection points out that it is the understanding and use of *intentional explanations* that does most of the work here. After all, how can one conceptualise something as a reason if one does not understand what it means to intentionally pursue a goal? One might thus argue that one can conceptualise reasons as reasons only if one can conceptualise intentions as such. The latter would, for instance, involve the ability to deliberately ascribe intentions to others (or oneself). Hence, one would first have to establish that Ruth can ascribe an intention to herself before one can jump to the conclusion that she can conceptualise the termites as sources of conative salience.

4.2 The Indirect Argument: Conceptualising Intentions

So, suppose the following conditional holds true: If animals can conceptualise reasons as reasons, then they can conceptualise intentions as such – the latter thus being a necessary condition for the former. If what I have just argued is correct, then the antecedent holds. This would already suffice to establish that the consequent holds as well. This notwithstanding, I will now present a further argument for the plausibility of the consequent – for the claim that certain animals can conceptualise intentions according to Glock's account. So even if the critic were correct in pressing this conditional, we have independent reason to believe that animals can satisfy its consequent. This should, in turn, grant further plausibility to the claim that animals can conceptualise reasons as reasons. To argue for this, I will rely on empirical data gathered by Held and colleagues.[52] More empirical evidence for the idea that various animals ascribe intentional states to one another is available.[53] I will rely on Held and colleagues merely for illustrative purposes.[54]

Held and colleagues conducted experiments that suggest that pigs can conceptualise intentions.[55] For this, they let a hierarchically dominant pig (DP) and a hierarchically subordinate pig (SP) forage for food in a specific area that was appropriately

[50] Cf. Krupenye and Call 2019; Seyfarth and Cheney 2015; Wild 2008, pp. 162-79.
[51] Glock 2019, p. 669.
[52] Held et al. 2001.
[53] Cf. Call and Tomasello 2008; Hare 2011; Krupenye and Call 2019; Krupenye et al. 2016, 2017; Legg, Ostojic, and Clayton 2016.
[54] Note that Held et al. 2001 studied pigs. They did this purposely because they wanted to show that it is not just chimpanzees that display such sophisticated mental capacities (cf. p. 210). I take it that this also further strengthens my overall point.
[55] Held et al. 2001.

prepared (by placing food at certain locations, for instance). In a first round, both pigs simply searched for the food. Then they were put back, the food was replaced, and only the SP was allowed to search again. The food was located at the exact same spots. Thus, the SP was trained to believe that the food will be placed at the exact same location as immediately before. The DP remained ignorant about this. In a second round, the experimenters put the food in different locations and let only the SP search for it. When they then let both pigs forage again, the DP had to search for food, while the SP merely had to *relocate* it – it already knew where the food was. Now, what happened was that the DP would very soon give up its search and follow the SP instead. This was a robust result. The best explanation for this is that the DP realised that the SP was *knowledgeable* about the location of the food. But this alone is certainly insufficient to explain the DP's behaviour. The DP also needed to realise that the SP *wanted* to locate the food. Notice, however, that this just *is* what it means to attribute *intentional action* to one's conspecific.

Again, intentional action is action which is "to be explained by reference both to what they want […] and to what they believe […]".[56] Beliefs and desires together constitute intentions. The DP abandoned its search and went along with the SP *because* it realised both that the SP *knew* where the food was and *wanted* to go there. Both attributions are necessary to explain the DP's actions in this way.[57] That is, had it believed that the SP merely *guessed* where the food was, it plausibly would not have abandoned its own search. Similarly, had the DP believed that the SP knew where the food was without *wanting* to find it, it plausibly would not have followed it either. But what this means is that the DP had to effectively discriminate between the SP's being knowledgeable rather than ignorant, and as desiring rather than not desiring the food – it recognised the other pig "*as* being *F* rather than non-*F* or *G*".[58] If this is true, however, the DP displayed "some mastery of the idiom of intentional explanations".[59] This is because attributing such a belief-desire pair *just is* attributing an intention. All of this put differently: the best explanation for why the DP abandoned its own search and followed the SP is because it attributed an appropriate belief and desire (i.e. an intention) to the SP.[60] In doing so, the DP discriminated between the SP's wanting rather than not wanting to find the food and knowing rather than guessing where it was. But cognitive and conative states together constitute *intentions*. According to Glock's own criteria, the DP thus conceptualised the SP's intentions.[61]

---

[56] Glock 2019, p. 660. See also Section 2.
[57] For further evidence that lower-level, non-intentional explanations will not do see Krupenye et al. 2016, 2017, as well as Krupenye and Call 2019.
[58] Glock 2010, p. 30.
[59] Glock 2019, p. 669.
[60] See footnote 56.
[61] Notice that, at the bare minimum, the DP would have to recognize the SP's goals. That is, would it not understand that the SP's goal was the food, it would not have abandoned its own search. Now,

Possibly, if one can conceptualise reasons as reasons, then one can conceptualise intentions as such. Given this entailment relation, I have now presented two separate arguments for the idea that animals can conceptualise reasons as reasons. One argument directly for the truth of the antecedent, implying the consequent, and one for the plausibility of the consequent, providing further reasons to believe in the truth of the antecedent. In other words: if what I have argued is correct, then we have reason to believe that animals can conceptualise reasons as reasons. The idea is that animals are plausibly capable of identifying certain objects as that in virtue of which something became conatively salient to them – as that in virtue of which something appeared *good in their eyes*. Their behaviour would certainly suggest that this was the case. I have also presented an independent argument for the claim that animals are capable of conceptualising intentions. The two arguments work independently from one another, but also support each other. I thus submit that Glock can claim that animals can conceptualise reasons *as* reasons. Since animals are non-linguistic, he can thus reject (P4), as well as (P1). Notice, however, that the above argument is still moot when it comes to (P3). Conceptualising reasons as reasons is necessary for reflecting upon them, but not necessarily sufficient. This notwithstanding, it goes some way towards our getting a better grasp of animals' possible ability to reflect upon the reasons they have.

5. Conclusion

Glock argues that animals are capable of acting rationally and for reasons. He rejects the claim that they can reflect upon their reasons. This is because reflecting upon one's reasons requires one's being able to conceptualise them as reasons. I have argued, however, that Glock can actually claim that animals can conceptualise reasons as reasons. This is because in his picture behaviour fulfils basically the same function as language when it comes to problem-solving and answer-giving. When Ruth picks the termite-tool rather than the ant-tool, she does this *because* she wants termites for lunch. The termites are that in virtue of which the termite-tool became conatively salient to Ruth. Plausibly, she can deliberately discriminate between the termites and non-termites or ants *as* that which made the termite-tool seem *good in her eyes*. But this just *is* what it means to conceptualise reasons *as* reasons. Moreover, empirical research suggests that some animals are capable of deliberately attributing intentions to others. This implies their ability to conceptualise intentions according to Glock's account.[62] Thus, I have presented two independent arguments for the idea that animals can conceptualise reasons as reasons. The first is a direct argument for

recall that intentional explanations might also just imply reference to goals. If this were the case, then there is another sense in which the DP understood the SP's intention – by understanding that it *pursued a goal* rather than a *non-goal* or thinking it merely *wandered around*.

[62] Glock 2010.

this, while the second is indirect in so far as it suggests that a possible necessary condition for conceptualising reasons as reasons is fulfilled by animals.

As such, I take it that my arguments imply three expansions upon Glock's own account. The first is, of course, that in using Glock's account, one can actually claim that non-linguistic animals are capable of conceptualising reasons *as* reasons. The second and third improvements revolve around the different conceptions of rationality that Glock provides. More specifically, I believe that he can account for more than he himself has claimed. As already suggested in Section 4.1, and this is the second expansion, I submit that Glock can account for the practical aspect of the kind of rationality that demands the capacity to justify one's actions and beliefs (iv). This is because reference to one's motivating reasons also involves some minimally justificatory behaviour. The third and final expansion is the following: if what I have said thus far is correct, then I have also provided some initial grounds for the claim that animals have the capacity to *reason* – i.e. Glock's conception (iii). Reasoning, as Glock sketches it, involves the capacity to "draw theoretical and/or practical inferences and to avoid inconsistencies".[63] Now, cashing out practical-inference-drawing in my terms would amount to something like the idea that conative salience can (or should) be transmitted if the objects/actions/etc. stand in the right relation towards one another. Conative salience should, for instance, be transmitted from ends to means – if I value an end, I ought to value its means. But yet again, capacity precedes normativity. So, let's reconsider the case of Ruth. The reason why Ruth wants termites for lunch is, in the first instance, the fact that she is hungry. The fact that she is hungry makes termites conatively salient to Ruth. Now, plausibly, this is also what gives rise to her wanting to use the termite-tool. But this just means that the conative salience attached to the termites in virtue of her being hungry was transmitted to the termite-tool. Just as the termites became conatively salient because of the fact that Ruth is hungry, the termite-tool gained conative salience in virtue of its being a means towards Ruth's end. This, I believe, is a first sketch of what we might want to call a *minimal practical inference* – the transfer of conative salience.[64]

[63] Glock 2019, p. 665.

[64] I would like to thank Gabriel Levc, Marlene Valek, Judith Martens, Leonie Holzner, Julia Kremser the organisers and participants of the ic.SoAP 2021 Graduate Conference, the organisers and participants of the Vienna Forum for Analytic Philosophy Internal Workshop, the UPJA editorial board, and three anonymous referees for their thought-provoking comments, vigorous conversations, and immensely insightful ideas. I would also like to thank Bill Withers for his soothing music.

References

Alvarez, M. (2010). *Kinds of Reasons: An Essay in the Philosophy of Action.* Oxford: Oxford University Press.

Brandom, R. (2010). 'Conceptual Content and Discursive Practice.' *Grazer Philosophische Studien 81*(1), pp. 13-35.

Call, J., and Tomasello M. (2008). 'Does the Chimpanzee Have a Theory of Mind? 30 Years Later.' *Trends in Cognitive Science 12*(5), pp. 187-92.

Davidson, D. (1963). 'Actions, Reasons, and Causes.' *Journal of Philosophy, 60*(23), pp. 685-700.

Davidson, D. (1982). 'Rational Animals.' *Dialectica 36*(4), pp. 317-27.

Davidson, D. (2001). *Subjective, Intersubjective, Objective.* Oxford: Oxford University Press.

Dretske, F. (2006). 'Minimal Rationality.' In *Rational Animals?*, edited by S. Hurley and M. Nudds, pp. 107-16. Oxford: Oxford University Press.

Frankfurt, H. G. (2004). *The Reasons of Love.* New Jersey: Princeton University Press.

Glock, H. J. (2010). 'Can Animals Judge?' *Dialectica 64*(1), pp. 11-33.

Glock, H. J. (2019). 'Agency, Intelligence and Reasons in Animals.' *Philosophy, 94*(4), pp. 645-71.

Hacker, P. M. S. (2007). *Human Nature.* Oxford: Blackwell.

Hampshire, S. (1959). *Thought and Action*. London: Chatto and Windus.

Hare, B. (2011). 'From Hominoid to Hominid Mind: What Changed and Why?' *Annual Review of Anthropology 40*(1), pp. 293-309.

Held, S., Mendl M., Devereux C. L., and Byrne R. W. (2001). 'Studies in Social Cognition: From Primates to Pigs.' *Animal Welfare 10,* pp. 209-17.

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2016). 'Great Apes Anticipate That Other Individuals Will Act According to False Beliefs.' *Science 354*, pp. 110-14.

Krupenye, C., Kano, F., Hirata, S., Call, J., and Tomasello, M. (2017). 'A Test of the Submentalizing Hypothesis: Apes' Performance in a False Belief Task Inanimate control.' *Communicative & Integrative Biology 10*(4), e1343771.

Krupenye, C., and Call, J. (2019). 'Theory of Mind in Animals: Current and Future Directions.' *WIREs Cogn Sci 10*(6), e1503.

ISSN: 2653-3146

Legg, E. W., Ostojic, L., and Clayton, N. S. (2016). 'Caching at a Distance: A Cache Protection Strategy in Eurasian Jays.' *Animal Cognition 19*(4), pp. 753-58.

Marcus, E. (2012). *Rational Causation.* Cambridge: Harvard University Press.

Millikan, R. G. (1995). 'Pushmi-Pullyu Representations.' *Philosophical Perspectives 9*, pp. 185-200.

Millikan, R. G. (2005). *Language: A Biological Model.* Oxford: Oxford University Press.

Millikan, R. G. (2006). 'Styles of Rationality.' In *Rational Animals?*, edited by S. Hurley and M. Nudds, pp. 117-26. Oxford: Oxford University Press.

Scanlon, T. (1998). *What We Owe to Each Other.* Cambridge: Harvard University Press.

Schmidt, E. (2015). "Content, Concepts, Concept Possession." In *Modest Nonconceptualism: Epistemology, Phenomenology, and Content*, pp. 7-30. Dordrecht: Springer.

Seyfarth, R. M., and Cheney, D. L. (2015). "Social Cognition." *Animal Behaviour 105*, pp. 191-202.

Stoecker, R. (2009). 'Why Animals Can't Act.' *Inquiry 52*, pp. 255-71.

Wild, M. (2008). *Tierphilosophie zur Einführung.* Hamburg: Junius Verlag.

ISSN: 2653-3146

Founded in 2019, the *Undergraduate Philosophy Journal of Australasia* (UPJA) is the first undergraduate philosophy journal run by students from the Australasian region. We publish two issues and host two virtual conferences annually: one mid-year and another at the end of the year. Our calls for papers for each issue open roughly in March and September, respectively. We aim to be an inclusive and diverse journal that welcomes submissions on any philosophical topic attempting to make a substantive contribution to contemporary debate. Submissions from women and other members of underrepresented groups in philosophy, including those for whom English is not their first language, are particularly encouraged.