

Volume XIII ◇ **Issue 4**

2022

Logos & Episteme

an international journal
of epistemology

**Romanian Academy
Iasi Branch**



**“Gheorghe Zane” Institute
for Economic and Social
Research**

Founding Editor

Teodor Dima (1939-2019)

Editorial Board

Editor-in-Chief

Eugen Huzum

Executive Editors

Vasile Pleșca
Cătălina-Daniela Răducu

Assistant Editors

Irina Frasin
Bogdan Ștefanachi
Ioan Alexandru Tofan

Web&Graphics

Codrin Dinu Vasiliu
Virgil-Constantin Fătu
Simona-Roxana Ulman

Contact address:

Institutul de Cercetări
Economice și Sociale „Gh.Zane”

Iași, str.T.Codrescu, nr.2, cod 700481

Tel/Fax: 004 0332 408922

Email: logosandepisteme@yahoo.com

<http://logos-and-episteme.acadiasi.ro/>

https://www.pdcnet.org/pdc/bvdb.nsf/journal?openform&journal=pdc_logos-episteme

Advisory Board

Frederick R. Adams
University of Delaware, USA

Scott F. Aikin
Vanderbilt University, USA

Daniel Andler
Université Paris-Sorbonne, Paris IV, France

Panayot Butchvarov
University of Iowa, USA

Mircea Dumitru
Universitatea din București, România

Sanford Goldberg
Northwestern University, Evanston, USA

Alvin I. Goldman
Rutgers, The State University of New Jersey, USA

Susan Haack
University of Miami, USA

Stephen Hetherington
The University of New South Wales, Sydney, Australia

Paul Humphreys
University of Virginia, USA

Jonathan L. Kvanvig
Baylor University, USA

Thierry Martin
Université de Franche-Comté, Besançon, France

Jürgen Mittelstrab
Universität Konstanz, Germany

Christian Möckel
Humboldt-Universität zu Berlin, Germany

Maryvonne Perrot
Université de Bourgogne, Dijon, France

Olga Maria Pombo-Martins
Universidade de Lisboa, Portugal

Duncan Pritchard
University of Edinburgh, United Kingdom

Nicolas Rescher
University of Pittsburgh, USA

Rahman Shahid
Université Lille 3, France

Ernest Sosa
Rutgers University, USA

John F. Symons
University of Texas at El Paso, USA

TABLE OF CONTENTS

RESEARCH ARTICLES

Ryan MILLER, Nonrational Belief Paradoxes as Byzantine Failures.....	343
Rogelio MIRANDA VILCHIS, Improving Conceptual Engineering by Differentiating the Functions of Concepts.....	359
Ragnar VAN DER MERWE, Rational Decision-Making in a Complex World: Towards an Instrumental, Yet Embodied, Account.....	381

DISCUSSION NOTES/ DEBATE

Timothy KIRSCHENHEITER, Objecting to the 'Doesn't Justify the Denial of a Defeater' Theory of Knowledge: A Reply to Feit and Cullison.....	407
Gustavo PICAZO, On The Persistence of Absolute Metaphysics.....	417
Lukas SCHWENGERER, Defending Joint Acceptance Accounts of Group Belief Against the Challenge from Group Lies.....	421
Notes on the Contributors.....	429
Notes to Contributors.....	431
<i>Logos and Episteme</i> . Aims and Scope.....	435

RESEARCH ARTICLES

NONRATIONAL BELIEF PARADOXES AS BYZANTINE FAILURES

Ryan MILLER

ABSTRACT: David Christensen and others argue that Dutch Strategies are more like peer disagreements than Dutch Books, and should not count against agents' conformity to ideal rationality. I review these arguments, then show that Dutch Books, Dutch Strategies, and peer disagreements are only possible in the case of what computer scientists call Byzantine Failures—uncorrected Byzantine Faults which update arbitrary values. Yet such Byzantine Failures make agents equally vulnerable to all three kinds of epistemic inconsistencies, so there is no principled basis for claiming that only avoidance of true Dutch Books characterizes ideally rational agents. Agents without Byzantine Failures can be ideally rational in a very strong sense, but are not normative for humans.

KEYWORDS: Dutch books, Dutch strategies, Reflection, ideal rational agents, Byzantine generals, peer disagreements

1. Consistency Paradoxes for Ideal Rational Agents

1.1 Paradoxes of Rational Requirements

The following characteristics are often taken to characterize an ideally rational agent (Grüne-Yanoff 2007):

- 1.1.1 the agent's preference ordering over her prospects¹ is complete
- 1.1.2 the agent's preference ordering over her prospects is transitive
- 1.1.3 the agent's preference ordering over her prospects is continuous
- 1.1.4 the agent's preference ordering over her prospects is independent of irrelevant alternatives
- 1.2.1 the agent's set of probabilistic beliefs is coherent (they satisfy the Kolmogorov axioms)
- 1.2.2 the agent's set of probabilistic beliefs is complete
- 1.2.3 the agent updates her probabilistic beliefs by conditionalization

Frank Ramsey and Bruno de Finetti discovered a natural way of unifying these perhaps seemingly disparate characteristics through the phenomenon of Dutch Books. In a Dutch Book, a bettor faces a guaranteed loss (regardless of the outcome

¹ The set of prospects at any time is fixed, and each prospect is either a future state of the world which occurs with certainty or a probability distribution over such states.

of any risks hazarded), when making a series of synchronic bets at her fair betting quotient² against a competent bookie who possesses no evidence not also in the possession of the bettor (Vineberg 2016). Characteristics 1.1.1-1.1.4 are the axioms of von Neumann and Morgenstern's Expected Utility Theory (1953), which gives the standard method for assigning value under conditions of risk, and hence for interpreting the notion of a guaranteed loss. Characteristic 1.2.2 ensures that the better actually has a fair betting quotient for all of the bets offered by the bookie. Ramsey (1964) and de Finetti (1964) then show that unless the bettor possesses characteristic 1.2.1, she can face a Dutch Book. While the pragmatic connections among guaranteed losses, optimal bets, and ideal rationality are perhaps tenuous and difficult to define, the possibility of a Dutch Book is nonetheless a plausible illustration of a failure of ideal rationality (Skyrms 1987). When the series of bets is offered diachronically, a guaranteed-loss situation is called a Dutch Strategy,³ which Teller (1973) and Armendt (1980) show results for any agent lacking characteristic 1.2.3. Since Dutch Books and Strategies connect the Expected Utility Theory axioms, the Kolmogorov axioms for probability theory, and Bayesian reasoning—each of which has been enormously fruitful—they seem to have explanatory power for characterizing ideally rational agents. The characteristics they demand can be summed up as “epistemic consistency” (Christensen 1991).

Bas van Fraassen (1984) and Jordan Sobel (1987) show that avoiding Dutch Strategies also justifies another proposed characteristic of ideally rational agents: Reflection. The principle of Reflection demands strong diachronic consistency in judgments, such that “the agent's present subjective probability for proposition A, on the supposition that his subjective probability for this proposition will equal r at some later time, must equal this same number r ” (van Fraassen 1984).⁴ David Christensen (1991) worries that Reflection leads to paradoxes—most seriously a

² A fair betting quotient is the odds at which the bettor is equally willing to take either side of the bet.

³ Skyrms (1993) gives the exact conditions for such a diachronic series.

⁴ In other words, a change in credence requires a change in evidence. Credences of ideally rational agents, like stock prices in efficient markets, must “already reflect the effects of information based both on events that have already occurred and on events which, as of now, the market expects to take place in the future...the full effects of new information on intrinsic values [will] be reflected ‘instantaneously’ in actual prices...[so]...successive price changes in individual securities will be independent...[and]...the future path of the price level of a security is no more predictable than the path of a series of cumulated random numbers” (Fama 1965). In fact, because prediction market prices can be interpreted as credences (Wolfers and Zitzewitz 2006), the theory of efficient markets (where traders get no free lunch) and ideally rational agents (where bookies get no free lunch) have the same constraints.

contradictions with the Kolmogorov axioms in a situation where an agent has a small-but-non-zero credence that she will in the future have credence .95 that she has no credences greater than .90.⁵ Reflection means that an agent who will be irrational in the future must be irrational today, a result Christensen takes as absurd. W.J. Talbott (1991) improves on Christensen's argument in two regards. First, he shows that the general formula for generating Christensen cases is any situation in which an agent expects that she will violate Conditionalization (characteristic 1.2.3). Second, he gives an everyday example in which an agent expects that she will violate Conditionalization without doing anything obviously irrational: all the agent has to do is (1) have credence r about the contents of her breakfast on day T (today) and (2) expect that on day $T+365$ she will have a credence less than r about the contents of her breakfast on day T . We are clearly all ineluctably vulnerable to Dutch Strategies.

1.2 Equivalence of Single-Agent Diachronic Consistency and Two-Agent Synchronic Consistency

Christensen (1991) shows that single-agent Dutch Strategies are equivalent to Double Agent Dutch Books. In a Double Agent Dutch Book, a bookie makes a sure profit on a set of synchronic bets with a pair of bettors whose credences differ. We can easily convert any Dutch Strategy into a Double Agent Dutch Book by simply replacing the future agent in the description with a parallel agent. If the parallel agents' prospects are entangled (e.g. by joint finances), then the bookie's sure gain implies a sure loss for both of them. In a further (unnamed) variation, which Christensen discusses as an inconsistency without actually giving a Dutch Book, the agents' credences need not actually differ as long as one agent believes that they differ. If I am willing to bet 3:1 odds-on that Reflection is a true characterization of all rational agents and also willing to bet 3:1 odds-on that Christensen will bet odds-against this claim, then the bookie makes a sure profit no matter whether Christensen (having come around) prefers 3:1 odds-on for Reflection or (still holding out) 3:1 odds-against Reflection. The bookie's payoffs are given in Table 1 (when she varies her stakes as indicated there).

⁵ Perhaps because a typically reliable informant has informed her that her drink was spiked with the drug LSP which has this unusual psychedelic effect, though in this case the informant erred. Such cases implicate not just Reflection but also deductive closure of justification (Backes 2019).

Table 1

	Reflection is True and Christensen bets 3:1 odds-on	Reflection is True and Christensen bets 3:1 odds-against	Reflection is False and Christensen bets 3:1 odds-on	Reflection is False and Christensen bets 3:1 odds-against
My bet on Reflection (7x stake)	-7	-7	21	21
My bet on Christensen's bet on Reflection (5x stake)	15	-5	5	-15
Christensen's bet on Reflection (5x stake)	-5	15	-5	-5
Total	3	3	21	1

The bookie has developed a Double Agent Dutch Book just by knowing that I think I disagree with Christensen. In a way this is unsurprising: Dutch Books are tests of epistemic consistency, and peer disagreement seems like it can be characterized as group inconsistency.⁶ Christensen, however, stresses that such group inconsistency is not indicative of any failure of ideal rationality in the agents who make up the group—perhaps, for instance, the agents have reasonably differing priors.

1.3 Limitations on Expectations of Consistency in Ideal Rational Agents

Christensen (1991) argues that since Dutch Strategies lead to paradoxes and their structurally-identical Double Agent Dutch Books do not indicate failures of ideal rationality, Dutch Strategies themselves should not be interpreted as constraints on ideally rational agents. This nonetheless comes at a cost for Christensen, since such Dutch Strategies are the leading support for Conditionalization (characteristic 1.2.3) which Christensen accepts. Since Talbott (1991)'s examples show that humans cannot always expect to obey Conditionalization (yet he thinks we ought to be rational and ought-implies-can), he jettisons that principle along with Reflection and Dutch Strategy avoidance in general. Talbott takes it that only Dutch Books and Strategies where the agent is aware of the guaranteed loss constrain rationality, but this renders them fruitless as tests of general epistemic

⁶ This point has been formalized much earlier by Ryder (1981).

consistency. Surely rationality requires more than avoiding explicit guaranteed losses.

Christensen himself later brings pressure from two directions against this approach of relaxing constraints on ideal rationality. First, he treats peer disagreement as a source of epistemic concern for rational agents (Christensen 2000; 2007b). Second, in the presence of irrational beliefs even purely Synchronic Reflection also leads to paradoxes, even though it is supported by a simple single agent Dutch Book (Christensen 2007a). Christensen releases this pressure by weakening the constraints yet further: we shouldn't expect perfect synchronic meta-consistency, either (2007a). The arguments for it aren't a true Dutch Book, Christensen says, because the bookie has *contingent* knowledge that the agent doesn't have—it just happens to be knowledge about the agent's own credences (Christensen 2007a). Credences—whether synchronic or diachronic, first-party or third-party—are just ordinary evidence (Christensen 2007a). Sherrilyn Roush (2009) uses the idea that credences are just ordinary evidence to develop a Re-Cal variant of Conditionalization for rational updating of credences even in the face of first-order Conditionalization failures. Because this method relies on principled distinctions between first-, second-, and higher-order evidence, credences, and Conditionalization, it is of no assistance for resolving cases where the non-rational first-order credences are not governed by higher-order credences and thus subject to revision. Peer disagreement is just a special case of this latter situation: neither of the peers' credences are higher-order with respect to the other, so there is no rational way to resolve the incoherence (Roush 2009).

These arguments naturally lead to a three-fold categorization of epistemic consistency demands: strict constraints on rationality supported by true Dutch Books, broader principles supported by Dutch Strategies that should be used when reality doesn't conspire against us (Vineberg 1997), and cases of pure inconsistency lacking any principled method for resolution. Ideally rational agents should be untroubled by peer disagreement, avoid Dutch Strategies whenever they can do so without paradox, and avoid Dutch Books at all costs. Only vulnerability to true Dutch Books should worry us concerning an agent's characterization as ideally rational.

2. The Byzantine Failure Explanation of Consistency Paradoxes

2.1 Byzantine Generals and Byzantine Failures in Computer Science

The large philosophical literature generating and analyzing the paradoxes that result when supposedly ideal rational agents are confronted with nonrational

beliefs can be understood as instances of what computer scientists call the Byzantine Generals problem. The thought experiment given by Lamport, Shostak, and Pease (1982) runs as follows. A number of generals from Byzantium are encamped around a city they have under siege, each with his own army. They are trying to decide whether to storm the city or retreat until the next campaign season, but face the difficulty that some of their number may be traitorous. The constraints on their decision-making are that all loyal generals must adopt the same plan (lest their forces be scattered and routed) and that plan must be the one that a majority of loyal generals privately think best (lest the traitors control the army's strategic decision-making to their advantage).⁷ Under what conditions can these constraints be met? Given Kenneth May (1952)'s theorem in favor of simple-majority voting for two-candidate ballots, a first instinct is to assume that the constraints are met as long as the super-majority among loyal generals is greater than the number of traitors. The trouble is that in the Byzantine scenario there is no neutral arbiter to count the ballots, and a traitorous general may send different responses to different loyal generals in order to sow disarray.

Lamport et al. (1982) derive three important results from the Byzantine Generals problem. The first is that it is equivalent to the Byzantine Lieutenants problem, wherein all loyal Lieutenant Generals adopt the same plan, and it is the plan ordered by the Field Marshal as long as the Field Marshal is loyal. Hierarchy in place of anonymity provides no assistance if the hierarchy cannot be trusted. The second result is that the problem cannot be solved without $3t + 1$ generals, where t is the number of traitors. The third result is that if traitors can be caught when forging messages (e.g. by enforcing cryptographic signing), then the naïve supermajority solution holds, because each general can report every message he receives to every other general without possibility of deception.

While the canonical form of the Byzantine Generals problem involves malicious actors, Lamport et al. (1982) are clear that it applies just as strongly to ordinary hardware failures which result in different signals being received by different processors. In fact their earlier more rigorous and less didactic paper (Pease, Shostak, and Lamport 1980) mentions only faulty processors and not traitorous generals. Here the constraint is merely that "independent processes"

⁷ One may note a certain analogy to Kenneth Arrow (1950)'s impossibility theorem for converting individual ordinal preferences to community ordinal preferences under conditions of unrestricted domain, non-dictatorship, Pareto efficiency, and independence of irrelevant alternatives. Decision theory has already been analyzed in these terms by Briggs (2010). In the Byzantine Generals case, the domain has been restricted, but the non-dictatorship requirement has been strengthened.

must “arrive at an exact mutual agreement of some kind” (Pease, Shostak, and Lamport 1980). A system which meets this constraint exhibits “interactive consistency” (Pease, Shostak, and Lamport 1980). A faulty processor can play the role of a traitorous general merely by reporting different values to different peer processors. When two processors disagree about the value of an input, this is merely the Lieutenants version of the problem (Lamport, Shostak, and Pease 1982). Further, “processor” means nothing more than a peer agent in a parallel system (Lamport, Shostak, and Pease 1982) or even a subsequent independent state of a single system (Biely and Hutle 2009). Later papers on the Byzantine Generals problem thus often recast it in terms of “Byzantine Faults” which “present different symptoms to different observers” and “Byzantine Failures” in which systems requiring interactive consistency cannot achieve it due to Byzantine Faults (Driscoll et al. 2004). If a Byzantine Fault is detected and corrected, whether by a trusted meta-process or a robust consensus protocol, then it will not result in a Byzantine Failure (Arora and Kulkarni 1998).⁸ Since arbitrary hardware failures lead to arbitrary processing results, any arbitrary hardware failure can easily lead to a Byzantine Fault (Lamport, Shostak, and Pease 1982; Driscoll et al. 2004). This leads Arora and Kulkarni (1998) to simply define Byzantine Faults as those which “corrupt processes permanently⁹ and undetectably¹⁰ such that the corrupted processes execute arbitrarily nondeterministic¹¹ actions.” Such processes will obviously be inconsistent with the correctly-functioning processes. Biely and Hutle (2009) call Byzantine Faults “arbitrary value faults” because the result is that there is no constraint on the output value of the process. Byzantine Faults are the most general model of faults because they do not assume that any degree of detection and correction is possible (Biely and Hutle 2009).

⁸ Kuznets et al. (2019) provide an epistemic logic for checking whether Byzantine Faults can be caught.

⁹ I have left out the complicated discussion of timing in the Byzantine Generals literature because unlike real carbon or silicon agents, Dutch Strategies operate on a turn-based system. Permanent in this context merely means extending beyond the time-out in a real-time system or until the end of the turn in a turn-based system.

¹⁰ Undetectable by the system itself, because if a process detects its own fault, then it will not report it, whereas if a neutral arbiter does so, then that process is no longer a peer. This does not mean that the fault is undetectable in principle by an arbiter outside the system.

¹¹ Arbitrary and nondeterministic not in the strong sense of appealing to irreducible objective chance but in the sense that the result cannot be predicted by knowing the algorithm used by the processor.

2.2 Peer Disagreement Cases as Byzantine Failures

Peer disagreement cases are the most obvious instances of Byzantine Failure in human agents. In the check-splitting case (Christensen 2007b), two peers need to come to consensus about the total bill so that each pays the correct amount. The peers produce inconsistent answers. If each interpreted a smudged line on the bill differently, we have the faulty-input Byzantine Lieutenants problem. Since both know how to perform arithmetic, if one has added incorrectly then it is due to an arbitrary, non-deterministic fault like skipping a line, adding a line twice, failing to carry, etc. The agent did not catch this fault before making her report. There is no detector available (e.g. a trusted third party, or a checksum algorithm). It does not matter whether the error leads to forged responses or not,¹² because there are not enough agents available to perform even the naïve majoritarian consensus protocol. The Byzantine Fault has led to a Byzantine Failure where there is no correct procedure for achieving consensus—the system lacks interactive consistency.

Analysis of the check-splitting case in more traditional terms yields the same result. If both agents stand fast then there is a Double Agent Dutch Book against them—they are epistemically inconsistent. The parties can take each other's credences as evidence and use Conditionalization to update their own credences, but doing so won't generally result in convergence since their priors differ. In fact, it can lead to paradoxical situations where credences cross over (Lang 2014). Meta-methods like Re-Cal won't work because the situation is symmetric (Roush 2009). The parties can merely decide to split the difference, but now they are assuming that both have made errors rather than only one, and that those errors are precisely canceling—a highly unlikely set of events, for which there is no evidence. If that were a rational requirement, then rationality would be anti-truth-conducive. In short, the agents are stuck in a situation of epistemic inconsistency without any generalizable and reliable means of escape.

The other Double Agent Dutch Book cases Christensen (1991) discusses are relevantly similar. He portrays himself as holding a trusted meta-role when he explains his wife's differing meteorological credences by her "pessimism," but unless she accepts him as a checker and corrector of her views rather than an epistemic peer, she has no reason to concede to that judgment. If she fails to concede to his judgment and holds fast to her credences, then a clever bookie can do guaranteed damage to their joint bank account. A narrator who accepts

¹² As Driscoll et al. (2004) make clear for the silicon case, this should not be taken for granted as it often is. If a hardware error can make a person calculating a total read a line incorrectly while doing the sum, could not the same or similar error make a person read the line incorrectly while reporting the results of her calculation?

Christensen's view that she is unduly pessimistic will interpret her pessimism as an arbitrary hardware failure, where she fails to match her credences to the objective chances in accord with Lewis (1980)'s Principal Principle. Since there is a Dutch Strategy available in favor of the Principal Principle (Howson 1992), this serves to identify the agent experiencing the Byzantine Fault to third parties. What it does not do, given the unavailability of both a checker actually trusted by both parties and additional peer parties, is prevent the Byzantine Fault from leading to a Byzantine Failure where the parties exhibit interactive inconsistency.

Peers exhibit unresolvable epistemic inconsistency (vulnerability to a Double Agent Dutch Book) just in case they exhibit interactive inconsistency (Byzantine Failure). When agents exhibit interactive inconsistency, they have no reliable strategy available for achieving consensus, so they will be subject to Double Agent Dutch Books. When agents exhibit unresolvable epistemic inconsistency, they face guaranteed losses through Double Agent Dutch Books which both parties would wish to avoid if they had some reliable strategy available for achieving consensus.

2.3 Dutch Strategy Paradoxes as Byzantine Failures

As Christensen (1991) suggested, there is nothing fundamentally different about single-agent diachronic cases. Any Double Agent Dutch Book can be converted into a Dutch Strategy by merely transferring the properties of the second agent to the first agent at a later time. If we expect time consistency from rational agents then this is a problem, otherwise not.

The same goes for the Byzantine Failure analysis of such cases. If I sum my own restaurant bill twice and get two different answers, I have an interactive inconsistency because the result should be the same and I have no more tools to resolve the failure than in the two-agent synchronic case. The agent who knows he will be unwarrantedly pessimistic in the future can only avoid treating the future self as a peer if the future self can be convinced that he is unduly pessimistic—but if the future self is aware of his pessimism and able to act on that knowledge then he can update using Roush's Re-Cal to escape the problem. If the future self is unconvinced of his own irrationality, then I am stuck treating him as a peer. If I assume that neither of us has experienced a Byzantine Fault, then he must have evidence that I lack and have updated his credences by Conditionalizing, so I should use Reflection to incorporate that information. If I assume that he has experienced a Byzantine Fault then I don't have a long enough time series (treating each temporal snapshot as a peer processor) to avoid Byzantine Failure. If I know

Ryan Miller

that my undue pessimism will wear off, after all, then I can use Reflection to update directly to that post-pessimism correct value and there is no paradox.

Christensen (1991)'s catalog of psychological failures all amount to arbitrary hardware faults. In each case, the agent comes to believe something for some reason other than updating on evidence by Conditionalization, which is the rational algorithm that (as shown by Dutch Strategy) prevents diachronic epistemic inconsistency. In each case, the agent is unable to detect and correct his non-rational update. In each case, the resultant credence is essentially an arbitrary value. While less obvious, this is even true for Talbott (1991)'s forgetting case. When I forget what I had for breakfast, I have to update my credence, and I do not do so by Conditionalization on new evidence. What of Talbott's ought-implies-can argument? In order to have a high credence in my choice of breakfast I need not remember the gestalt of consuming the breakfast—I need only store the credence from when I did remember the gestalt and refuse to update except by Conditionalization on new evidence. Characteristic 1.2.2 stated that ideal rational agents have a complete set of probabilistic beliefs—otherwise they might have no fair betting quotients for bookies to discover, be unwilling to take bets, and hence lack susceptibility to Dutch Books and Strategies not through rational success but rather through inadequacy. The agent with the fewest beliefs would be the most rational. If I have a complete set of probabilistic beliefs, however, then I must have adequate memory to store those, and cannot lose credences by memory pressure. If I lose credences and have to regenerate them from nearby credences (about e.g. what I usually have for breakfast), then I have experienced an arbitrary hardware failure. Surely Talbott is correct that this does not describe the human situation, in which such failures are inevitable, but it fails to do so in a way that is not unique to Dutch Strategies. In the other direction, we should expect arbitrary hardware faults to lead to vulnerability to Dutch Strategies. Memory faults do so, as Talbott showed. Computation faults would lead to incorrect Conditionalization—the only allowed update operation—which also results in a Dutch Strategy.

Christensen is therefore correct that not much separates Double Agent Dutch Book cases and Dutch Strategy cases. Not only are both subject to equivalent betting losses (assuming that consistency is demanded in the Double Agent case by e.g. entangled finances), but both are generated by Byzantine Faults. Both can be avoided by the same degree of enhanced redundancy.

2.4 Dutch Books as Byzantine Failures

Whereas Christensen draws a close analogy between Double Agent Dutch Books and Dutch Strategies, he distinguishes both sharply from true Dutch Books (1991;

2000; 2007a). The latter he considers as genuine constraints on the credences of ideal rational agents. But what kind of irrationality is indicated by susceptibility to a Dutch Book? Brian Weatherson (2005) indicates that susceptibility to mathematical error is a sufficient kind of irrationality to make an agent vulnerable to Dutch Books. Prospects, after all, are probability distributions over payoffs. If you do the math wrong, you can easily find yourself in a Dutch Book.¹³ And why might you do the math wrong? Well, you experienced an input, memory,¹⁴ or calculation error that you didn't detect and correct: a Byzantine Fault. And as in the two-agent synchronic case, in the single-agent synchronic case every Byzantine Fault is trivially a Byzantine Failure. There is no justification for imputing some stronger form of irrationality to agents vulnerable to Dutch Books when math errors are both common and sufficient for such vulnerability. Conversely, every Byzantine Failure will lead to a Dutch Book. If the hardware failure isn't in credences—the arena subjected to a consistency demand by Dutch Books—then it isn't Byzantine. If the failure is in credences, then an arbitrary change to the credence for p , which leaves credences for q , $p \& q$, etc. unaffected, will lead to a Dutch Book. Even explicit Dutch Books, of the type demanded by Talbott (1991), can be accepted in the event of Byzantine Failures: the fault need only erase the memory of the bookie presenting the guaranteed loss before accepting the series of bets.

Peer disagreement cases and Dutch Strategy paradoxes both presume Byzantine Failures. Unless there is an arbitrary value fault, there is no explanation for why the peers disagree or why the supposedly rational agent updates her credences other than by Conditionalization on new evidence. In fact, other human biases and limitations can be assimilated to Conditionalization by varying the payoffs, ensuring that in such non-Byzantine situations no Dutch Book is possible (Williams 2021). In the presence of Byzantine Failures, however, agents cannot guarantee that they will avoid Dutch Books. Agents can only satisfy ideal rationality if they can avoid Byzantine Failures—if, in Susan Vineberg (1997)'s phrasing, the universe declines to conspire against them.

Perhaps Christensen could respond that true Dutch Books test for epistemic consistency of states, rather than consistency of agents. Maybe the Dutch Book can only be offered while the putatively rational agent is in a constant state with

¹³ Weatherson (2004) argues that since Dutch Books only bind when consistency is expected, they do not mandate assigning a credence of 1 to all logical truths. Therefore there's no reason to assume that agents merely have credences rather than calculating them—certainly if humans can be ideal rational agents they would be the sort who sometimes have to calculate their credences.

¹⁴ For the role of memory in deduction see Genot and Jacot (2020).

Ryan Miller

respect to all her credences. Now, however, there can be no talk of bookies eliciting fair betting quotients—they must have direct access to the credences of the agent, and they must perform all the calculations with respect to the agent's preference ordering. The trouble with this approach is that states don't have preferences—agents do. Even more clearly, states do not experience payoffs. There is a reason that ideal rationality is an attribute of agents, rather than states.

3. Conclusion: A Stricter Model of Ideal Rationality

The conclusion is that, if ideal rationality is to mean anything at all, agents experiencing Byzantine Failures cannot count as ideally rational. In the absence of paradoxes generated by such failures, however, we have no reason to reject Dutch Strategy-motivated constraints on rationality. Such Dutch Strategies then provide a path to a stricter model of ideal rationality than that envisioned by Christensen and summarized in characteristics 1.1.1-1.2.3 at the start of this paper. The first characteristic which can be added is David Lewis's Principal Principle, supported by a Dutch Strategy given by Colin Howson (1992). Then, since the Principal Principle is incompatible with contingent priors (Milne 1991), another additional characteristic of ideally rational agents is that their priors will be necessary. Necessary a posteriori truths are discovered by evidence, so their priors would be necessary a priori. The most promising scheme for necessary a priori priors is Indifference (Pettigrew 2016), which assigns the same priors to all agents.¹⁵ Since ideally rational agents' credences are only functions of priors and evidence (Teller 1973; Armendt 1980), in the absence of Byzantine Failures inconsistencies among agents would all be due to different evidence. Then Conditionalization and Reflection have no trouble meeting Christensen (2000)'s demand for impartiality, and no Double Agent Dutch Books are possible against such strictly rational agents.

This is an extremely strict model for ideal rationality. Philosophers who want to take ideal rationality as normative for humans may naturally rebel at such a model.¹⁶ But humans are subject to Byzantine Faults. A model of bounded rationality intended to be normative for humans must show how those faults can be prevented from developing into Byzantine Failures. This will inevitably mean deciding that in certain situations insufficient parallelism is available for any claim to consistency. In other words, there will be situations in which agents with such bounded rationality will not bet. It is irrational to visit a bookie with your partner

¹⁵ Necessary a priori priors combined with diachronic Dutch strategies and Conditionalization also solves the Sleeping Beauty problem (Milano 2022), another point in favor of such a demanding standard.

¹⁶ Though as John Broome (2007) points out, it can be quite difficult to justify such desires.

if you think you have opposing beliefs and a joint checking account, and it is just as irrational to bet when you suspect that you are experiencing a psychological difficulty that impedes your rationality. Nor should we have expected human-like agents to accept bets at some fair betting quotient on all propositions, since human-like agents obviously lack the complete set of probabilistic beliefs necessary to have such quotients. The characteristics of ideal rational agents are closely intertwined, and rejecting some of those characteristics on the strength of arbitrary value faults without considering what the possibility of such faults says about the system as a whole only leads to confusion.¹⁷

References

- Armendt, Brad. 1980. "Is There a Dutch Book Argument for Probability Kinematics?" *Philosophy of Science* 47 (4): 583–88. <https://doi.org/10.1086/288958>.
- Arora, A., and S. S. Kulkarni. 1998. "Detectors and Correctors: A Theory of Fault-Tolerance Components." In *Proceedings of the 18th International Conference on Distributed Computing Systems*, 436–43. <https://doi.org/10.1109/ICDCS.1998.679772>.
- Arrow, Kenneth J. 1950. "A Difficulty in the Concept of Social Welfare." *Journal of Political Economy* 58 (4): 328–46.
- Backes, Marvin. 2019. "A Bitter Pill for Closure." *Synthese* 196 (9): 3773–87. <https://doi.org/10.1007/s11229-017-1620-8>.
- Biely, Martin, and Martin Hutle. 2009. "Consensus When All Processes May Be Byzantine for Some Time." In *Stabilization, Safety, and Security of Distributed Systems*, edited by Rachid Guerraoui and Franck Petit, 120–32. Lecture Notes in Computer Science. Berlin: Springer.
- Briggs, R. 2010. "Decision-Theoretic Paradoxes as Voting Paradoxes." *The Philosophical Review* 119 (1): 1–30.
- Broome, John. 2007. "Is Rationality Normative?" *Disputatio* 2 (23): 161–78. <https://doi.org/10.2478/disp-2007-0008>.
- Christensen, David. 1991. "Clever Bookies and Coherent Beliefs." *The Philosophical Review* 100 (2): 229–47. <https://doi.org/10.2307/2185301>.

¹⁷ I would like to thank Peter Milne and Philip Ebert for reading previous drafts of this paper, and the audiences at Formal Methods and Science in Philosophy IV in Dubrovnik and the philosophy department at Nanyang Technical University in Singapore for their helpful comments—of course all errors remain my own. This paper is dedicated to the memory of Katherine Hawley.

Ryan Miller

- . 2000. “Diachronic Coherence versus Epistemic Impartiality.” *The Philosophical Review* 109 (3): 349–71. <https://doi.org/10.2307/2693694>.
- . 2007a. “Epistemic Self-Respect.” *Proceedings of the Aristotelian Society* 107 (1pt3): 319–37. <https://doi.org/10.1111/j.1467-9264.2007.00224.x>.
- . 2007b. “Epistemology of Disagreement: The Good News.” *The Philosophical Review* 116 (2): 187–217.
- Driscoll, K., B. Hall, M. Paulitsch, P. Zumsteg, and H. Sivencrona. 2004. “The Real Byzantine Generals.” In *Proceedings of the 23rd Digital Avionics Systems Conference*, 2:6.D.4-61. <https://doi.org/10.1109/DASC.2004.1390734>.
- Fama, Eugene F. 1965. “Random Walks in Stock Market Prices.” *Financial Analysts Journal* 21 (5): 55–59.
- Finetti, B. de. 1964. “Foresight: Its Logical Laws, Its Subjective Sources.” In *Studies in Subjective Probability*, edited by Henry E. Kyburg, Jr. and Howard E. Smokler, 1st edition. John Wiley and Sons.
- Fraassen, C. van. 1984. “Belief and the Will.” *The Journal of Philosophy* 81 (5): 235–56. <https://doi.org/10.2307/2026388>.
- Genot, Emmanuel J., and Justine Jacot. 2020. “The Brain Attics: The Strategic Role of Memory in Single and Multi-Agent Inquiry.” *Synthese* 197 (3): 1203–24. <https://doi.org/10.1007/s11229-018-1743-6>.
- Grüne-Yanoff, Till. 2007. “Bounded Rationality.” *Philosophy Compass* 2 (3): 534–63. <https://doi.org/10.1111/j.1747-9991.2007.00074.x>.
- Howson, Colin. 1992. “Dutch Book Arguments and Consistency.” *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association* 1992 (2): 161–68. <https://doi.org/10.1086/psaprocbienmeetp.1992.2.192832>.
- Kuznets, Roman, Laurent Proserpi, Ulrich Schmid, and Krisztina Fruzsá. 2019. “Epistemic Reasoning with Byzantine-Faulty Agents.” In *Frontiers of Combining Systems*, edited by Andreas Herzig and Andrei Popescu, 259–76. Lecture Notes in Computer Science. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-29007-8_15.
- Lampert, Leslie, Robert Shostak, and Marshall Pease. 1982. “The Byzantine Generals Problem.” *ACM Transactions on Programming Languages and Systems* 4 (3): 382–401. <https://doi.org/10.1145/357172.357176>.
- Lang, Patrick. 2014. “Bayesian Epistemology of Disagreement.” M.A. Thesis, Vienna, Austria: University of Vienna. http://othes.univie.ac.at/31638/1/2014-01-15_0749501.pdf.

- Lewis, David. 1980. "A Subjectivist's Guide to Objective Chance." In *Studies in Inductive Logic and Probability, Volume II*, edited by Richard C. Jeffrey, 263–93. Berkeley: University of California Press.
- May, Kenneth O. 1952. "A Set of Independent Necessary and Sufficient Conditions for Simple Majority Decision." *Econometrica* 20 (4): 680–84. <https://doi.org/10.2307/1907651>.
- Milano, Silvia. 2022. "Bayesian Beauty." *Erkenntnis* 87 (2): 657–76. <https://doi.org/10.1007/s10670-019-00212-4>.
- Milne, Peter. 1991. "A Dilemma for Subjective Bayesians? And How to Resolve It." *Philosophical Studies* 62 (3): 307–14. <https://doi.org/10.1007/BF00372396>.
- Neumann, John von, and Oskar Morgenstern. 1953. *Theory of Games and Economic Behavior*. 3rd Edition. Princeton University Press. <https://www.jstor.org/stable/j.ctt1r2gkx>.
- Pease, M., R. Shostak, and L. Lamport. 1980. "Reaching Agreement in the Presence of Faults." *Journal of the Association for Computing Machinery* 27 (2): 228–34. <https://doi.org/10.1145/322186.322188>.
- Pettigrew, Richard. 2016. "Accuracy, Risk, and the Principle of Indifference." *Philosophy and Phenomenological Research* 92 (1): 35–59. <https://doi.org/10.1111/phpr.12097>.
- Ramsey, F.P. 1964. "Truth and Probability." In *Studies in Subjective Probability*, edited by Henry E. Kyburg, Jr. and Howard E. Smokler, 1st edition. John Wiley and Sons.
- Roush, Sherrilyn. 2009. "Second Guessing: A Self-Help Manual." *Episteme* 6 (3): 251–68. <https://doi.org/10.3366/E1742360009000690>.
- Ryder, J. M. 1981. "Consequences of a Simple Extension of the Dutch Book Argument." *The British Journal for the Philosophy of Science* 32 (2): 164–67. <https://doi.org/10.1093/bjps/32.2.164>.
- Skyrms, Brian. 1987. "Coherence." In *Scientific Inquiry in Philosophical Perspective*, edited by Nicholas Rescher, 225–42. Pittsburgh, PA: University of Pittsburgh Press. http://fitelson.org/probability/skyrms_coherence.pdf.
- . 1993. "A Mistake in Dynamic Coherence Arguments?" *Philosophy of Science* 60 (2): 320–28.
- Sobel, Jordan Howard. 1987. "Self-Doubts and Dutch Strategies." *Australasian Journal of Philosophy* 65 (1): 56–81. <https://doi.org/10.1080/00048408712342771>.
- Talbott, W. J. 1991. "Two Principles of Bayesian Epistemology." *Philosophical Studies* 62 (2): 135–50.

Ryan Miller

- Teller, Paul. 1973. "Conditionalization and Observation." *Synthese* 26 (2): 218–58. <https://doi.org/10.1007/BF00873264>.
- Vineberg, Susan. 1997. "Dutch Books, Dutch Strategies and What They Show about Rationality." *Philosophical Studies* 86 (2): 185–201.
- . 2016. "Dutch Book Arguments." In *The Stanford Encyclopedia of Philosophy*, edited by Edward N. Zalta, Spring 2016. Metaphysics Research Lab, Stanford University. <https://plato.stanford.edu/archives/spr2016/entries/dutch-book/>.
- Weatherson, Brian. 2004. "Some Dutch Book Arguments." *Thoughts Arguments and Rants* (blog). September 30, 2004. <http://tar.weatherson.org/2004/09/30/some-dutch-book-arguments/>.
- . 2005. "What Do Dutch Book Arguments Prove." *Thoughts Arguments and Rants* (blog). July 10, 2005. <http://tar.weatherson.org/2005/07/10/what-do-dutch-book-arguments-prove/>.
- Williams, Daniel. 2021. "Epistemic Irrationality in the Bayesian Brain." *The British Journal for the Philosophy of Science* 72 (4): 913–38. <https://doi.org/10.1093/bjps/axz044>.
- Wolfers, Justin, and Eric Zitzewitz. 2006. "Interpreting Prediction Market Prices as Probabilities." National Bureau of Economic Research Working Paper 12200. <https://doi.org/10.3386/w12200>.

IMPROVING CONCEPTUAL ENGINEERING BY DIFFERENTIATING THE FUNCTIONS OF CONCEPTS

Rogelio MIRANDA VILCHIS

ABSTRACT: The leading assumption of this paper is that we can improve the methodology of conceptual engineering if we differentiate between the different functions of our concepts. There is a growing body of research that emphasizes the revisionist virtues of conceptual engineering against the descriptive task of conceptual analysis. Yet, it also has faced severe critiques. Among the difficulties raised are the problems of conceptual identification and continuity. That is why several philosophers are trying to resolve these problems and improve the methodology by calling attention, for example, to the functions that concepts can play. I follow this line of argument and argue that we can increase the chances of success if we also clarify and differentiate them. Identifying and assessing the relationship between functions will help us avoid confusion, inconsistencies, and possible verbal disputes. Doing this not only serves our theoretical and practical purposes but helps us reconsider the potentialities and limits of the conceptual engineering program.

KEYWORDS: conceptual engineering, concepts, functions, emotion concepts, metaphilosophy

1. Introduction

Despite some difficulties found along the way, conceptual engineering has been gaining more and more attention from the philosophical community (Burgess & Plunkett 2013a; 2013b; Cappelen 2018; 2020; Plunkett & Cappelen 2020; Floridi 2011). Frequent worries like the discontinuity problem and the challenge posed by externalism are receiving a lot of different and interesting responses (Bach 2016; 2019; Brigandt 2010; Koch 2021; Riggs 2020; Sawyer 2020a; 2020b). Among them, the functionalist response has gotten considerable attention (Brigandt 2010; Nado 2019; Prinzing 2017; Kelp & Simion 2019), but I believe that this solution brings in another set of difficulties. I present these new problems and argue that we can resolve, or at least clarify, them by differentiating between the functions of our concepts. This differentiation will shed new light on familiar and avoidable confusions, as well as on the more general problems that threaten the methodology

of conceptual engineering. We will get a clearer picture of the place that this methodology has in philosophical theorizing.

In the second section of this paper, I distinguish between conceptual engineering and what Nado calls “functional conceptual engineering.” I show that the latter avoids some of the difficulties that the first faces. It will also be evident that functional conceptual engineering must face a new set of problems. In the third section, I suggest that the best strategy to answer this difficulty is to distinguish the varied functions of our concepts. The most significant distinction is between the representational function and other roles. Finally, the goal of the fourth section is to highlight the implications of this distinction, taking the “theory of constructed emotion” as a case study. I argue that if we do not carry out any distinction between our emotion concepts, we must deal with a host of unnecessary confusions, inconsistencies, and perhaps verbal disputes. On the other hand, making a distinction can prevent these problems, and consequently, improve the conditions to make conceptual engineering a more fruitful enterprise.

2. From Conceptual Engineering to Functional Conceptual Engineering

Conceptual engineering is an increasingly popular way to make sense of part of what philosophers are doing when they philosophize. Unlike the essentially more descriptive goals of conceptual analysis, conceptual engineering is a method that enables us to improve our concepts and, therefore, to make philosophical progress that would be otherwise impossible. Scientists have improved concepts like MASS, GENE, and MENTAL DISORDER; philosophers are trying to do the same with concepts like TRUTH (Chihara 1979; Eklund 2002; Scharp 2013; 2014). Yet, the Strawsonian objection against Carnap’s explication (Strawson 1963) still carries weight. One of the main worries is that revising our concepts seems to imply a change of subject. There is also the problem of defining what a concept is.

There is a lively debate on what concepts are (Margolis & Laurence 2014) and, surprisingly, on whether they even exist (Machery 2009). Given that conceptual engineering is in the business of improving, eliminating, or developing new concepts, the methodology of conceptual engineering seems to be subject to similar worries. However, these criticisms are typically directed to the adequacy of specific characterizations of concepts: whether it is better to characterize concepts as exemplars, prototypes, theories, or a combination of them. A general characterization would not be subject to the same worries. We can avoid them if we employ a general characterization of concepts like “multiple realizable functional kinds” (Isaac 2020).

Improving Conceptual Engineering by Differentiating the Functions of Concepts

We can consider concepts as cognitive tools with diverse functions like the representation of the world, the improvement of social practices, and as a way to stimulate our cognitive capabilities. Although there are influential non-representational views about the concepts employed in science, philosophy, and ordinary life (Blackburn et al. 2013; Rorty 1979; 1990), the predominant view is that we can think about the world because concepts play a representative role (I think that the arguments that I present below could apply to a non-representational take on concepts. But I lack the space to defend that view here). We ordinarily conceive of concepts as representational devices (Plunkett & Cappelen 2020) that track specific features of the world. They do this job, no matter whether they provide an accurate representation or not. When they track real worldly traits, I call them “accurate-representational concepts” (for simplicity, I usually refer to them only with the adverb “accurately” and kindred modifiers). The concepts CHAIR, HEART, and PERSON, for instance, represent certain traits of reality: chairs, hearts, and persons. But concepts can also represent inexistent things like centaurs, story characters (e.g., Sherlock Holmes), and chemical compounds made of exotic elements like XYZ. I call these concepts “merely referential” or “representational” (of course, both kinds of concepts are representational, but only one accurately represents).

Concepts also play a practical, interactive role. We use them to interact with the world and, if successful and with their assistance, we can modify reality. Concepts have the power to change the attitudes and thoughts of individuals and, therefore, their behavior. All that is required is that a substantive portion of the population entertains certain concepts to produce sociopolitical and even economic changes (Hacking 1995, 1999). Certain concepts bring about the social and political roles described in those very concepts. Indeed, believing that there is only one way of economic arrangement will cause that arrangement to exist.

But, even if a broad functional definition along these lines can avoid the problem of the existence of concepts, conceptual engineering still faces several challenges. Among them, we find Strawson’s challenge: Conceptual amelioration seems to imply a change of subject. We face a discontinuity problem. But this difficulty can also be surmounted by a functional characterization of concepts. Besides Isaac (2020), several philosophers have proposed to shift our attention from referents to functions (Brigandt 2010; Nado 2019; Prinzing 2017; Thomasson 2020). As I indicated above, one well-known function of concepts is to accurately represent distinct aspects of reality. Other functions are the encouragement of particular patterns of social behavior, the stimulation of creativity, aesthetic pleasure, etc. Although the reference of concepts usually varies for several reasons

(like semantic drift and conceptual engineering), they preserve their function. The function of a concept can remain the same even if its reference and meaning have changed. It is, then, more advisable to focus our attention on functional conceptual engineering instead of the more usual form of conceptual engineering that is more closely concerned with referents and meanings.¹ The concept FISH, for example, now excludes whales from its reference but its function is the same: to provide a taxonomy of animals (indeed, we can say that the new concept FISH performs this function better than the old one).

This functional approach seems to be a promissory way to meet Strawson's challenge. The problem is that if we do not discriminate between the different functions that a concept could have, we could fall prey to idle theoretical disputes or merely verbal disputes. We must get clear on questions like the following: Does the concept *c* have the function of representing some aspect of reality? Does it stimulate a particular kind of behavior? Is it intended just for entertainment?

Consider the concepts that appear in fairy tales and myths. Surely all, or most of them, lack a referent, so they cannot be employed to represent some aspect of reality. Concepts such as MINOTAUR or CYCLOPS continue to play a recreational function. These concepts are also useful to stimulate our imagination, and consequently, to foster our creativity. They could be extremely helpful for storytellers, moviemakers, and fiction writers. Other concepts like FREE MARKET, SUPPLY, and DEMAND have produced a new set of economic and social practices.

From these examples, we can see that concepts can play many different functions, but we must ask ourselves if this is an advantage or a disguised obstacle for our engineering goals. This situation is potentially dangerous because we can mistake one function for another. This is not a problem peculiar to philosophy but can arise in every theoretical domain. If we employ, for example, the term *minotaur* with the clear intention to describe something in the world, we would not achieve our goal. There is not anything in the world that can be referred to by this expression. It cannot play an accurate-representational function. Analogously, other concepts can play a practical role, but they might not perform an accurate-representational role. We can employ the concept MONEY and still hold genuine worries about the existence of the reference of this concept (Barrett 2017, 133-4;

¹ According to Prinzing (2017), concepts are functional kinds that are preserved through conceptual revision of, for example, extension. Nado (2019) rightly points out that it is very troublesome to define what a function is. I remain neutral on what exactly it is. After all, we do not need a precise definition to insist, as I do in the next sections, that there are different kinds of functions.

Goldstein 2020). We can mistake the accurate-representational function of a concept with its aesthetical, recreational, or social role (undoubtedly, we can entertain concepts with the single purpose of playing a game: a concept-guessing game).

We can see now that, despite all the virtues that functional conceptual analysis may bring about, we need to address some problems. In section fourth, I analyze these and other difficulties in more detail and show how we can avoid and lessen some of the potential harmful effects. But first, we need to distinguish the different kinds of conceptual functions.

3. Differentiating the Diverse Functions of Concepts

In the previous section, we have seen that conceptual engineers can overcome the discontinuity problem if they focus on the functions of concepts. This shift of focus is very promising, but we have seen that we face a new challenge: the potential confusion and inconsistency brought about by distinct and conflicting functions or by the failure to recognize the proper function of a given concept. In this section, I suggest that the obvious way to overcome these problems is by differentiating as far as possible the different kinds of functions that concepts play.²

Among other goals, philosophy and science are theoretical enterprises that try to describe reality accurately. Their sentences must be *true* or *approximately true* to achieve this goal. They must *accurately represent the world*. Like scientists, many philosophers seek to accurately describe what the world is like (Eklund 2014, 295). However, we saw that we can make one valuable distinction between the following two functions: representation and accurate-representation. Many concepts that do not represent the world accurately (like MINOTAUR) are, nonetheless, accomplishing their representational function (representing a minotaur. It does not matter that it is a fictional entity that exists only in people's minds). The accurate-representational function can be part of the representational function but not vice versa. In any case, as we will see, the primary function of concepts is to represent (I leave aside the problem of whether a representing function variant can work in non-correspondence theories of truth). They may represent non-existent things, but they owe their existence as concepts to this representational function. All other functions derive from this primordial role.³

² The differentiation may not be completely clear, but we must try to do the best we can even if "it is unlikely that description and prescription can be clearly separated" (Griffiths 2002, 908).

³ I think that this emphasis on the representational function – instead of the accurate-representational one – is valuable because it allows us to avoid skepticism about truth. Insisting on the representational role is weaker (and hence more useful for the purposes of this paper)

At this point, someone (e.g., Nado 2019) may object that – concerning the accurate-representational function – only sentences can be true or false. But this is not problematic. Although isolated concepts cannot have a truth value, they can represent a feature of the world. We say that they are adequate concepts if they accurately represent what is in the world and, because of that property, they help to make sentences true. And that is all we need to maintain that concepts have an accurate-representational function that is not only very important but that, most of the time, grounds the possibility of many other functions. But what are those other functions?

I do not attempt to offer a sharp and complete taxonomy of concepts' functions because there are many, and some might not have been invented yet. A rough characterization is sufficient for the present purposes. Based on the mere representational function, we can find the following ones: 1) the accurate-representational; 2) the epistemic; 3) the sociopolitical; 4) the moral; 5) the emotionally-and-cognitively-stimulating. This list is not exhaustive, but it is sufficient for clarifying the distinct functions of concepts and their relationships. We have already discussed the merely representational function and the accurate-representational one. Let us examine the remaining four.

For “epistemic function,” I mean the function that involves cognitive processes like abstraction (see, e.g., Cartwright 1989, chapter 5), idealization (see, e.g., McMullin 1985), and in general, the result of the modification of concepts that offer us a more simple and manageable representation of reality (they still play the mere representational function). The resulting concepts now have the epistemic functions of representing in more general, simple, and unifying ways. These entail some departure from the accurate-representational function, but the result is highly beneficial for beings with limited computational capacities like us.⁴ Nonetheless, most epistemic functions overlap with the accurate-representational function. We can appreciate this in our scientific and philosophical concepts. But these functions need not overlap. We can have simple and general concepts that lack a referent in the world (at least in the actual one). These concepts could be describing an alternative reality with a set of wholly different laws of nature. Perhaps they refer to fictional entities and characters.

than demanding true sentences and accurate concepts. For a defense of the latter view, see Simion (2018).

⁴ Distinguishing the epistemic function from the accurate-representational function has proved to be extremely difficult (Brigandt (2010) seems to conflate both. See Giere (2006) for a helpful perspective on these issues). But there is, intuitively, a worth distinction to make here.

Improving Conceptual Engineering by Differentiating the Functions of Concepts

We do not have to go very far to appreciate the potential lack of overlapping between the epistemic and the accurate-representational functions. Many scientific and philosophical concepts do not accurately represent the world because their epistemic function will be compromised if they do. A general and simple concept must idealize and abstract away many real worldly features. Consider the concepts FRICTIONLESS PLANE, POINT-PARTICLE, ISOLATED SYSTEMS, or MARKETS IN PERFECT EQUILIBRIUM. Yet, these concepts are very useful for understanding the world.

Concepts play sociopolitical functions too. Most of our concepts about what kinds of behavior, attitudes, and thoughts a person must have, shape sociopolitical reality. Having a concept representing a specific political structure reinforces a particular political behavior. The same is true of legal, social, and economic interactions. If a person believes that some juridical law is right, he will obey it because he recognizes it as true. When a person thinks that his concept of, for example, LAW is appropriate, then he will exhibit a certain kind of behavior that reflects his understanding of this concept towards other human beings.

The sociopolitical function can overlap with accurate-representational and epistemic ones (the diverse functions of concepts can overlap but need not). Instances of the concept DISEASE represent real sets of disorders in organic structures and faculties. These concepts do not merely represent the world but serve to shape it and shape the social and cultural relationships between individuals. DISEASE also plays the epistemic function of being an abstract summary of countless disease instances that allows theory construction and smooth communication.

The status of moral concepts is different. There is an intense debate about the accurate-representational or the mere-representational functions of concepts like GENEROUS. Cognitivists claim that moral concepts play an accurate-representational function, non-cognitivists deny it. One can argue that, although moral concepts do not play an accurate-representational function, they play a mere representational one because people bring certain images of properties to their mind when they think about generous persons. But it need not concern us whether non-cognitivists are right. The important point now is that it is possible to have a concept with a moral function without having an epistemic or accurate-representational function.

In the same way, a moral concept need not have any emotionally-and-cognitively-stimulating function as does the concept INFINITY. A concept can promote peace and respect without having to arouse curiosity, feelings of joy, sadness, surprise, or foster the incubation of new ideas (of course, this concept can

play all these emotionally-and-cognitively-stimulant functions but need not). Furthermore, a concept can have some of these functions without trying to accurately describe reality (accurate-representational function) to be epistemically useful (epistemic function) or to play any social and moral function.

INFINITY plays a cognitively stimulating role, promoting in some individuals the construction of new philosophical theories that are relevant to assess the status of other concepts like CAUSALITY and NUMBER. SADNESS or ANXIETY can have an emotionally therapeutic function in people if we re-engineer them as concepts that refer to emotive states that do not necessarily arise in situations of adversity and pain (see section 4).

After this rough characterization of the functions of concepts, I want to point out some of the most important relations between them and concepts. They can relate at least in the following ways: a) a function lacks a concept; b) a concept lacks a function; c) one function has two or more concepts; d) one concept has two or more functions, and e) a concept has only one function.

In the first relation, a function can lack a concept. In that case, we can devise new concepts that fulfill that function. If we need a concept that can describe a feature of reality, we can design it to provide us with a better description of the latter. Perhaps we would not require a concept that accurately represents reality but one that plays an epistemic role and can help us handle the large amount of information that reaches our senses. Concepts like FRICTIONLESS PLANE, POINT-PARTICLE, and ISOLATED SYSTEMS are doing this job. They abstract from the specific characteristics of frictionless planes, point-particles, and isolated systems. Functions like these offer us a perfect opportunity to engineer new concepts from scratch.

Some concepts lack a function (b). This situation may mean the inexistence of any function whatsoever or the lack of a second one. But there are no concepts without a function because they (as I stated above) at least merely represent (it is hard to think of a concept that lacks any representational property whatsoever). Concepts can lack a second function when, for instance, a concept has an accurate-representational function but lacks a sociopolitical, moral, or epistemic one.

A function can have two or more concepts (c), therefore producing a redundancy. It may not be problematic, but in some cases, we may also want this functions to be simpler and then it is better to eliminate one or more of the concepts. It can also happen that a concept can have two or more functions (d). DISEASE has the function of accurately describing physiological reality, and at the same time, it may have a social role.

Finally, although (e) is very controversial, there may be concepts that only have one function: They, accurately or not, merely represent. These concepts elicit some representation in our minds. Even those concepts that depict things that do not exist have a representational function.⁵

The distinctions just made are of great importance to avoid confusion, inconsistencies, and other difficulties when we engineer concepts. Making distinctions can be very useful to design concepts purposefully. Very likely, concepts designed with a clear purpose can more easily fulfill it than those designed to achieve vague and tangled goals. In the same way, a clear distinction between functions can pave the way to understand and revise existing concepts. We will cover this topic in the next section.

4. The Importance of Differentiating the Diverse Functions of Concepts: Emotion Concepts, a Case Study

In this section, I introduce some of the difficulties that arise in the absence of a clear distinction between the different functions of concepts and why it is crucial that we differentiate them. I argue that clarifying and differentiating the conceptual functions can help us understand and implement them when doing conceptual engineering. Let us begin with an analysis of some of the risks that arise when we do not differentiate between functions.

Lisa F. Barrett (2016; 2017) argues that emotion concepts like ANGER and FEAR are not natural kinds but constructions of mind and culture. Her revolutionary “theory of constructed emotion” challenges the classic theory of emotion that posits a set of basic universal emotive states: anger, fear, sadness, disgust, surprise, and happiness. She denies any biological and psychological basis that determines emotions. Emotions are real as other human constructions like money but are not determined by discrete sets of firing neurons nor by specific changes in the autonomic nerve system, groups of facial expressions, or body movements. On the contrary, the social constructivist hypothesis says that we construct them on the spot and “love (or curiosity, hunger, etc.) is an emotion as long as people agree that its instances serve the functions of an emotion” (Barrett 2017, 138).

According to the theory of constructed emotion, there are not emotive natural kinds but internal sensations that we individuate as instances of specific

⁵ Indeed, concepts like SQUARED-CIRCLE probably also have a representational function. A case can be made that they arouse curiosity, creativity, and distinctive emotional states. I am not going to settle the issue of the existence of concepts with only one function here. But if there is any, this seems to be a merely representational one.

emotions. In a process called “interoception,” the interoceptive network (composed of body-budgeting regions and the primary interoceptive cortex) represents all the sensations from our internal organs, tissues, hormones, immune system, etc. The name for this set of sensations is “affect” and is composed of two parameters: valence and arousal. Valence is the quantity of pleasure or displeasure. Arousal is the low or high level of arousal. With allostasis in view (the regulation of our body for growth, survival, and reproduction), the visceromotor regions of the interoceptive network inform us about the diverse combinations of high and low arousal, pleasure, and displeasure. But this information does not strike us as pure internal sensation. Concepts mediate it.

The body-budgeting part of the interoceptive network employs our conceptually organized past experiences about situations, events, persons, angles, places, times, and feelings involved in anger or sadness episodes and makes multiple predictions (applies and constructs concepts) about what is happening in the world in a given situation before any sensory input reaches our senses. Outside conscious awareness, from the conceptually organized past experiences and the clues in the immediate environment, a set of firing neurons unpacks our conceptually organized experiences about, say, anger into multiple fine-grained concepts about angles, places, times, and feelings. Then the neurons of the interoceptive cortex test whether these concepts fit the actual affective sensory input. Part of the fine-grained concepts is discarded because they do not conform with anger sensations of high arousal and displeasure, in which case the brain constructs new predictions, and the process begins again. But the “confirmed” set, as it usually happens, is assembled into a concept that represents the emotion of anger (indeed, representation is essential for Barrett’s theory. Concepts play the accurate-representational function regarding events and objects because concepts are “populations of representations that correspond to those events or objects” (Barrett 2016, 10)). The fine-grained concepts that compose the set are assembled into the concept ANGER for this specific situation, and we experience anger.

This explanation suggests that “emotional experiences have no objective fingerprints in the face, body, or brain that would enable us to compute an answer” (Barrett 2017, 107). Barrett has conducted many experiments that support the theory, and many other studies challenge the classical view of emotions as universal traits of human nature and stress the variability across cultures. In one study, participants had to identify emotions from three sources: reading about scenarios that usually enact a particular emotion, seeing photos of facial expressions, or both. She found that those who only read the scenario or read it and saw the face correctly predicted the emotion sixty-six percent of the time.

Those that only saw the facial expression got it right thirty-eight percent of the time. Tassinary et al. (1992) used a technique called electromyography (EMG) and did not find that participants displayed the same facial expression when they felt the same emotion. The electrodes on the surface of the skin detected the electrical signals that make the muscles responsible for the facial expressions, but people did not move the same facial muscles in the same pattern when experiencing a given emotion. There is evidence of “degeneracy” or, in philosophical terms, “indetermination” of emotive states by affective states (Wilson-Mendenhall et al. 2011; 2013) or populations of neurons (Whitacre & Bender 2010). Human emotions are compatible with different affective states. Besides, some meta-analyses show that numerous studies have not found the “fingerprint” of emotion in changes in the autonomic nervous system, specific sets of neurons, facial expressions, or body movements (Lindquist et al. 2012). Other studies support the hypothesis that emotion is constrained by culture (Barrett, Mesquita, & Gendron 2011; Dixon 2008; Frijda & Mesquita 1995; Harré 1986; Mesquita & Frijda 1992; Russell 1991; Williams 1977).⁶

If the theory of constructed emotion is right, then we are the architects of our own emotional experiences. As individuals and as societies, we have the power to create our emotions because we decide what configurations of persons and situations bring about particular emotive states. If we, as individuals, look at persons and situations from different sociopolitical, moral, and emotionally-and-cognitively-stimulating perspectives, we can change our emotions. We can change them even more if we further change how our society looks at persons and situations. We can transform or construct, that is, we can re-engineer existing concepts or engineer new ones for specific purposes like treating arachnophobia and anxiety, fostering positive emotions like happiness and positive habits like doing more exercise, or using anger to perform better at sports events.

The constructivist theory fits nicely with the conceptual engineering program because emotion concepts are not restricted by their biological and psychological foundations: “Emotion concepts are goal-based concepts” (Barrett 2017, 92). There is not a biological glue that holds together sensations; our individual and cultural goals do it. Depending on the goal, we can construct a feeling of displeasure and high arousal as fear, anger, or intense sadness. We can create the sensations of pleasure and high arousal as anger or excitement. One can also engineer new emotion concepts by combining existing ones.

⁶ Japanese use the word “itoshii” for a feeling of longing for an absent loved one and Bengali use “obhiman” for sorrow caused by the insensibility of a loved one (Russell 1991, 426), for example.

According to Barrett (2017), we can stop categorizing pain as exhaustion and then exercise more or deconstruct the concept ANXIETY and re-engineer it as the concept EXCITEMENT to treat anxiety. She argues that “people who recategorize anxiety as excitement show similar effects with better performance and fewer classic symptoms of anxiety when speaking in public and even when singing karaoke” (2017, 189). Categorization allows an affective sensation “to become an emotional experience such as happiness or fear, giving it additional meaning and functions understood within your culture. Categorization bestows new functions on biological signals” (2017, 126). With practice, Barrett says,

You can dissolve anxiety into a fast-beating heart. Once you can deconstruct into physical sensations, then you can recategorize them in some other way, using your rich set of concepts. Perhaps that pounding in your chest is not anxiety but anticipation, or even excitement. (2017, 188)

The theory implies that we can engineer and re-engineer emotion concepts from our basic affective states. Engineering “bestows new functions on biological signals, not by virtue of their physical nature but by virtue of your knowledge and the context around you in the world” (Barrett 2017, 126). But can we really engineer emotion concepts at will? This theory faces some difficulties.

First, the theory of constructed emotion is not uncontroversially confirmed by the data. In the same way that many studies support it, there are also multiple studies and meta-analyses confirming the existence of emotion fingertips in the autonomic nervous system, facial expressions, and neural patterns (Elfenbein & Ambady 2002; Norenzayan & Heine 2005; Phan et al. 2002; Vytal & Hamman 2010)

Second, if emotions are constructed at will, then the logical consequence is that it is highly probable that there are human communities that cannot communicate their emotive states because there is no need for two dissimilar cultures to share any emotive concept. But there is no evidence of any culture whose emotion concepts are not partially shared with the researcher or with other cultures. Translation would be impossible in such a relativistic framework, but we translate emotion words from different cultures. Communication is not total and translation is not exact, but there must be something to emotions that explains why intercultural communication and translation exceed chance. Barrett (2017, 38) suggests that “emotions are not inborn, and if they are universal, it’s due to shared concepts,” but the theory of constructed emotion cannot explain why we share them. On the other hand, the classical theory that posits a universal emotional biological basis across cultures can easily explain why we tend to share the same categories.

Finally, the theory of constructed emotions posits the existence of four fundamental affective states or dimensions that are not exclusive of humans but occur in all species: pleasure, displeasure, high and low arousal (Barrett 2016, 17).⁷ Indeed, arousal and valence are practically impossible to separate because if one changes, the other also changes (Barrett & Bliss-Moreau 2009) and scientists have found that valence is common to all cultures (Farroni 2007; Wierzbicka 1992). We can accept the idea that these elementary states do not determine emotion concepts, but it is compatible with affective states *partially* determining them.⁸ It is compatible with the fact that we share a set of *structural ranges* that underlie the possibility of translations and “they occur over a very wide range of cultures, and probably is universal, even if no words exist for certain of the structures in a particular language” (Frijda et al. 1995, 124).

“Negative” emotions like anger, fear, and sadness are displeasing, and “positive” emotions like happiness and love are pleasing. Unless we allow certain twists of meaning in our usual understanding of the basic emotions, we cannot say that fear is an emotion concept that can be constructed from pleasing and low arousal affective states. Anger cannot be an emotion concept that arises from low arousal and pleasing states, and sadness is not a category that rests on a high arousal and pleasing affective basis because they share the displeasing valence.⁹ In other words, the concept FEAR tracks an affective feeling that is displeasing and provokes high arousal. The properties of displeasure and high arousal do not determine the concept FEAR, and there is room for re-engineering it. But the partial determination partially restricts the engineering activity. On pain of inconsistency, confusion, and possible verbal disputes, the accurate-representational function of FEAR represents, though imperfectly, a region in the affective state that is characterized by high arousal and displeasure. It does not matter that we do not perfectly identify the corresponding area. Perhaps no region in nature has precise boundaries, but it does nothing to deter us from identifying gold, quarks, and atoms.

It is true that the neurological and physiological facts by themselves cannot determine the emotional facts. Emotive states are indetermined by the most,

⁷ Barrett et al. (2011) even describe how a patient with semantic dementia categorizes emotions as “positive” and “negative.”

⁸ Griffiths (2002) advances a similar view. According to Griffiths, emotion words “partially refer,” for example, to affect programs and to socially sustained practices.

⁹ As the Ilongot word “liget” demonstrates in Russell’s paper (1991), there can be overlapping between, for example, anger and sadness because they share a displeasing valence, but that does not mean that they are the same.

ontologically speaking, basic neurological and physiological ones, but this is perfectly compatible with partial determination. Emotion concepts fulfill their accurate-representational function when partially referring to specific affective regions even when they are blurry because of the varying set of individual and cultural practices.¹⁰

The accurate representational function does not depend on the individual or on culture. The world determines it. Yet, we can engineer what is not determined by the world to achieve other functions besides the accurate-representational. As long as these functions do not conflict with each other, all is fine. Besides playing the accurate-representational function of a region of high arousal and low pleasure in the affective state, redefining FEAR so that certain situations like speaking in public and meeting new people are culturally understood as non-threatening events plays the social function of diminishing individuals' fear states. A given concept can unproblematically play the accurate-representational and social, recreative, or moral functions at the same time. But is it always like that? I do not think so.

Concepts like CHAIR, GOLD, and ANIMAL purport to refer, primarily, to some features of reality: chairs, gold, and animals. This function is independent of its potential sociopolitical, aesthetical, and recreational uses. We can use GOLD to promote a specific social behavior. If, for some strange reason, we want people to behave towards chairs and animals as they do with gold, we can re-engineer the concept of GOLD to refer not only to gold but to chairs and animals. But we would have completely destroyed its accurate-representational function of representing objects that are yellow, shining, and with atomic number 79. In the same way, if some robots acquire enough external and behavioral similarity with animals, we may also want to preserve the meaning of ANIMAL as referring to organisms composed of cells and re-engineer (expand its extension) it to refer to robots. But then we would end up with a defective inconsistent concept because its accurate-representational function still includes referring to organisms composed of cells but now also includes robots in its extension. The deviation from the original concepts would be so great that the concept ANIMAL would be a new concept competing with its previous version, and it will lose. If we want to re-engineer or engineer

¹⁰ It is true that Barrett (2012) acknowledges the contribution of biology to emotion construction. The problem is that a concession like that means that we cannot construct concepts at will and on the spot. The theory of constructed emotion must be amended to accommodate the partial determination of the biological substrate.

concepts to fulfill other functions, we must differentiate the diverse roles that they can consistently play.¹¹

Thus, we can distinguish two functions in Barrett's ameliorated concept of ANXIETY as excitement. One function is the accurate-representational of referring to a high arousal and displeasing affective state, and the other is a therapeutic and performance improving one (that are instances of the emotionally-and-cognitively-stimulating function). If this re-engineered concept is successful, individuals and society alike will improve their mental health and performance in various ways. But ANXIETY faces the same problems that GOLD and ANIMAL.

The re-engineering of ANXIETY as excitement conflates two distinct but conflicting functions. If ANXIETY means that one feels excitement, then one experiences an affective state of high arousal and pleasure, but the accurate-representational function of ANXIETY depends on referring to an affective state of high arousal and displeasure. They are incompatible.¹² Perhaps ANXIETY refers to a sensation of pleasure and displeasure at the same time, but this new concept also conflicts with our original English concepts of ANXIETY and EXCITEMENT that pick up unpleasing and pleasing states, correspondingly. And unless we defend the existence of inconsistent concepts, we must fix the conflict.¹³ We can hold back our re-engineering pretenses, add a new compatible therapeutic function to ANXIETY, or engineer a different concept.

The lack of attention to the functions we assign to concepts engenders inconsistencies and confusion. If a patient uses EXCITEMENT as an instance of the feeling of anxiety can confuse psychologists and vice versa. If a group of psychologists that adhere to the theory of constructed emotion employ ANXIETY as a concept that picks up pleasing affective states, it will engender a verbal dispute with psychologists fond of the classical theory. One of the main sources of these difficulties is the over-optimistic idea that we can construe emotions and that "we can impose functions that would not otherwise exist, thereby inventing reality" (Barrett 2017, 135). This idea obscures the accurate-representational function that many concepts possess and how it can conflict with other "imposed functions." If ANXIETY accurately describes a feature of reality, the high arousal and displeasing

¹¹ Wishful thinking, for example, can promote better health; "nevertheless, practical reasons are not good reasons for belief: wishful thinking is bad believing" (Simion 2018, 95).

¹² We can also think of future incompatibility. Whether a compatible relationship between two or more functions continues or not is an empirical question. If, for example, future scientific research reveals that the accurate-representational function of a concept conflicts with others, then we would need to fix the problem.

¹³ We do conceptual revision with amelioration in view, not inconsistency or "perversion" (Marques 2020).

affective state, then we cannot invent new anxiety concepts but only modify them to the extent that the slack between the world and concepts allows. Understanding this can aid us to design emotion concepts and concepts overall with clearer functions, and then more easily fulfill goals like treating anxiety, avoiding negative emotions, increasing positive ones, and using them to improve performance in sports and life. We can construe our concepts, but we lack complete engineering control over them and what we want them to be.

In summary, when we do conceptual engineering, it will pay off to differentiate the distinct functions of our concepts as far as possible. This distinction will be useful to improve the understanding of our current concepts, foresee the possible ways to ameliorate them, avoid inconsistencies, confusion, and possible verbal disputes.

5. Conclusion

We have seen that functional-conceptual-engineering is a good candidate for overcoming the discontinuity problem. We have also noted that this way to understand conceptual engineering must face other difficulties. I pointed out that one way to avoid these obstacles is to differentiate between the different kinds of functions of concepts, as well as the multiple relationships between them.

We can engineer or re-engineer most of our concepts, but there are limits to this activity. If we do not examine the functions that we are adding to our concepts, we can end up with conceptual and functional inconsistency. We must be aware that adding sociopolitical, moral, or therapeutic functions to concepts that fulfill an accurate-representational or epistemic function, or vice versa, may be detrimental. One way to improve the activity of conceptual engineering is to differentiate between the different functions that our concepts can play.

One point that has been overlooked is that, although there is some slack between the world and our concepts, we must acknowledge that the world partially determines concepts – one exception is perhaps fictional concepts. We also need to be aware that our assessment of this relation between the world and concepts can change with the advancement of science. Our evaluation of the independence of sociopolitical, epistemic, entertaining, and therapeutic practices from their neurological, physiological and psychological substrate will change with our scientific knowledge. Surely, there is necessary much more work in clarifying and differentiating the numerous functions played by our current concepts – and

Improving Conceptual Engineering by Differentiating the Functions of Concepts perhaps future ones. But, I hope, this rough exploration can shed some light on the practice and improvement of conceptual engineering.¹⁴

References

- Bach, Theodore. 2016. "Social Categories Are Natural Kinds, Not Objective Types (and Why It Matters Politically)." *Journal of Social Ontology* 2 (2): 177-201. <https://doi.org/10.1515/jso-2015-0039>.
- Bach, Theodore. 2019. "Real Kinds in Real Time: On Responsible Social Modeling." *The Monist* 102 (2): 236-58. <https://doi.org/10.1093/MONIST/ONZ008>.
- Barrett, F. Lisa, and Eliza, Bliss-Moreau. 2009. "Affect as a Psychological Primitive." *Advances in Experimental Social Psychology* 41, 167-218. [https://doi.org/10.1016/S0065-2601\(08\)00404-8](https://doi.org/10.1016/S0065-2601(08)00404-8).
- Barrett, F. Lisa, Batja Mesquita, and Maria Gendron. 2011. "Context in Emotion Perception." *Current Directions in Psychological Science* 20 (5): 286-90. <https://doi.org/10.1177/0963721411422522>.
- Barrett, F. Lisa. 2012. "Emotions Are Real." *American Psychological Association* 12 (3): 413-29. <https://doi.org/10.1037/a0027555>.
- Barrett, F. Lisa. 2016. "The Theory of Constructed Emotion: An Active Inference Account of Interoception and Categorization." *Social Cognitive and Affective Neuroscience* 12 (1): 1-23. <https://doi.org/10.1093/scan/nsw154>.
- Barrett, F. Lisa. 2017. *How Emotions Are Made: The Secret Life of the Brain*. Boston: Houghton Mifflin Harcourt.
- Brigandt, Ingo. 2010. "The Epistemic Goal of a Concept: Accounting for the Rationality of Semantic Change and Variation." *Synthese* 177 (1): 19-40. <https://doi.org/10.1007/s11229-009-9623-8>.
- Burgess, Alexis, and David Plunkett. 2013a. "Conceptual Ethics I." *Philosophy Compass* 8 (12): 1091-101. <https://doi.org/10.1111/phc3.12086>.
- Burgess, Alexis, and David Plunkett. 2013b. "Conceptual ethics II." *Philosophy Compass* 8 (12): 1102-10. <https://doi.org/10.1111/phc3.12085>.
- Cappelen, Herman. 2018. *Fixing Language: An Essay on Conceptual Engineering*. Oxford: Oxford University Press.
- . 2020. "Conceptual Engineering: The Master Argument." In *Conceptual Engineering and Conceptual Ethics*, edited by Herman Cappelen, David Plunkett, and Alexis Burgess, 1-34. Oxford: Oxford University Press.
- Cartwright, Nancy. 1989. *Nature's Capacities and Their Measurement*. Oxford: Oxford University Press.

¹⁴ **Acknowledgments:** I thank Herman Cappelen for his helpful suggestions.

- Chihara, Charles. 1979. "The Semantic Paradoxes: A Diagnostic Investigation." *The Philosophical Review* 88 (4): 590-618.
- Dixon, Thomas. 2008. *The Invention of Altruism: Making Moral Meanings in Victorian Britain*. Oxford: Oxford University Press.
- Eklund, Matti. 2002. "Inconsistent Languages." *Philosophy and Phenomenological Research* 64 (2): 251-75. <https://doi.org/10.1111/j.1933-1592.2002.tb00001.x>
- Eklund, Matti. 2014. "Replacing Truth." In *Metasemantics: New Essays on the Foundations of Meaning*, edited by Alexis Burgess and Brett Sherman, 293-310. Oxford: Oxford University Press.
- Elfenbein, A. Hilary, and Nalini Ambady. 2002. "On the Universality and Cultural Specificity of Emotion Recognition: A Meta-analysis." *Psychological Bulletin* 128 (2): 203-35. <https://doi.org/10.1037/0033-2909.128.2.203>.
- Farroni, Teresa, Enrica Menon, Silvia Rigato, and Mark H. Johnson. 2007. "The Perception of Facial Expressions in Newborns." *European Journal of Developmental Psychology* 4 (1): 2-13. <https://doi.org/10.1080/17405620601046832>.
- Floridi, Luciano. 2011. "A Defence of Constructionism." *Metaphilosophy* 42: 282-304. <https://doi.org/10.1111/j.1467-9973.2011.01693.x>.
- Frijda, H. Nico, Suprapti S. Markam, Kaori Sato, and Reinout Wiers. 1995. "Emotions and Emotion Words." In *Everyday Conceptions of Emotion: An Introduction to Psychology, Anthropology and Linguistics of Emotion*, edited by James A. Russell, José M. Fernández-Dols, Anthony S. R. Manstead, and Jane C. Wellenkamp, 121-43. Madrid: Springer.
- Frijda, H. Nico, and Batja Mesquita. 1995. "The Social Roles and Functions of Emotions." In *Emotion and Culture: Empirical Studies of Mutual Influence*, edited by Shinobu Kitayama, and Rose R. Hazel, 51-87. American Psychological Association.
- Giere, N. Ronald. 2006. *Scientific Perspectivism*. Chicago: The University of Chicago Press.
- Goldstein, Jacob. 2020. *Money: The True Story of a Made-Up Thing*. New York: Hachette Books.
- Griffiths, E. Paul. 2002. "Emotions as Natural and Normative Kinds." *Philosophy of Science* 71 (5): 901-11. <https://doi.org/10.1086/425944>.
- Hacking, Ian. 1995. "The Looping Effects of Human Kinds." In *Causal Cognition: A Multidisciplinary Debate*, edited by Dan Sperber, David Premack, and James Ann Premack, 351-94. Oxford: Clarendon Press.
- . 1999. *The Social Construction of What?* Harvard: Harvard University Press.
- Harré, Rom. 1986. *The Social Construction of Emotion*. Oxford: Blackwell.

Improving Conceptual Engineering by Differentiating the Functions of Concepts

- Isaac Gustavo, Manuel. 2020. "How to Conceptually Engineer Conceptual Engineering?" *Inquiry: An Interdisciplinary Journal of Philosophy*. <https://doi.org/10.1080/0020174X.2020.1719881>.
- Koch, Steffen. 2021. "The Externalist Challenge to Conceptual Engineering." *Synthese* 198 (1): 327-48. <http://doi.org/10.1007/s11229-018-02007-6>.
- Lindquist, A. Kristen, Tor D. Wager, Hedy Kober, Eliza Bliss-Moreau, and Lisa F. Barrett. 2012. "The Brain of Emotion: A Meta-analytic Review." *Behavioural and Brain Sciences* 35 (3): 121-202. <https://doi.org/10.1017/S0140525X11000446>.
- Machery, Edouard. 2009. *Doing Without Concepts*. Oxford: Oxford University Press.
- Margolis, Eric, and Stephen Laurence. 2014. "Concepts" *The Stanford Encyclopedia of Philosophy*. Edited by Edward N. Zalta. <https://plato.stanford.edu/archives/spr2014/entries/concepts/>.
- Marques, Teresa. 2020. "Amelioration vs Perversion." In *Shifting Concepts: The Philosophy and Psychology of Conceptual Variability*, edited by Teresa Marques and Åsa Wikforss, 260-84. Oxford: Oxford University Press.
- Mesquita, Batja, and Nico H. Frijda. 1992. "Cultural Variations in Emotions: A Review." *Psychological Bulletin* 112 (2): 179-204. <https://doi.org/10.1037/0033-2909.112.2.179>.
- McMullin, Ernan. 1985. "Galilean Idealization." *Studies in the History and Philosophy of Science* 16: 247-73. [https://doi.org/10.1016/0039-3681\(85\)90003-2](https://doi.org/10.1016/0039-3681(85)90003-2).
- Nado, Jennifer. 2019. "Conceptual Engineering, Truth and Efficacy." *Synthese* 198: 1507-27. <https://doi.org/10.1007/s11229-019-02096-x>.
- Norenzayan, Ara, and Steve J. Heine. 2005. "Psychological Universals: What Are They and How Can We Know?" *Psychological Bulletin* 131 (5): 763-84. <https://doi.org/10.1037/0033-2909.131.5.763>.
- Phan, K. Luan, Tor Wager, Taylor F. Stephan, and Israel Liberzon. 2002. "Functional Neuroanatomy of Emotion: A Meta-analysis of Emotion Activation Studies in PET and fMRI." *Neuroimage* 16 (2): 331-48. <http://doi.org/10.1006/nimg.2002.1087>.
- Plunkett, David, and Herman Cappelen. 2020. "A Guided Tour of Conceptual Analysis and Conceptual Ethics." In *Conceptual Engineering and Conceptual Ethics*, edited by Herman Cappelen, David Plunkett, and Alexis Burgess, 1-34. Oxford: Oxford University Press.

Rogelio Miranda Vilchis

- Prinzing, Michael. 2017. "The Revisionist Rubric: Conceptual Engineering and the Discontinuity Objection." *Inquiry: An Interdisciplinary Journal of Philosophy* 61 (8): 854-80. <https://doi.org/10.1080/0020174X.2017.1385522>.
- Riggs, Jared. 2020. "Conceptual Engineers Shouldn't Worry about Semantic Externalism." *Inquiry*. <https://doi.org/10.1080/0020174X.2019.1675534>.
- Rorty, Richard. 1979. *Philosophy and the Mirror of Nature*. Princeton: Princeton University Press.
- Rorty, Richard. 1990. *Objectivity, Relativism, and Truth*. Cambridge: Cambridge University Press.
- Russell, A. James. 1991. "Culture and the Categorization of Emotions." *Psychological Bulletin* 110 (3): 426-50.
- Sawyer, Sarah. 2020a. "The Role of Concepts in Fixing Language." *Canadian Journal of Philosophy* 50 (5): 555-65. <https://doi.org/10.1017/can.2020.5>.
- Sawyer, Sarah. 2020b. "Truth and Objectivity in Conceptual Engineering." *Inquiry* 69 (9-10): 1001-22. <https://doi.org/10.1080/0020174X.2020.1805708>.
- Scharp, Kevin. 2013. "Truth, the Liar, and Relativism." *Philosophical Review* 122 (3): 427-510.
- Scharp, Kevin. 2014. "Replacing Truth." *The Philosophical Quarterly* 64 (256): 535-37. <https://doi.org/10.1093/pq/pqu019>.
- Simion, Mona. 2018. "Epistemic Trouble for Engineering 'Woman'." *Logos & Episteme* 9 (1): 91-8. <https://doi.org/10.5840/logos-episteme2018916>.
- Simion, Mona, and Christopher Kelp. 2019. "Conceptual Innovation, Function First." *Noûs* 54 (4): 1-18. doi:10.1111/nous.12302.
- Strawson, F. Peter. 1963. "Carnap's Views on Constructed Systems Versus Natural Languages in Analytic Philosophy." In *The philosophy of Rudolf Carnap*, edited by Paul A. Schilpp, 503-18. La Salle: Open Court.
- Tassinary, G. Louis, and John T. Cacioppo. 1992. "Unobservable Facial Actions and Emotion." *Psychological Science* 3 (1): 28-33. <https://doi.org/10.1111/j.1467-9280.1992.tb00252.x>.
- Thomasson, Amie. 2020. "A Pragmatic Method for Conceptual Ethics." In *Conceptual Ethics and Conceptual Engineering*, edited by Herman Cappelen, David Plunkett, and Alexis Burgess, 435-58. Oxford: Oxford University Press.
- Vytal, Katherine, and Stephan Hamman. 2010. Neuroimaging Support for Discrete Neural Correlates of Basic Emotions: A Voxel-based Meta-analysis. *Journal of Cognitive Neuroscience* 22 (12): 2864-85. <https://doi.org/10.1162/jocn.20b09.21366>.

Improving Conceptual Engineering by Differentiating the Functions of Concepts

- Whitacre, James, and Axel Bender. 2010. "Degeneracy: A Design Principle for Achieving Robustness and Evolvability." *Journal of Theoretical Biology* 263 (1): 143-53. <https://doi.org/10.1016/j.jtbi.2009.11.008>.
- Wierzbicka, Anna. 1992. *Semantics, Culture, and Cognition: Universal Human Concepts in Culture-Specific Configurations*. Oxford: Oxford University Press.
- Williams, Raymond. 1977. "Keywords: A Vocabulary of Culture and Society." *Science and Society* 41 (2): 221-24.
- Wilson-Mendenhall, D. Christine, Lisa F. Barrett, Kyle W. Simmons, and Lawrence W. Barsalou. 2011. "Grounding Emotion in Situated Conceptualization." *Neuropsychologia* 49 (5): 1105-127. <https://doi.org/10.1016/j.neuropsychologia.2010.12.032>.
- Wilson-Mendenhall, D. Christine, Lisa F. Barrett, and Lawrence W. Barsalou. 2013. Neural Evidence that Human Emotions Share Core Affective Properties. *Psychological Science* 24 (6): 947-56. <https://doi.org/10.1177/0956797612464242>.

RATIONAL DECISION-MAKING IN A COMPLEX WORLD: TOWARDS AN INSTRUMENTAL, YET EMBODIED, ACCOUNT

Ragnar VAN DER MERWE

ABSTRACT: *Prima facie*, we make successful decisions as we act on and intervene in the world day-to-day. Epistemologists are often concerned with whether rationality is involved in such decision-making practices, and, if so, to what degree. Some, particularly in the post-structuralist tradition, argue that successful decision-making occurs via an existential leap into the unknown rather than via any determinant or criterion such as rationality. I call this view radical voluntarism (RV). Proponents of RV include those who subscribe to a view they call Critical Complexity (CC). In this paper, I argue that CC presents a false dichotomy when it conceives of rationality in Cartesian – i.e. ideal and transcendental – terms, and then concludes that RV is the proper alternative. I then outline a pragmatist rationality informed by recent work in psychology on bounded rationality, ecological rationality, and specifically embodied rationality. Such a pragmatist rationality seems to be compatible with the tenets of post-structuralism, and can therefore replace RV in CC.

KEYWORDS: voluntarism, bounded rationality, ecological rationality, embodied rationality, complexity

Introduction

An important question in epistemology relates to how successful decision-making is possible when we act on or intervene in the world. Is there perhaps some nomological principle – some norm of rationality – that guides us? Or, is there no such norm; successful decision-making results from an act of unbridled volition? How we answer this question has significant import for, not only philosophical inquiry, but also our practical socio-political affairs. If a norm of rationality exists, then we should presumably let it determine our decisional practices. Surely, we want our decisions to be rational. A norm of rationality could however be considered constraining and exclusionary; it may conflict with putative desirables like human freedom and diversity. In this paper, I aim to contribute towards a resolution in this debate.

Epistemologists in the so-called post-modern tradition often maintain that successful decision-making can occur in the absence of rationality (see Searle 1977; Derrida 1978; Habermans 1990; Foucault 2001). Exemplary of this view is a post-structuralist approach to complexity theory called Critical Complexity (CC).¹ Proponents of CC include notably Paul Cilliers, Rika Preiser, and Minka Woermann. CC draws on both complexity theory and Derrida's post-structural semantics to argue that the world's manifest complexity radically overdetermines rational decision-making. When we act on or intervene in the world, there are no non-contextual, non-provisional norms – *viz.* criteria or constraints – that can determine our choices. For CCists, we are nonetheless ethically compelled to act on and intervene in the world (Woermann and Cilliers 2012; see also Derrida 1999). Despite the absence of determinant norms, successful decision-making is possible through an *existential leap* forwards into the unknown. We can think of such a leap as an act of pure will or volition in the face of radical uncertainty caused by the world's overwhelming complexity.

Let us say, for example, that I am walking to work and a panhandler asks me for money. There is a moment when I must decide what to do: stop to give the panhandler some money or look the other way and continue walking. For CCists, RV dictates that my choice cannot be a strictly rational one. I *will* decide one way or the other, and then act accordingly, but this decision will be the product of a kind of unanalysable volitional compulsion rather than determined by any prescriptive principles. In this context, for principle P to *determine* decision D is for P to force or dictate D. Or, more aptly, for principle P to determine decision D is for P to play a necessary (if not sufficient) role in D. P is then primary (if not alone) in realising or instantiating D.

Specific to our discussion, we can say that rationality determines some decision-making activity if it plays the primary affective role (amongst sundry 'affectors') in the outcome of that decision-making activity. So, my decision to help or not help the panhandler is rational if rationality determines the outcome of that decision. CCists would deny this. Following Derrida, decision-making is not determined in this way. As mentioned, an existential leap forwards instead plays the primary affective role in the outcome of some decision. I will call this post-structural take on decision-making *radical voluntarism* (RV). I introduce and explicate CC and RV in section 1. I also argue that RV does not describe how we *de facto* make successful decisions. Were RV correct, we should behave in a random and erratic fashion whenever faced with two or more choices in some actional or interventive encounter with complexity. This is however not what we witness.

¹ Following Cilliers (1998), I take post-structuralism to be a kind of post-modernism.

Instead, we seem to quite easily make successful decisions day-to-day as we navigate our complex world.²

In section 2, I cash out *successful* decision-making in terms of agents attaining goals. My decision regarding the panhandler, for example, will be successful to the degree that my resultant action (give or ignore) is concordant with my pertinent goals. Perhaps, I will give if I have previously found token acts of kindness towards strangers emotionally rewarding and I am feeling emotionally drained. Perhaps, I will ignore if I am concerned for the general well-being of society and I have heard experts urge people not to give money to panhandlers. Our successful day-to-day decisions can be perfectly rational if rationality is understood in this deflated way. That is, if rationality is understood in *pragmatist* – i.e. naturalised and instrumental – terms. In developing such a pragmatist rationality, I draw on recent work in psychology on instrumental rationality, ecological rationality, and specifically embodied rationality. Originally outlined in Spellman and Schnall (2009), embodied rationality has been little discussed in the philosophical literature (Gupta 2021 is a notable exception), and my co-opting of it for pragmatist ends should therefore make a novel contribution to epistemology.

In section 3, I outline how my pragmatist rationality might be incorporable into CC. CC associates rationality with the kind of Cartesian or transcendental and deterministic rationality that flourished in 18th century epistemology. Given the putative deficiencies of this view, CC concludes that RV is the proper alternative. I argue that this presents a false dichotomy between two extremes (see also van der Merwe 2021). Thinking of rationality in pragmatist terms may permit CCists to embrace the idea that we are capable of rational decision-making without abandoning the core tenets of post-structuralism. As per embodied rationality, we can ‘ground’ successful decision-making in the sensory-motor capabilities we employ during goal-attainment, where ‘grounds’ is cashed out in suitably weak naturalised terms, rather than in metaphysically constitutive terms. In explicating what he calls “embodied heuristics,” Gerd Gigerenzer (2021) invokes the example of a baseball outfielder catching a flyball. The outfielder does not perform anything like a mathematical calculation related to measurements of height, distance, mass, acceleration, and like. Instead, she follows what Gigerenzer calls the “gaze heuristic:” “Fixate your eyes on the ball, run, and adjust your speed so that the angle of gaze remains constant” (2021, 5). To engage in rational (i.e. goal-attaining) decision-making, the outfielder need only have the ability to (1) hold her gaze on

² It is debatable whether the world as a whole is complex or only parts of it (see Ladyman and Wiesner 2021 for an overview of the debate). For the purposes of this paper, I will anyhow follow CC in supposing that the world as a whole is complex (presumably by degrees).

the ball, (2) run, and (3) adjust her running speed. The gaze heuristic, says Gigerenzer, is thus *embodied* in the outfielders sensory-motor capabilities.³ This embodiment is what I will call ‘grounding.’

This paper centres around notions of decision-making and rationality. These are, of course, big topics. My aim here is not to settle once-and-for-all the nature of decision-making and rationality, nor to offer necessary and sufficient conditions for their instantiation. Following CC, I am instead specifically concerned with actional decision-making in the face of complexity; that is, decisions that precede some action or intervention in our complex world. Many of our actional decisions involve an encounter with complexity in some or other form. Examples include seemingly mundane tasks like choosing what groceries to buy (in the complex economic system) or choosing to cross the road (in the complex traffic system).

For the purposes of this paper, I take decision-making to involve the actual moment an agent makes some choice in the face of some variety of alternatives, and not the moment/s immediately prior to or proceeding such a choice. When I choose whether to give to or avoid the panhandler, for example, there is presumably a moment – an instant in time – where I transition from a cognitive state of non-decision to one of decision-made. This moment is the moment of *choice*, and it is where CC considers RV to apply.⁴ A choice occurs immediately posterior to mental deliberation (*viz.* contemplation and prediction) and immediately prior to physical action (*viz.* tactile engagement with the world). Our concern is thus with the liminal moment where the former transitions to the latter. I, for example, deliberate the panhandler’s request; I make a choice; and then act accordingly. I argue that this transition from deliberation to action does not occur via Cartesian rationality nor via RV, but rather via a kind of pragmatist rationality.⁵ I will call this *experiential rationality*. Experiential rationality is closely related to embodied rationality, but differs by incorporating the philosophical notion of grounding (or ‘grounding’).⁶

³ Hawks and bats likewise utilise the gaze heuristic while intercepting prey during flight (Gigerenzer 2021, 7-9).

⁴ Arguably, decision-making is ongoing rather than contained in an instant. We are constantly updating our decisional states as we navigate the world. For the sake of argument, I will nonetheless grant CC that decision-making occurs in an instant.

⁵ There may be other ways that successful decision-making occurs. The two options of (1) rationality (whether of the Cartesian or pragmatist kind) and (2) RV do not necessarily exhaust the possibilities. However, since CCists specifically contrast RV with rationality, I will attempt to defend rationality, and then argue for its compatibility with post-structuralism.

⁶ It should be apparent that the pragmatism I have in mind here is inspired by the so-called

Also note the following provisos. I will gloss over much of the nuance related to the similarities and differences between bounded rationality, ecological rationality, and embodied rationality (see however the collection in Viale 2021 for the status of the current debate). This is because my aim is simply to draw support from embodied rationality, and not to develop a detailed psychology (or physiology for that matter) of rational decision-making. Further, although my aim is prescriptive in advising CCists to adopt experiential rationality, my account of rational decision-making is itself descriptive. I aim to explicate how human agents *de facto* utilise rationality in their successful decision-making practices, and not necessarily how they ought to do so. I also take rationality to be a capacity exercised by an individual agent. Social or collective rationality is thus a special application of, rather than constitutive of, rationality. Social or collective rationality is individual rationality exercised in a social context. Note also that I will not argue for whether and/or how rational decision-making may be specifically related to belief, knowledge, understanding, and truth. Although important in their own right, these issues are not our direct concern here (see however the collection in Knauff and Spohn 2021 for the status of the current debate).

In section 4, I engage with a possible response: CCists may claim that we should embrace *aporetic* logic – a kind of post-structural dialetheism – instead of my pragmatist rationality. In response, I argue that aporetic logic creates more problems than it solves.

1. Critical Complexity (CC) and Radical Voluntarism (RV)

My goal in this section is to briefly outline CC and its post-structural understanding of successful decision-making. I emphasise CC's criticism of the claim that rationality can serve as a norm for decision-making and that the proper alternative – RV – involves an existential leap into the unknown. I proceed as follows. Firstly, I briefly outline a key Derridean notion – *différance* – that is foundational to CC's view (section 1.1). Secondly, I explicate CC's criticism of rationality (section 1.2). Lastly, I critique RV (section 1.3).

experience pragmatism of e.g. Putnam, McDowell, and Misak rather than the *linguistic* pragmatism of e.g. Davidson, Brandom, and Price (see Misak 2014; Levine 2019 ch. 1; van der Merwe forthcoming for more on this distinction).

1.1 Derrida's Notion of *Différance*

At the heart of Derridean post-structuralism is the claim that we can never truly capture the meaning of a linguistic sign or network of signs in a semantic system such as a language. This is because of “*différance*.”

For Derrida (e.g. 1982; 1988), a semantic system has no centre, no locus or ground of meaning. Instead, following Saussure (1974), meaning is constituted by the many differences between signs making up the system. Meaning is generated by the endless and iterative interaction of these differences. Deviating from Saussure, Derrida however attributes the source of this meaning-generation to *différance*. The notion of *différance* is notoriously difficult to define. We can nonetheless think of it as an ontologically significant, yet ethereal and nebulous, kind of oscillation or “movement” (as Derrida puts it) that both creates and destroys semantic differences. *Différance* should be understood as both noun and verb, both present and absent. *Différance*, says Derrida, is

the systematic play of differences, of the traces of differences, of the spacing by means of which elements are related to each other. This spacing is the simultaneously active and passive... (1981, 27).

Différance plays or “dances” between signs. It produces, or rather *is* the production of, fleeting instances of meaning, meaning that is always elusive to epistemic capture (Derrida 1981; see also Cilliers 1998 ch. 3; Woermann 2016 ch. 3).⁷

Following Derrida, CCists consider meaning to be generated by, but not grounded in, the play of *différance*. According to de Villiers-Botha and Cilliers,

meaning is not static or final – it is always deferred... The sign is produced by the system, but at the same time the meaning that is generated for it through the process of *différance* reverberates through the system, influencing other signs (2010, 31).

The meaning in a semantic system cannot be codified into an ordered nomological structure. Meaning is never fully present to an epistemic inquirer; there is no ‘transcendental signified.’ The force of *différance*, says Woermann, “destroys the... possibility of saturated meaning” (2016, 100). Meaning is necessarily provisional; it cannot be “closed;” closure of meaning is always “deferred.”

⁷ In the context of complexity theory, Woermann thinks of *différance* as “the play of disorder... and entropy” within a complex system (2016, 64)

1.2 CC on Rationality

CC takes the ungroundedness of meaning to have wide-reaching implications for knowledge, truth, and – most importantly for our purposes – decision-making. Since knowledge, truth, and decision-making rely on capturing the meaning of concepts, they are – like meaning – prone to *différance*'s disruptive influence. Given the play of *différance*, decision-making is never determined or calculable. There are no transcendental criteria or constraints – e.g. an ideal of rationality – that we can fix on to secure certainty (Woermann and Cilliers 2012). Derrida's semantics dispels the Cartesian dream that the “world can be made rationally transparent and can yield objective and universal knowledge” (Woermann 2016, 88; see also Cilliers 2000b). On CC's account, our decisional actions are inescapably arational.

Note that CC is specifically against what Woermann (2016) calls a “strong” or “modernist” rationality, i.e. rationality that serves as an infallible, yet epistemically accessible, guide to decision-making. On such a Cartesian view,

agents are believed to make decisions based on reasonable [i.e. rational] principles and calculations, and the trajectory from decision to outcome is viewed in terms of a linear causality (Woermann 2016, 126; see also Woermann et al. 2018).

However, because of

the non-closure of meaning... our decisions and actions cannot be objectively described. Instead, we must engage in contingency, alterity, and the over-determinations that characterise our contexts (all of which involve judgement and sense-making that surpass calculation and pure rational argumentation) (Woermann 2016, 8).

This engagement involves RV.

1.3 RV: A Leap into the Unknown

According to Woermann, we undergo a “terrible experience of undecidability” prior to acting on or intervening in the world (2016, 180). When engaging in decision-making, we must, says Derrida, “go through an ordeal of undecidability in order to decide. So, to that extent the result, by definition, is unpredictable, unknown” (Derrida in Cilliers et al. 2016, 173; see also Human 2016). This ordeal results from the absence of any determinants for decision-making, e.g. norms of rationality. Consequently, “in order for a decision to be a decision it has to go through a moment when irrespective of what you know, you make a leap into the decision” (Derrida 1999, 280). The outcome of my decision regarding whether to give to versus ignore a panhandler is then radically uncertain. I cannot appeal to

rationality or similar principles during decision-making; I must “just do it” (as Nike marketers like to say).

That said, Derrida and CC do recognise that we somehow make decisions that lead to successful actions on and interventions in the world. For Derrida and CC, we do so – we overcome the ordeal of undecidability – through the leap mentioned above. This leap is blind in the sense that it occurs independent of any determining criterion or constraint. It is also unanalysable and unquantifiable using traditional – i.e. modernist – methods, yet it somehow propels us from undecidability to decidability. Without the ‘invisible hand’ of rationality, we experience a moment of pure will – a moment of compulsion, rather than guidance – towards some course of action. This is RV, and, according to Derrida, it “not only threatens a break with science in the strict sense, but with philosophy as ontology, as knowledge...” (Derrida in Cilliers et al. 2016, 173). My decision regarding the panhandler is then supposed to involve an ordeal of undecidability that results in a leap to action, a leap that is *au fond* arational. The same putatively applies to all decisions and actions we make in our complex world, even those involving everyday activities like buying groceries and crossing the road (I discuss in section 4 why CCists cannot draw a distinction between cases where RV applies versus cases where it does not).

Importantly, for CC, RV introduces *freedom*. According to Woermann and Cilliers, rationality is radically overdetermined by the world’s complexity, and “it is these overdeterminations that generate freedom...” (2012, 455). We are not bound by decisional principles or linear rules for action; instead, we are the existential deciders of our modal future. For CC, this kind of freedom also has unavoidably *ethical* implications. According to Preiser et al., the “ethical moment is situated in the moment in which we take the leap from that which is known to that which is uncertain or unknown” (2013, 271). This moment “is born once we enter into the gap of the infinite abyss that is created by the limits of our models”, i.e. the limits of our capacity to capture meaning (Preiser et al. 2013, 271). For CC, the loss of meaning introduces freedom, and freedom introduces ethics. This because with freedom comes *responsibility* (Derrida 2002; Cilliers, 2005; Woermann, 2016). The decisional leap at the core of RV is inherently ethical given its nondeterminate nature. We cannot defer accountability for the consequences of our decisions onto self-extrinsic factors, such as norms of rationality.

However, what exactly this decisional leap – this “ethical moment” – entails remains largely mysterious on CC’s account. Why do we decide one way rather than another at any given moment? Attempting to answer this question would

presumably introduce the kind of criterial or constraining norms RV rules out.⁸ But, one naturally wonders how *successful* decision-making is possible if all decisions ultimately result from RV rather than from being principled by or grounded in something more exacting, something like rationality. If there are no discernible norms for decision-making, how is it that we can make decisions that generate a preferable or beneficial, rather than aberrant or random, outcome? Today, for example, I decided to get out of bed, I decided to come to work,⁹ and I then decided to continue writing this paper where I left off yesterday. These are just three of the countless decisions I made today that most would agree are successful on any non-trivial definition of ‘success’ (I argue in section 3 that we should think of ‘success’ in this context in terms of goal-attainment). I further made these decisions without anything outwardly resembling Derrida’s “ordeal of undecidability” or Woermann’s “terrible experience of undecidability.” In fact, I performed these decisions without much contemplation or effort at all.

As we act on and intervene in the world moment-to-moment we repeatedly make decisions that are *prima facie* successful. However, this should be impossible were RV correct. Without some minimal determinant/s for decision-making, we should mostly make erratic or arbitrary decisions proceeded by random or akratic actions. Yet, this is not what we outwardly experience nor what we witness in the behaviour of others. CC seemingly cannot account for how and why we function successfully moment-to-moment as decision-making agents despite the world’s evident complexity.

2. Experiential Rationality: Naturalised, Instrumental, and Embodied

I have outlined CC and its alternative to rational decision-making: RV. I have also argued that RV insufficiently accounts for our everyday decision-making practices.

Edgar Morin – who has partly inspired CC (Woermann 2016) – endorses freedom, but also recognises the need for decisional norms or what he calls “determinations.” “Free action,” he says, “depends upon the knowledge and utilization of determinations (constants, structures, laws)” (Morin 2008, 114). Determinations are “conditions” for decision-making:

Freedom also presupposes two conditions. To begin with, there is an internal condition, involving the cerebral, mental, and intellectual ability to consider a situation and establish choices and chances of success. Then there are external

⁸ Derrida does at times suggest that there is a kind of quasi-theological force operant in the world, a force that can compel our ethical decisions (see Derrida in Cilliers et al. 2016). CCists do not follow Derrida in this regard, however.

⁹ I decided not to give to the panhandler, by the way.

conditions which render the choices possible (Morin 2008, 78).

Morin's internal condition equates to what we would normally call rationality: our capacity for rationality grants us the ability to "consider a situation and establish choices and chances of success." Morin's external conditions are states of the world 'out there' independent of us: what we might call "facts" or "states of affairs." My concern in this paper is specifically with Morin's internal condition. While acknowledging its presence, CC does not consider any such internal condition to be determinate of decision-making. RV, rather than rationality, plays the primary affective role in the outcome of some decision-making activity.

In this section, I outline a pragmatist conception of rationality that is potentially incorporable into CC. Such a pragmatist rationality should, on the one hand, constrain decision-making without the rigidity of Cartesian rationality; and, on the other hand, allow for some degree of decisional freedom without the laxity entailed in RV. To be a pragmatist kind of rationality, rationality must, I propose, satisfy two conditions:

C1: *Naturalised*, in the sense of taking into the account the Darwinian insight that human agents – including their cognitive faculties – are the product of biological evolution. Being a cognitive ability, rationality is therefore a product of biological evolution.¹⁰ Like other human faculties, rationality must 'emerge' somehow in both ontogeny and phylogeny.¹¹ Rationality is a natural outcome of our Darwinian genealogy, as are hunger, desire, and similar physiological processes (see also Campbell 1974; Dennett 1995; Wilke and Todd 2010).

C2: *Instrumental*, in the sense of being centred around goal-attainment, where attaining a goal involves getting something we want (see Goldman 1970, ch. 4; Okasha 2018, ch. 7). Some decision-making activity is therefore rational when its outcome aligns with some pertinent goal, a goal that is consistent with the kind of goals human agents tend to have (i.e. not aberrant goals premised on psychotic, hyper-emotional, or self-destructive tendencies, for example¹²).

A kind of rationality that satisfies C1 and C2 involves neither a top-down executive commander of decision-making (as Cartesians might suggest) nor being

¹⁰ See Okasha (2018 ch. 6) for an informative discussion on how rationality may have evolved by natural selection (see also Godfrey-Smith 2002). Gigerenzer and Sturm (2012) argue at length that rationality can be both descriptively *and* normatively naturalised.

¹¹ How exactly this kind of emergence might occur is not our concern here (see however the collection in Bedau and Humphreys 2008).

¹² Such aberrant goals are the exception rather than the norm. Kenrick and Giskevicius (2013 ch. 6) and Buss (2019 ch. 10) argue nonetheless that some risk-taking behaviour can serve an evolutionary adaptive function.

lost in a sea of semantic overdetermination (as CCists suggest). It is instead a natural product of decision-making processes employed during goal-attainment.

I now discuss some contemporary theories of rationality that align with and inspire the kind of pragmatist rationality I have in mind. I focus on psychological accounts of rationality developed by Steven Pinker (section 2.1) and Gerd Gigerenzer (section 2.2). Most important for our purposes is a recent derivative of Gigerenzer's view that has come to be known as *embodied rationality* (section 2.3).

2.1 Pinker's Instrumental Rationality

Pinker thinks of rationality primarily in instrumental terms. Rationality, he says, equates to "the ways an intelligent agent ought to reason, given its goals and the world in which it lives" (Pinker 2021 ch. 1 para. 14 emphasis removed; see also Haselton et al. 2009; Broome 2013; Kenrick and Griskevicius 2013).¹³ Rationality is also neither reducible to deductive logic nor does it answer to some presiding meta-rationality (Pinker 2002). Instead, the rational operations our minds perform are foundational on our biological neural hardware. Rationality is ongoing as we engage in and overcome real-life, sometimes messy, worldly decision-making challenges (see also Campbell 1974; Churchland 1987).

In response to celebrated demonstrations of supposedly widespread human irrationality (e.g. Ariely 2008; Thaler and Sunstein 2008; Kahneman 2011), Pinker shows how measures of irrationality drop significantly when tasks designed to highlight irrationality are reframed in ways that align with our everyday concerns, rather than being contrived in artificial scenarios specifically designed to fool our decision-making capabilities (see also Gigerenzer 2008; Haselton et al. 2009; Spellman and Schnall 2009; Kenrick and Griskevicius 2013). Dan Mercier and Hugo Sperber (2017) argue that, since rationality must have evolved by natural selection, it is unlikely to be systematically maladaptive (see also Wilke and Todd 2010). Although our reasoning (*viz.* rational inquiry) sometimes falters, it is generally reliable in helping us attain of the kind of generic goals that human beings tend to pursue (see also Haselton et al. 2009; Pinker 2010; Buss 2019; Edis and Boudry 2019). These generic goals include environmental navigation, thirst and hunger satiation, social cooperation, and the like.

So-called cognitive illusions – the gambler's fallacy, confirmation bias, priming, framing effects, and similar errors of reasoning – do not demonstrate that we are irrational or even mostly irrational. "They lead to incorrect answers, yes,

¹³ Giovanni Rolla states necessary and sufficient conditions for rationality in instrumental terms: "S is a rational agent iff S is able to achieve a specific goal through the exercise of the relevant capabilities in suitable conditions" (2016, 20).

but they are often correct answers to different and more useful questions” (Pinker 2021 ch. 1, the moral from cognitive illusions section, para. 8; see also Godfrey-Smith 1996, 2002). Granted, we are sometimes prone to irrationality, but this must be the exception rather than the norm, otherwise our generic decision-making activities should largely fail (recall section 1.3). Most of us however regularly and reliably make goal-attaining decisions – i.e. successful decisions – such as those involved in grocery buying and road-crossing.

2.2 Gigerenzer’s Ecological Rationality

Gigerenzer’s *ecological rationality* or what he calls “rationality for mortals” is an extension of Herbert Simon’s (1983; Newell and Simon 1972) much-discussed *bounded rationality*. Bounded rationality, says Gigerenzer,

is the study of how humans and other animals rely on heuristics to achieve their goals in situations of uncertainty. It differs from axiomatic rationality, which asks whether humans conform to logical principles [as in the Cartesian approach] (2021, 1).

Such heuristics compose an “adaptive toolbox” for successful decision-making. They are “fast and frugal” rules-of-thumb of the sort we should expect imperfect biological beings to employ. We are not angels; our cognitive capabilities have been tinkered together in a kludgy and piece-meal fashion by natural selection over millennia.

Heuristics, says Gigerenzer, “work in real-world environments of natural complexity... where an optimal strategy is often unknown or computationally intractable” (2008, 8 emphasis removed; see also Gigerenzer and Selton 2002; see Gigerenzer and Sturm 2012, 247-251 for a list of typical heuristics). Gigerenzer uses the example of playing chess. We play chess using a kind of intuitive reasoning, and sometimes play it very well, without having to calculate all possible outcomes and without making a blind decisional leap at every move. Some sort of ‘algorithm’ is running – some sort of ‘calculation’ is going on – but this only approximates anything like an ideal Cartesian rationality (Gigerenzer and Brighton 2009; Gigerenzer and Sturm 2012; see also Vlerick and Broadbent 2015). As Giovanni Dosi and colleagues put it,

[h]uman agents tackle every day, with varying degrees of success, highly complex and ‘hard’ (in the sense of computability theory) problems with their highly limited computational capabilities... we cannot handle more than a very limited number of the overwhelming number of interdependencies that characterize our world, but nevertheless we go along, sometimes decently well, with simple but useful representations and simple but effective heuristics (2021, 493).

Gigerenzer's heuristics are similar to what Leda Cosmides and John Tooby call "reasoning instincts." Reasoning instincts "make certain kinds of inferences just as easy, effortless, and 'natural' to humans as spinning a web is to a spider or building a dam is to a beaver" (Cosmides and Tooby 1994, 330). Reasoning instincts employ a kind of fallible and adaptive Darwinian reasoning. Like other animals, we follow intuitive rules-of-thumb of the sort that proved useful to our ancestors, and that can be successfully applied to much of our modern environment (see Dennett 2009; Haselton et al. 2009; Kenrick and Griskevicius 2013; Mercier and Sperber 2017; Pinker 2021; Mastrogiorgio et al. 2022). Heuristic-based decision-making is rational *qua* rationality understood in a suitably naturalised and instrumental way.

A philosophical question nonetheless remains regarding what *grounds* rationality. We want to ground rationality since (as argued in section 1.3) successful decision-making requires constraint/s. If successful decision-making involves rationality, then rationality cannot be a *laissez-faire* matter, otherwise success would be arbitrary (in the way RV seems to imply). Grounding in philosophy is conventionally understood in metaphysical terms. According to Ricki Bliss and Kelly Trogdon metaphysical grounding is "a form of constitutive (as opposed to causal or probabilistic) determination or explanation" (2021, np). Some superficial phenomenon of interest is constitutively – i.e. necessarily – determined or explained by some more fundamental grounding base, e.g. simples, dispositions, bare particulars, or similar fundamentalia. This is not the kind of grounding that applies to my pragmatist rationality, *viz.* experiential rationality. As I outline in sections 2.3 and 3, a pragmatist kind of grounding – or 'grounding' – is provisional and contextual in that it applies to biological agents making everyday decisions here and now. It does apply to generic agents (including AI systems and aliens perhaps) engaged in (Turing machine-like) decision-making *simpliciter*. As we will see, experiential rationality is naturalistically 'grounded' in those *sensory-motor capabilities* we instrumentally employ during goal-attainment.

2.3 Embodied Rationality

Although psychologists do not usually invoke the philosophical notion of grounding, we can think of proponents of embodied rationality as seeking to ground rationality in the *sensory-motor capabilities* we employ when engaged in successful actions on and interventions in the world. Sensory-motor capabilities are those biological bodily skills employed in receiving sensory information from the world and then generating an appropriate motor response. As mentioned, grounding rationality in sensory-motor capabilities will involve a weak, naturalised, and instrumental kind of grounding – 'grounding' – rather than a

strong metaphysically constitutive kind of grounding. Rationality – *viz.* the utilisation of reasoning heuristics – is ‘grounded’ specifically in those motor-sensory capabilities we employ during goal-attainment. Central to experiential rationality is the idea that such a ‘grounded’ rationality determines successful decision-making (where ‘determination’ recall involves playing the primary affective role in the outcome of a decision, and ‘success’ is cashed out in terms of goal-attainment).

In arguing for embodied rationality, Antonio Mastrogiorgio and Enrico Petracca note that Gigerenzer’s ecological rationality treats heuristics as “formal rules for information processing,” rules that are “implemented through ‘computer programs’” in the mind (2016, 225). This, they argue, is inconsistent with a Darwinian understanding of human cognition, where cognition should be non-algorithmic and kludgy (see also Kauffman 2019; Mastrogiorgio et al. 2022). Embodied rationality is an attempt to overcome this ostensible deficiency in Gigerenzer’s view.

As the name suggests, embodied rationality holds that the body plays a central role in rational inquiry. According to Mastrogiorgio et al., embodied rationality

invites us to abandon a third person rationality (where cognitive processes can be expressed as objectified, algorithmic rules for information processing) and calls into account the biological realm... [E]mbodied rationality emphasizes the constitutive dependence of heuristics on the human body and in particular on the sensory-motor system... [C]ognitive processes can be understood precisely as they are grounded on the sensory-motor system, and not prescinding from it [sic] (2022, 12; see also Rolla 2016; Gupta 2021).

‘Demoting’ rationality from the transcendental to the natural in this way renders it, not only compatible with Darwinism, but also potentially incorporable into CC. CCists reject any notion of a transcendental (i.e. Cartesian) rationality, but can potentially embrace the weaker suggestion that rationality is embodied or ‘grounded’ in the sensory-motor system (I argue to this effect in section 3).

Influenced by proponents of embodied rationality, Gigerenzer (2021) has recently suggested that his reasoning heuristics be thought of as “embodied heuristics.” Embodied heuristics are “rules of thumb that exploit specific sensory and motor capacities in order to facilitate high-quality decisions in an uncertain world” (Gigerenzer 2021, 2). And the “ecological rationality of a heuristic is measured by the degree to which it can attain a goal” (Gigerenzer 2021, 5). This intimates at the kind of naturalised and instrumental rationality entailed in what I am calling experiential rationality. Ecological rationality further

analyzes the match between the adaptive toolbox of an individual or species, and the environment. A *match* refers to the likelihood that a given heuristic achieves a given goal in a given environment (Gigerenzer 2021, 4 original emphasis).

As mentioned in the introduction, a baseball outfielder catching a flyball is an apt example. Recall that the outfielder does not perform any formal calculations when catching the ball. Instead, she simply follows what Gigerenzer calls the “gaze heuristic:” “Fixate your eyes on the ball, run, and adjust your speed so that the angle of gaze remains constant” (2021, 5). Rationality plays a central role in such processes; it is embodied in the sensory-motor capabilities the outfielder employs during goal-attainment (see also Gallagher 2018). For Gigerenzer, the relevant sensory-motor capabilities are part of both our phylogenetic and ontogenetic endowment. Phylogenetically, they are a product of Darwinian evolution (see also Jonsson and von Hofsten 2003; Mastrogiorgio and Petracca 2016). We obviously did not evolve to catch flyballs, but the ability to do so is an *exaptation* from capabilities our ancestors employed during activities like hunting for food (Gigerenzer 2021; see also Kauffman 2019; Mastrogiorgio et al. 2022). Regarding ontogeny, Amitabha das Gupta states that

an infant acquires her capacity to reason based on her embodied experience which she attains due to the interplay of certain bodily structures or modalities along with certain emotive elements... Reason [*viz.* rationality] thus emerges out of embodied experience (2021, 14).

Gupta thus invokes a suitably weak (i.e. naturalised), rather than a strong, sense of emergence that is consistent with experiential rationality (see O’Connor 2021 for more on the distinction between weak and strong emergence). In both phylogeny and ontogeny then, rationality (weakly) emerges from human beings’ everyday sensory-motor interactions with the world. Embodied rationality thus satisfies C1: *Naturalised*. Further, rationality does not obtain in any old sensory-motor capabilities. Sensory-motor capabilities must be of the right sort, the sort employed during goal-attainment (e.g. catching a fly-ball). Embodied rationality thus satisfies C2: *Instrumental*.

Embodied rationality also allows us to constrain or ‘ground’ rationality in a way that is compatible with C1 and C2. Experiential rationality differs from embodied rationality in emphasising the role of the philosophical notion of grounding (or ‘grounding’ when suitably pragmatized). When it comes to decision-making then, experiential rationality states as follows:

Successful actional or interventive decision-making is rational to the degree that it utilises reasoning heuristics, where reasoning heuristics render rationality ‘grounded’ in those sensory-motor capabilities we employ during goal-

attainment.¹⁴

I now argue that experiential rationality is compatible with CC (and therefore presumably with post-structuralism more generally).

3. Merging Experiential Rationality with CC

We have seen how recent work in psychology suggests that rationality can be naturalised and instrumental. I have called this kind of rationality experiential rationality. Experiential rationality allows for some decisional freedom in that we can pursue variable goals (excluding aberrant goals, as per C2) and we may employ a variety of heuristics in attaining those goals. Experiential rationality is also fallible to the degree that we *qua* biological agents are fallible (goal-attainment via randomness or lucky guesses will however be arational). Experiential rationality also allows that we can be rational by degrees (Gigerenzer 2021) (I will not give an account of degrees of rationality here however). Most importantly, experiential rationality circumvents RV by ‘grounding’ – i.e. constraining – rationality, and this is done without invoking Cartesian-style transcendental norms. Rationality is ‘grounded,’ but not grounded.

As outlined in section 1, CC is averse to strict deterministic rules or norms for decision-making and to the idea that we can get an epistemic fix on meaning, knowledge, or truth to secure certainty. Pragmatists mostly share this aversion. CC however thinks that the proper alternative is RV, which, as I have argued, cannot account for how we *de facto* make decisions day-to-day. Meaning, knowledge, truth, and decision-making may be overdetermined (to varying degrees) by the world’s complexity, but this does not necessarily imply RV. A properly pragmatized notion of rationality can potentially succeed where RV fails. As argued, we regularly and reliably employ reasoning heuristics during successful goal-attainment despite the world’s evident complexity.

As far as I can tell, experiential rationality is compatible with the post-structuralist implications of Derrida’s semantics. To incorporate experiential rationality into CC, CCists need simply accept the following putative truism:

Decision-making is performed by Darwinian agents and is therefore constrained by biology to some degree.

This notion of being “constrained by” is what I have referred to as ‘grounded,’ and “biology,” in this context, refers specifically to agents’ sensory-motor capabilities

¹⁴ This suggests that non-human animals are capable of rationality to the extent that they employ experiential rationality (see however Okasha, 2018 ch. 6 for an overview of the debate around whether animals are capable of rationality).

employed during goal-attainment. Nothing here involves Cartesian norms or epistemic certainty. Experiential rationality's compatibility with CC is particularly noticeable if we rephrase the above claim in the following conditional form:

If decision-makers are Darwinian agents, then decision-making will be constrained by biology to some degree.

Experiential rationality is therefore a rationality specific to Darwinian agents like us. To say that decision-making is not constrained by biology to some degree is to contradict Darwinism. This is because, as mentioned, our cognitive faculties are evolved, and therefore fallible biological kludges. When it comes to decision-making, we are not free to defy our Darwinian constitution. We cannot decide to levitate or spontaneously combust, for example; or, if we did, the relevant decisional effort would fail, it would be unsuccessful. The choices (and resultant actions or interventions) we are capable of making are limited to what biological agents, like us, are *de facto* capable of. And the goals we pursue are limited to those that non-aberrant biological agents *de facto* pursue and can *de facto* attain (as per C2). It is in this sense that experiential rationality is naturalised and instrumental while also invoking constraints on decision-making. This contradicts RV where decision-making is unbounded. Thus, if CCists accept the putative truism stated in the above conditional, then they must give up RV.

Experiential rationality has notable elements of contextuality and provisionality. It is however not *radically* contextual and provisional in the way that RV is. Experiential rationality is contextual and provisional in the sense that it applies to Darwinian agents like us engaged in decision-making here and now,¹⁵ and not to agents *simpliciter* (recall section 2.2). Accepting experiential rationality does not commit us to universal claims about rationality. It is instead a more modest attempt to describe the way we engage in successful decision-making related to actions on and interventions in the world.

In sum, experiential rationality may be fairly easily incorporable into CC. CCists would have to give up RV, but this seems a relatively small price to pay all things considered.

I now engage with a possible objection. CCists may claim that a suitable alternative to both Cartesian rationality and RV is not experiential rationality, but rather what might be called *aporetic* rationality.

¹⁵ Including perhaps so-called higher non-human animals (recall footnote 14).

4. Possible Response: Aporetic Rationality

According to Oliver Human, CC “harbours a somewhat ironic dimension” (2016, 53 fn. 9). This is its endorsement of what Woermann calls an “aporetic logic,” a logic that embraces paradoxes and contradictions (2016, 67-81; see also Derrida 1988, 116). Aporetic logic is a kind of post-structural dialethism where one ostensibly deals with “uncertainty through the use of reason... defined as a wager between the calculable and the incalculable” (Human and Cilliers 2013, 34). Such a wager involves making decisions based on mutual considerations of antithetical or contradictory concepts. Woermann (2010) refers to this as “both/and” logic. We must think both yes and no, both random and predictable, both P and $\sim P$ (Woermann 2016, 118; see also Hurst 2010, 243-246). Here, logical contradiction can be the locus of epistemic illumination rather than a dead-end for inquiry (as so-called analytic philosophers might suppose). Aporetic logic, says Andrea Hurst, calls for a “new paradigm of complexity that enables us to think in terms of mutually negating opposites joined in relations of co-implication” (2010, 241). According to Preiser et al.,

the logic of [CC-style] thinking proposes a type of thinking that necessitates a double movement... It suggests that the concept and its counterpart (the *yes* and the *no*) are thought simultaneously (2013, 269 original emphases; see also Woermann 2016, 68-71).

CCists may claim that aporetic logic can be employed in successful decision-making instead of experiential rationality.

The problem is that CCists do not explain how exactly we are to simultaneously think in terms of “mutually negating opposites,” in terms of “the yes and the no.” It is questionable whether we can simultaneously think antithetical or contradictory concepts. Attempting to do so would presumably involve concurrently holding both concepts in conscious awareness. I am not sure if this can be done. A colleague who teaches introductory logic to undergraduates asks her students to think of a square circle. One or two students always claim to be capable of the task. Yet, on interrogation, they turn out to be either thinking of a square on top of a circle or thinking ‘square’ then ‘circle’ then ‘square’ then ‘circle’ etc. They are not thinking ‘square’ and ‘circle’ at the same time. The task is designed to show that certain things are a priori impossible. Now, ‘square’ and ‘circle’ are, of course, not antithetical concepts, but this anecdote does suggest that we just cannot think certain things. Antithetical concepts (P and $\sim P$) are plausibly even harder to think simultaneously than ‘square’ and ‘circle.’ Try to simultaneously think ‘square’ and ‘ \sim square’ for example. I predict certain failure (thinking of a shimmering or a faded square does not count). CCists however claim

to be, not only capable of simultaneously thinking P and \sim P, but also advise others to do so. The problem is that they do not explain how exactly this task is to be performed, nor how it is that they can do it while others cannot.

CCists also cannot claim that rationality applies in certain circumstances but not in others (recall section 1.3). At times, CCists distinguish between what they call *general* complexity and *restricted* complexity (approximately the standard distinction between genuinely complex systems versus merely complicated systems [see Poli 2013]). As Woermann et al. put it,

[i]n the restricted paradigm, complexity is treated as a problem that can be overcome (complex problems are understood as complicated problems); whereas in the general paradigm, complexity is treated as an ontological fact, which holds certain epistemological and cognitive implications for the manner in which we deal with complexity (2018, 5; see also Cilliers 2010).

It may then be tempting for CCists to state that rationality only applies when we deal with restricted complexity, while RV applies when we deal with general complexity.

Drawing such demarcations is however at odds with the implications of *différance*. According to Derrida, *différance* disrupts all (non-provisional/non-heuristic) distinctions. We need to isolate meaning to draw demarcations, and *différance* ruins all attempts to do so (recall section 1.1) (Derrida 1988, 116; Woermann 2016, 173-176; see also Human and Cilliers 2013). Post-structuralism disallows meaningful delineation between one domain and another (Woermann et al. 2018, 7-10); that is, meaningful delineation between general complexity and restricted complexity (see also Hurst 2010). Claiming that there are two separate domains – one amenable to rationality and the other to RV – violates post-structuralism’s own taboo on such demarcations. On the post-structuralists’ own account, *différance* should render rationality as radically contingent and contextual as RV. Hence, the need to give up RV.

Conclusion

According to Morin, rationality “never has the ambition to exhaustively hold the totality of reality in a logical system,” yet it is “our only trustworthy instrument of knowledge...” (2008, 47). I have argued along similar lines that there are no universal and exacting norms for decision-making, but we *qua* biological agents are nonetheless constrained in our decision-making practices by rationality properly pragmatized. I have called this experiential rationality, and it seems consistent with (at least, some of) the tenets of post-structuralism. It is therefore potentially incorporable into CC.

Ragnar Van der Merwe

Although CCist's criticisms of Cartesian rationality are on point, their alternative – RV – overemphasises the role of undecidability and freedom in our actional and interventive encounters with complexity. I have argued that we should instead think of rationality in this context as the successful utilisation of embodied heuristics. Doing so 'grounds' rationality in the sensory-motor capabilities we employ during goal-attainment. It also renders rationality responsible for the kind of successful actional and interventive decisions we make day-to-day despite the world's evident complexity.

References

- Ariely, D. 2008. *Predictably Irrational: The Hidden Forces that Shape our Decisions*. New York: Harper Collins.
- Bedau, M. and P. Humphreys. 2008. *Emergence: Contemporary Readings in Philosophy and Science*. Cambridge, Mass.: MIT Press.
- Bliss, R. and K. Trogon. 2021. "Metaphysical grounding." In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Online publication, URL = <https://plato.stanford.edu/archives/win2021/entries/grounding/>.
- Buss, D. M. 2019. *Evolutionary Psychology: The New Science of the Mind, 6th Edition*. New York: Routledge.
- Campbell, D. T. 1974. "Evolutionary Epistemology." In *The Philosophy of Karl Popper, Volume I*, edited by P. A. Schlipp, 413–459. Illinois: La Salle.
- Churchland, P. S. 1987. "Epistemology in the Age of Neuroscience." *The Journal of Philosophy* 84 (10): 544–553.
- Cilliers, P. 1998. *Complexity and Postmodernism: Understanding Complex Systems*. London: Routledge.
- Cilliers, P. 2000. "Knowledge, Complexity, and Understanding." *Emergence* 2 (4): 7–13.
- Cilliers, P. 2005. "Complexity, Deconstruction and Relativism." *Theory, Culture & Society* 22 (5): 255–267.
- Cilliers, P. 2010. "Difference, Identity and Complexity." In *Issues in Business Ethics: Complexity, Difference and Identity*, edited by P. Cilliers and R. Preiser, 1–18. London: Springer.
- Cilliers, P., W. van der Merwe, and J. Degennar. 2016. "Justice, Law and Philosophy: An Interview with Jacques Derrida." In *Paul Cilliers, Critical Complexity: Collected Essays*, edited by R. Preiser, 171–180. Berlin: De Gruyter.
- Cosmides, L. and J. Tooby. 1994. "Better than Rational: Evolutionary Psychology and the Invisible Hand." *The American Economic Review* 84 (2): 327–332.

- De Villiers-Botha, T. and P. Cilliers. 2010. "The complex 'I': The Formation of Identity in Complex Systems." In *Issues in Business Ethics: Complexity, Difference and Identity*, edited by P. Cilliers and R. Preiser. 19–38. Dordrecht: Springer.
- Dennett, D. C. 1995. *Darwin's Dangerous Idea*. New York: Simon & Schuster.
- Dennett D. C. 2009. "Darwin's 'Strange Inversion of Reasoning'." *Proceedings of the National Academy of Sciences* 106 (Suppl. 1): 10061–10065.
- Derrida, J. 1978. *Writing and Difference*. Chicago: University of Chicago Press.
- Derrida, J. 1981. *Positions*. Chicago: University of Chicago Press.
- Derrida, J. 1982. "Différance." In *Margins of Philosophy*, edited and translated by A. Bass, 1–28. Chicago: University of Chicago Press.
- Derrida, J. 1988. *Limited Inc*, edited by G. Graff, translated by S. Weber. Evanston: Northern Western University Press.
- Derrida, J. 1999. "Hospitality, Justice and Responsibility: A Dialogue with Jacques Derrida." In *Questioning Ethics: Contemporary Debates in Philosophy*, edited by R. Kearney and M. Dooley, 65–68. London: Routledge.
- Derrida, J. 2002. "Ethics and Politics Today." In *Negotiations: Interventions and Interviews, 1971–2001*, edited and translated by E. Rottenberg, 295–314. Stanford: Stanford University Press.
- Dosi, G., M. Faillo, and L. Marengo. 2021. "Beyond 'Bounded Rationality': Behaviours and Learning in Complex Evolving Worlds." In *Routledge Handbook of Bounded Rationality*, edited by R. Viale, 491–506. New York: Routledge.
- Edis, T. and M. Boudry. 2019. "Truth and Consequences: When Is It Rational to Accept Falsehoods?" *Journal of Cognition and Culture* 19 (1–2): 147–169.
- Foucault, M. 2001. *Essential Works of Foucault 1954–1984, Volume 3: Power*, edited by J. D. Faubion, translated by R. Hurley and others. New York: New Press.
- Gallagher, S. 2018. "Embodied Rationality." In *The Mystery of Rationality: Mind, Beliefs and the Social Sciences*, edited by G. Bronner and F. di Iorio, 83–94. Cham: Springer.
- Gigerenzer, G. 2008. *Rationality for Mortals: How People Cope with Uncertainty*. New York: Oxford University Press.
- Gigerenzer, G. 2021. "Embodied Heuristics." *Frontiers in Psychology* 12: 4243.
- Gigerenzer, G. and R. Selten. 2002. *Bounded Rationality: The Adaptive Toolbox*. Cambridge Mass.: MIT Press.
- Gigerenzer, G. and H. Brighton. 2009. "Homo Heuristicus: Why Biased Minds Make Better Inferences." *Topics in Cognitive Science* 1 (1): 107–143.

Ragnar Van der Merwe

- Gigerenzer, G. and T. Sturm. 2012. "How (Far) can Rationality be Naturalized?" *Synthese* 187 (1): 243–268.
- Godfrey-Smith, P. 1996. *Complexity and the Function of Mind in Nature*. Cambridge: Cambridge University Press.
- Godfrey-Smith, P. 2002. "Environmental Complexity and the Evolution of Cognition." In *The Evolution of Intelligence*, edited by R. J. Sternberg and J. Kaufman, 233–249. Mahwah: Lawrence Erlbaum.
- Goldman, A. I. 1970. *A Theory of Human Action*. Princeton: Princeton University Press.
- Gupta, A. D. 2021. "On the Bodily Basis of Human Cognition: A Philosophical Perspective on Embodiment." *Frontiers in Human Neuroscience* 15: 745095.
- Habermas, J. 1990. *Moral Consciousness and Communicative Action*. Cambridge, Mass.: MIT Press.
- Haselton, M. G., G. A. Bryant, A. Wilke, D. A. Frederick, A. Galperin, W. E. Frankenhuis, and T. Moore. 2009. "Adaptive Rationality: An Evolutionary Perspective on Cognitive Bias." *Social Cognition* 27 (5): 733–763.
- Human, O. 2016. "Potential Novelty: Towards an Understanding of Novelty without an Event." *Theory, Culture & Society* 32 (4): 45–63.
- Human, O. and P. Cilliers. 2013. "Towards an Economy of Complexity: Derrida, Morin and Bataille." *Theory Culture & Society* 30 (5): 24–44.
- Hurst, A. 2010. "Complexity and the Idea of Human Development." *South African Journal of Philosophy* 29 (3): 233–252.
- Jonsson, B. and C. von Hofsten. 2003. "Infants' Ability to Track and Reach for Temporarily Occluded Objects." *Developmental Science* 6 (1): 86–99.
- Kahneman, D. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus & Giroux.
- Kauffman, S. 2019. *A World Beyond Physics: The Emergence and Evolution of Life*. New York: Oxford University Press.
- Kenrick, D. T. and V. Griskevicius. 2013. *The Rational Animal: How Evolution Made Us Smarter than We Think*. New York: Basic Books.
- Knauff, M. and W. Spohn. 2021. *The Handbook of Rationality*. Cambridge Mass.: MIT Press.
- Ladyman, J. and K. Wiesner. 2021. *What is a Complex System?* New Haven: Yale University Press.
- Levine, S. 2019. *Pragmatism, Objectivity, and Experience*. Cambridge: Cambridge University Press.
- Mastrogiorgio, A. and E. Petracca. 2016. "Embodying Rationality." In *Model-Based Reasoning in Science and Technology: Logical, Epistemological and*

- Cognitive Issues*, edited by L. Magnani and C. Casadio, 219–237. Cham: Springer.
- Mastrogiorgio, A., T. Felin, S. Kauffman, and M. Mastrogiorgio. 2022. “More Thumbs than Rules: Is Rationality an Exaptation?” *Frontiers in Psychology* 13: 805743.
- Mercier, H. and D. Sperber. 2017. *The Enigma of Reason*. Cambridge, Mass.: Harvard University Press.
- Misak, C. J. 2014. “Language and Experience for Pragmatism.” *European Journal of Pragmatism and American Philosophy*. Online publication, URL = <https://doi.org/10.4000/ejppap.295>.
- Morin, E. 2008. *On Complexity*, translated by S. M. Kelly. Cresskill: Hampton Press.
- Newell, A., and H. A. Simon. 1972. *Human Problem Solving*. Englewood Cliffs: Prentice Hall.
- O’Connor, T. 2021. “Emergent Properties.” In *The Stanford Encyclopedia of Philosophy*, edited by E. N. Zalta. Online publication, URL = <https://plato.stanford.edu/archives/win2021/entries/properties-emergent/>.
- Okasha, S. 2018. *Agents and Goals in Evolution*. Oxford: Oxford University Press.
- Pinker, S. 2002. *The Blank Slate: The Modern Denial of Human Nature*. New York: Viking.
- Pinker, S. 2010. “The Cognitive Niche: Coevolution of Intelligence, Sociality, and Language.” *Proceedings of the National Academy of Sciences* 107 (2): 8993–8999.
- Pinker, S. 2021. *Rationality: What it is, why it Seems Scarce, why it Matters* (EPUB version). New York: Allen Lane.
- Poli, R. 2013. “A Note on the Difference Between Complicated and Complex Social Systems.” *Cadmus* 2 (1): 142–147.
- Preiser, R. and P. Cilliers. 2010. “Unpacking the Ethics of Complexity: Concluding Reflections.” In *Issues in Business Ethics: Complexity, Difference and Identity*, edited by P. Cilliers and R. Preiser, 265–287. Dordrecht: Springer.
- Preiser, R., P. Cilliers, and O. Human. 2013. “Deconstruction and Complexity: A Critical Economy.” *South African Journal of Philosophy* 32 (3): 261–273.
- Rolla, G. 2016. “Epistemic Immodesty and Embodied Rationality.” *Manuscrito* 39 (3): 5–28.
- Saussure, F. de. 1974. *Course in General Linguistics*. London: Fontana.
- Searle, J. R. 1977. “Reiterating the Differences: A Reply to Derrida.” *Glyph* 1 (1): 198–208.
- Simon, H. A. 1983. *Reason in Human Affairs*. Stanford: Stanford University Press.

Ragnar Van der Merwe

- Spellman, B. and S. Schnall. 2009. "Embodied Rationality." *Queen's Law Journal* 35 (1): 117–164.
- Thaler, R. H. and C. R. Sunstein. 2008. *Nudge: Improving Decisions about Health, Wealth, and Happiness*. New Haven: Yale University Press.
- Van der Merwe, R. 2021. "On Paul Cilliers' Approach to Complexity: Post-structuralism versus Model Exclusivity." *Interdisciplinary Description of Complex Systems* 19 (4): 457–469.
- Van der Merwe, R. Forthcoming. "Whewell's Hylomorphism as a Metaphorical Explanation for how Mind and World Merge." *Journal for General Philosophy of Science*, DOI: 10.1007/s10838-021-09595-x.
- Viale, R. 2021. *Routledge Handbook of Bounded Rationality*. New York: Routledge.
- Vlerick, M. and A. Broadbent. 2015. "Evolution and Epistemic Justification." *Dialectica* 69 (2): 185–203.
- Wilke, A. and P. M. Todd. 2010. "Past and Present Environments: The Evolution of Decision Making." *Psicothema* 22 (1): 4–8.
- Woermann, M. 2010. "Corporate Identity, Responsibility and the Ethics of Complexity." In *Issues in Business Ethics: Complexity, Difference and Identity*, edited by P. Cilliers and R. Preiser, 167–191. Dordrecht: Springer.
- Woermann, M. 2016. *Bridging Complexity and Post-Structuralism: Insights and Implications*. Cham: Springer.
- Woermann, M. and P. Cilliers. 2012. "The Ethics of Complexity and the Complexity of Ethics." *South African Journal of Philosophy* 31 (2): 447–464.
- Woermann, M., O. Human, and R. Preiser. 2018. "General Complexity: A Philosophical and Critical Perspective." *Emergence: Complexity and Organization* 20 (2): 1–18.

DISCUSSION NOTES/DEBATE

OBJECTING TO THE ‘DOESN’T JUSTIFY THE DENIAL OF A DEFEATER’ THEORY OF KNOWLEDGE: A REPLY TO FEIT AND CULLISON

Timothy KIRSCHENHEITER

ABSTRACT: In this paper, I explain Neil Feit and Andrew Cullison’s two proposed theories of knowledge, their initial No Essential Falsehood-Justifying Grounds account and their ultimate ‘Doesn’t Justify the Denial of a Defeater’ account. I then offer original counterexamples against both of these theories. In the process of doing so, I both explain Feit and Cullison’s motivation for jointly offering their theories and recount counterexamples that others have offered against various theories that assert that knowledge is justified, true belief plus some condition concerning essential reliance.

KEYWORDS: knowledge, false beliefs, Gettier problem, Neil Feit, Andrew Cullison, Ted Warfield

I. Introduction

In this paper, I explain Neil Feit and Andrew Cullison’s (2011) two proposed theories of knowledge, the No Essential Falsehood-Justifying Grounds account and their ultimate ‘Doesn’t Justify the Denial of a Defeater’ account. I then offer original counterexamples against both of these theories. In the process of doing so, I both explain Feit and Cullison’s motivation for jointly offering their theories and recount counterexamples that others have offered against various theories that assert that knowledge is justified, true belief plus some condition concerning essential reliance.

II. The No Essential Falsehoods Account of Knowledge and Criticisms of This Account

In *Epistemology* (2003), Richard Feldman offers the No Essential Falsehoods account of knowledge (NEF). It is as follows:

- S knows $p = df$
(i) p is true
(ii) S believes p

Timothy Kirschenheiter

(iii) *S* is justified in believing *p*

(iv) *S*'s justification for *p* does not essentially depend on any falsehood¹

This account simply tacks an extra condition onto justified, true belief theory. This condition is meant to account for Gettier cases that show that having a justified, true belief is not equivalent to having knowledge.²

NEF can be criticized in at least two ways. First, there are cases that intuitively show that this account is too broad. That is, there are instances of non-knowledge that NEF considers to be knowledge. Second, cases offered by various authors have shown that sometimes you can have knowledge that derives from false beliefs – even false beliefs upon which you essentially depend. In other words, there are instances of knowledge that NEF cannot account for. So, this account also proves to be too narrow.

I will return to knowledge from false beliefs later, but let's first focus on cases where NEF proves to be too broad. In order to demonstrate this point, Neil Feit and Andrew Cullison, writing together, offered the following counterexample against NEF:

Uncle George: It is common knowledge in Smith's office that George is a wise and honest man. George has told Smith that he, George, is an uncle. He has a 'World's Greatest Uncle' mug on his desk, and so on. On the basis of all of this evidence, Smith believes that George is an uncle. In this particular instance, however, George has been pretending to be an uncle. The twist is that George now really is an uncle, unbeknownst to him. His estranged sister just had a baby boy.³

Smith has a justified, true belief that George is an uncle. Furthermore, George is actually an uncle. So, Smith's justification for his belief does not essentially depend on a falsehood. So, on NEF, Smith knows that George is an uncle. Intuitively, however, Smith does not know that Smith is an uncle.

Consider another case:

DontKnowHeGot: You have a generally trustworthy coworker in your office named DontKnowHeGot, who gives you a great deal of evidence that he owns a Ford vehicle. He talks about his Ford frequently, he has a Ford keychain, he has a Ford tattoo on his lower back, and he even named his firstborn son Ford and his firstborn daughter Forda. On the basis of all of this evidence, you justifiably come to believe that DontKnowHeGot owns a Ford. However, DontKnowHeGot has been trying to deceive you. He believes that he does not own a Ford. However, unbeknownst to DontKnowHeGot, the rusted-out shell of an old truck in his backyard is actually a 1939 Ford truck.

¹ For an explication of this theory, see Feldman (2003, 25-36).

² For an explanation of Gettier cases, see Gettier (1963).

³ Feit and Cullison (2011, 289-290).

Objecting to the ‘Doesn’t Justify the Denial of a Defeater’ Theory of Knowledge

You have a justified, true belief that DontKnowHeGot owns a Ford. And your justification for this belief does not essentially depend on a falsehood. Rather, you are essentially depending on a true claim – DontKnowHeGot in fact owns a Ford. Intuitively, however, you do not know that DontKnowHeGot owns a Ford.

Uncle George and DontKnowHeGot both show that NEF is too broad. The account determines that both of these instances of non-knowledge count as knowledge. But intuitively this conclusion is incorrect in both cases.

III. Feit and Cullison’s Response to These Cases

Feit and Cullison attempt to save the NEF by offering a slightly edited version. This version is meant to account for Uncle George and other potential, similar cases, like DontKnowHeGot. They refer to this account as the No Essential Falsehood-Justifying Grounds theory of knowledge (NEFJG). This account is as follows:

S knows *p* = df

(i) *S* believes *p*

(ii) *p* is true

(iii) *S* is justified in believing *p*

(iv) no ground that is essential to *S*’s justification for *p* justifies *S* in believing a falsehood.⁴

Condition (iv) accounts for our intuitive judgments in both Uncle George and DontKnowHeGot. In Uncle George, Smith’s justification for his belief that George is an uncle also justifies him in believing many falsehoods, including the claim that George believes that he is an uncle. So, on the NEFJG, Smith does not know that George is an uncle.

Similarly, in DontKnowHeGot you are justified in believing the false claim that DontKnowHeGot believes that he owns a Ford. So, your belief that DontKnowHeGot owns a Ford does not pass the fourth condition for knowledge offered by Feit and Cullison. So, you do not know that DontKnowHeGot owns a Ford. So, NEFJG can account for our intuitive judgments in these cases.

IV. A Counterexample to the No Essential Falsehood-Justifying Grounds Account

Though NEFJG matches our intuitions in Uncle George and DontKnowHeGot, it is too narrow in other respects. That is, it leaves out genuine instances of knowledge. Consider the following case:

⁴ Feit and Cullison (2011, 291).

Timothy Kirschenheiter

Aunt Kathy: Imagine that I see my Aunt Kathy wearing a Seahawks Super Bowl XX championship ring. On the basis of this evidence, I form the belief “My Aunt Kathy is wearing a Seahawks Super Bowl XX championship ring.” And, on the basis of that claim, I form the belief “If my Aunt Kathy is wearing a Seahawks Super Bowl XX championship ring, then she is wearing a ring.”

This appears to be a genuine instance of knowledge. However, it does not meet the four conditions offered by Feit and Cullison above. While I have a justified, true belief, condition (iv) is not met. In other words, one of the grounds essential to my justification for believing the conditional also justifies me in believing a falsehood. My ground that says that “My Aunt Kathy is wearing a Seahawks Super Bowl XX championship ring” justifies me in believing that the Seahawks won Super Bowl XX.⁵ This belief, however, would be false. In fact, the Chicago Bears won Super Bowl XX. So, according to NEFJG, I do not have knowledge of the claim “If my Aunt Kathy is wearing a Seahawks Super Bowl XX championship ring, then she is wearing a ring.” However, it is intuitively clear that this belief is a genuine instance of knowledge. So, NEFJG fails.

One potential way to object to counterexamples against theories like NEFJG is to argue that there is some other nearby claim that is actually being essentially relied upon. Perhaps one might think that I am essentially relying on the fact that my Aunt Kathy is wearing a ring or the logical truth that if one is wearing a ring, then they must be wearing a ring. However, neither of those beliefs, even taken in conjunction, can get to the actual conditional that is the instance of knowledge in this case. Even if there are other, nearby beliefs that are needed to reach the conclusion, the conditional that is the object of my knowledge cannot be reached without the belief that my Aunt Kathy is wearing a Seahawks Super Bowl XX championship ring.⁶

V. Another Objection to the No Essential Falsehood-Justifying Grounds Account – Knowledge from False Beliefs

Another criticism of NEF that also applies to Feit and Cullison’s NEFJG involves knowledge from false beliefs. There are numerous examples that philosophers have given of knowledge from false beliefs, but let’s focus on a famous case from Ted Warfield (2005):

⁵ This counterexample assumes that I lack the knowledge of who won this Super Bowl from another source but also realize how strange and rare a championship ring for a non-champion would be.

⁶ Peter Murphy (2013) made a similar point in this journal when offering a counterexample involving a conditional claim against knowledge-from-knowledge.

Objecting to the ‘Doesn’t Justify the Denial of a Defeater’ Theory of Knowledge

Professor: Counting with some care the number of people present at his talk, Warfield reasons: ‘There are 53 people at my talk; therefore my 100 handout copies are sufficient.’ His premise is false. There are 52 people in attendance – he double-counted one person who changed seats during the count.⁷

Even though Warfield is essentially depending on a falsehood – here “there are 53 people at my talk” – intuitively Warfield still has knowledge of his conclusion that “my 100 handout copies are sufficient.” On both NEF and NEFJG, Warfield would *not* have knowledge of his conclusion. So, these accounts fail to give the correct answer in this case. They both treat this instance of knowledge as non-knowledge. So, they are both too narrow.

One could potentially object to Warfield’s Professor case by claiming that Warfield is actually depending on the claim that “there are about 53 people at my talk.” So, since there are actually about 53 people at his talk, he is not essentially depending on a falsehood in order to reach his conclusion. Instead, he is essentially depending on a nearby, true claim.⁸

The problem with this objection is that it is offering a counterfactual situation (distinct from the actually offered case) whereby the theory would give the correct answer. But this counterfactual does nothing to disprove the fact that the theory cannot account for the original case, as stated. This sort of objection is basically building a strawman, by altering the cases offered and asserting that the altered cases can be accounted for by the theory in question. Again, this does nothing to show whether the theory can account for the original case, as offered.

So I grant that, in Professor, *if* Warfield had depended on the claim that “there are about 53 people at my talk,” then NEF and NEFJG *could* account for Warfield’s knowledge of his conclusion. However, that is not the claim upon which Warfield essentially depends in Professor. Rather, he essentially depends on the claim that “there are 53 people at my talk.” And *even when* he depends on this claim, he still has knowledge of his conclusion that “my 100 handout copies are sufficient.” NEF and NEFJG cannot account for this. So, the theories fail.

VI. Feit and Cullison’s Attempt to Account for Knowledge from False Beliefs

Feit and Cullison offer a new theory in place of NEFJG in order to account for the sort of examples offered by Warfield and others. That is, they offer an account

⁷ Warfield (2005, 407-408). I edited the wording of this example in order to make it refer to Warfield, as he offers the example in the first-person.

⁸ Martin Montminy (2014) offers this sort of objection against examples meant to show that knowledge can come from false beliefs, arguing that there are nearby beliefs of the subject by which they gain inferential knowledge.

Timothy Kirschenheiter

meant to permit some instances of knowledge from false beliefs. They attempt to do this through the use of defeaters. They define a defeater as follows:

d is a defeater (with respect to evidence *e* for *p*) =df. *d* is a true proposition such that *e* justifies *p* but the evidence set that results from adding *d* does not justify *p*.⁹

And they offer the following theory, which they call the ‘Doesn’t Justify the Denial of a Defeater’ account of knowledge (DJDD):

S knows *p* = df

(i) *S* believes *p*

(ii) *p* is true

(iii) *S* is justified in believing *p*

(iv) no ground that is essential to *S*’s justification for *p* justifies *S* in believing the negation of a defeater¹⁰

In other words, condition (iv) says that when the denial of the justified falsehood serves as a defeater, then there is no knowledge. DJDD accounts for Professor, because if one added the claim that “there are *not* 53 people at the talk” to Warfield’s set of evidence, he would still be justified in believing his conclusion. Even when the denial of the justified falsehood (“there are 53 people at my talk”) is added to Warfield’s overall evidence, he is still justified in believing that he has enough copies of his handout.

VII. Two Counterexamples against the ‘Doesn’t Justify the Denial of a Defeater’ Account of Knowledge¹¹

Though DJDD handles Professor, there are potential counterexamples against it. I will now consider two of these counterexamples. The first counterexample is the standard sort of counterexample given against more basic no defeaters views. It does not seem as if Feit and Cullison’s more detailed no defeaters view can handle even this standard counterexample. Consider the following:¹²

Grabit: You see your student Tom Grabit stick a DVD in his coat pocket and sneak out of the library. You recognize Tom easily, given your many interactions with him. Meanwhile, Tom’s crime is reported to Tom’s mother in her room at a psychiatric hospital. And she replies that Tom didn’t do it. She claims that it was

⁹ Feit and Cullison (2011, 295).

¹⁰ Feit and Cullison (2011, 295).

¹¹ The only robust challenge to DJDD offered thus far comes from Stephen Hetherington (2016). But Hetherington objects in a very different way than I do in this paper. His focus is on whether Feit and Cullison are offering a fallibilist or infallibilist account. Concluding that they are offering an infallibilist account, Hetherington then argues against DJDD on this basis.

¹² This counterexample is adapted from Feldman (2003, 35-36).

Objecting to the ‘Doesn’t Justify the Denial of a Defeater’ Theory of Knowledge

his twin brother Tim. However, Tom does not actually have a twin. The mother is simply deluded. But you are unaware of all of this information involving Tom’s mother.

Intuitively, it is clear that you know that Tom Grabit stole the DVD from the library. But DJDD cannot account for this. You do not meet condition (iv). You have an essential ground (something like “the person stealing the DVD looks exactly like my student, Tom Grabit”) that justifies you in believing the negation of a defeater. The defeater here is the claim that “Tom’s mother says his twin brother Tim stole the DVD.” Given your evidence, you are justified in believing the negation of that defeater. You are justified in believing that “it is not the case that Tom’s mother says his twin brother Tim stole the DVD.” So, condition (iv) is not met. DJDD would say that you do not have knowledge in Grabit. This is clearly counterintuitive. So, we have good reason to reject DJDD.

Consider another counterexample, one meant to show DJDD does not account for all instances of knowledge from false belief:

Blind Warfield: Warfield is blind and asks one of his students to count how many people are at his talk. The student tells him that he counted 53. On the basis of the student’s claim, Warfield concludes that the 100 copies of his talk are more than sufficient. However, the student accidentally miscounted. There are actually 52 people at the talk.

Does blind Warfield know that he has enough copies of his talk? Intuitively, I say yes. However, although he has a justified, true belief, condition (iv) on DJDD is not met. The denial of his justified, false belief that there are 53 people at the talk serves as a defeater. Given that he did not do the counting himself, he lacks the sort of evidence that would allow for him to conclude that the number is around 53, though not 53 exactly.

Perhaps one could object that Warfield can reasonably conclude that the student’s count was slightly off, but he was generally in the vicinity of the right count. In other words, he still has good reason to believe that the count is around 53. Though I think that this sort of objection is changing our actual scenario to a counterfactual one (as explained above), let’s avoid this concern by filling out the scenario a bit more so as to account for this worry. Consider the following case:¹³

Sometimes Prankster: Imagine that the student that a blind Warfield asks to count the number of people in the audience is a bit of a prankster. Now, he does not always play pranks on Warfield, but he does so occasionally. Imagine that

¹³ This counterexample can also be used against NEF and NEFJG. And it is immune from the objection offered against Professor that claims that there is some other nearby claim that Warfield is essentially relying on.

Timothy Kirschenheiter

when the student first reports the total as 53, Warfield is just over the required level of justification for his belief that there are 53 people at his talk. This lower level of justification is due to the student's past trickery. Warfield uses this ground to conclude that his 100 copies are more than sufficient. Of course, there are only 52 people at the talk. Now, if the denial of Warfield's justified, false belief that there are 53 people were added to his set of evidence, Warfield would no longer be justified in believing the student's count at all, as the chance that he is playing a prank on Warfield has gone up significantly. So, his level of justification, previously hovering just above sufficient justification, now falls below that threshold. However, in actuality the student was not tricking him and made a small, innocent mistake, causing his count to be off by one.

In this scenario, intuitively, blind Warfield knows that he has enough copies of his talk.¹⁴ Yet DJDD would not consider this case to be an instance of knowledge. The denial of the justified falsehood – the claim that “there are 53 people at my talk” – serves as a defeater of blind Warfield's belief that “my 100 handout copies are more than sufficient.” But this result is counterintuitive. So, we have good reason to reject DJDD. Yet again, another potential candidate for a correct conceptual analysis of knowledge is toppled.

References

- Feit, Neil and Andrew Cullison. 2011. “When Does Falsehood Preclude Knowledge?” *Pacific Philosophical Quarterly* 92(3), 283-304.
- Feldman, Richard. *Epistemology*. 2003. Upper Saddle River, NJ: Prentice Hall.
- Gettier, Edmund. 1963. “Is Justified True Belief Knowledge?” *Analysis* 23(6), 121-123.
- Hetherington, Stephen. 2016. “Understanding Fallible Warrant and Fallible Knowledge: Three Proposals.” *Pacific Philosophical Quarterly* 97(2), 270–282.
- Klein, Peter. 2008. “Useful Falsehoods.” In *Epistemology: New Essays*, edited by Quentin Smith, 25-61. New York: Oxford University Press.
- Montminy, Martin. 2014. “Knowledge Despite Falsehood.” *Canadian Journal of Philosophy* 44 (3/4), 463-475.
- Murphy, Peter. 2013. “Another Blow to Knowledge from Knowledge.” *Logos & Episteme*, 4(3), 311-317.

¹⁴This intuition is supported by the fact that we can assume that Warfield already factored in the likelihood that the student was tricking him before he decided to trust the student. In other words, the likelihood of his being tricked is already factored into his justification. So, if he is justified in the first place, the fact that his student miscounted should not harm Warfield's knowledge in this case.

Objecting to the 'Doesn't Justify the Denial of a Defeater' Theory of Knowledge
Warfield, Ted. 2005. "Knowledge from Falsehood." *Philosophical Perspectives*
19(1), 405-416.

ON THE PERSISTENCE OF ABSOLUTE METAPHYSICS

Gustavo PICAZO

ABSTRACT: Greenwood (2019) casts doubts upon whether a certain view about social groups (the view that social groups persist throughout changes in their membership, by virtue of the maintenance of their structure or function) is a fundamental metaphysical truth about social groups, rather than a theoretical truth about some or many social groups. In this note, I introduce a distinction between absolute and relative metaphysics, and argue that there are no ‘fundamental metaphysical truths’ (as Greenwood conceives of them) at all. If there is one thing that should not persist here, it is absolute metaphysics.

KEYWORDS: John D. Greenwood, absolute metaphysics, relative metaphysics, metaphysics of social groups, metaphysical truth, theoretical truth

John D. Greenwood (2019) evaluates ‘the common view that social groups persist throughout changes in their membership, by virtue of the maintenance of their structure and/or function’ (Abstract). He argues that ‘Despite the initial plausibility of this claim, there are reasons to doubt that this is a *metaphysical truth* about social groups, rather than a *theoretical truth* about some or many social groups’ (§I, my italics). Greenwood’s argument is based on two fictional counterexamples: ‘the Mooseville College Philosophy Department’ and a motorcycle club called ‘The Ravens.’ After a brief discussion of them, with special emphasis on how the members of these groups see themselves, Greenwood concludes that ‘continuity of structure and/or function is neither sufficient nor necessary for the persistence of social groups’ (§II).

Having done that, Greenwood goes on to consider one possible objection to his argument: what the members of these groups would say about themselves might be different from what neutral observers would say. However, he remarks, ‘Two reasonable responses suggest themselves’ to such an objection:

One is that it all depends upon *theoretical explanatory considerations*, as to whether one has to appeal to compositional or structural/functional similarities or differences to explain continuities or discontinuities in earlier and later behavior. The second is that there is no fact of the matter, since *our judgment in these matters depends upon the subjective weight we place* on continuity of composition versus continuity of

structure and/or function. But neither response supports the view that it is a *fundamental metaphysical truth* that social groups persist throughout changes in their membership, or that social groups persist because of continuities of structure/function. (§II, my italics)

In this discussion note, I do not want to focus on the question of the persistence of social groups, but on the contrast that Greenwood draws between a theoretical truth and a metaphysical truth (which he also calls ‘a fundamental metaphysical truth’ [§II, just quoted] and ‘a fundamental truth about the metaphysics of social groups’ [§I]). From what we have just read, it seems clear that, for Greenwood, neither ‘theoretical explanatory considerations’ nor ‘the subjective weight we place on continuity of composition versus continuity of structure and/or function’ pertains to fundamental metaphysical truths about social groups. Thus, it appears that such truths, as Greenwood conceives of them, lie outside the scope of what can be determined by means of theoretical or subjective considerations. And this being so, we must ask ourselves: how could metaphysical truths about social groups be determined if not by reference to theoretical or subjective considerations? Indeed, how could such truths come to be known, stated, or even glimpsed if not by reference to *human* considerations of one kind or another? The answer is, of course, that they could not.

I welcome Greenwood’s doubts about the view that it is a fundamental metaphysical truth that social groups persist through changes in their membership or because of continuities of structure/function. But I would like to invite him to extend such doubts to any view about social groups – in fact, to any view whatsoever. To that end, I suggest we distinguish between ‘absolute’ (or ‘fundamental’) metaphysical claims and ‘relative’ (or ‘local’) ones. Both absolute and relative metaphysical claims concern matters of ontology, such as the repertoire of existing objects of a particular kind or the existence and persistence conditions for those objects. However, the former are meant to hold unrestrictedly, while the latter are restricted to a particular fragment of discourse at a given time. Thus, absolute metaphysical claims attempt to describe the ontology of ‘the world in itself’, while relative metaphysical claims simply address the ontology of a particular domain of knowledge at a particular point in time (or of a particular theory, viewpoint, etc). Applying this distinction, a ‘theoretical truth’ about the existence or persistence conditions of social groups within a particular theory (or with respect to our current best social science) will be regarded as a *local* metaphysical truth, that is, a metaphysical truth relative to that theory (or to our current best social science).

In Picazo (2021a, §6.5), I have elaborated on a distinction similar to that between absolute and relative metaphysics, and in (Picazo 2021b, 2021c, 2021d), I have discussed at length a major philosophical preconception (semantic Platonism) that leads to the neglect of such distinctions. On reflection, the idea that there are absolute metaphysical truths – ie metaphysical truths that transcend any human consideration – is easily seen to be untenable. Hence, the claim that social groups persist through changes in their membership (or because of continuities of structure/function) can be ruled out as a fundamental metaphysical truth, simply because there are no such truths. But we could still hold that the claim is true of our current best social science (on the basis of, among other things, theoretical explanatory considerations) or of a particular viewpoint (depending on the subjective weight we place on continuity of composition versus continuity of structure and/or function). If there is one thing that should not persist here, at least as a respectable academic endeavour, it is absolute metaphysics.¹

References

- Greenwood, J. D. 2019. “On the Persistence of Social Groups.” *Philosophy of the Social Sciences* 50 (1): 78–81.
- Picazo, G. 2021a. “New Foundations (Natural Language as a Complex System, or New Foundations for Philosophical Semantics, Epistemology and Metaphysics, Based on the Process-Socio-Environmental Conception of Linguistic Meaning and Knowledge).” *Journal of Research in Humanities and Social Science* 9 (6): 33–44.
- Picazo, G. 2021b. “The Long Shadow of Semantic Platonism, Part I: General Considerations.” *Philosophia* 49 (4): 1427–1453.
- Picazo, G. 2021c. “The Long Shadow of Semantic Platonism, Part II: Recent Illustrations.” *Philosophia* 49 (5): 2211–2242.
- Picazo, G. 2021d. “The Long Shadow of Semantic Platonism, Part III: Additional Illustrations, from a Collection of Classic Essays.” *Disputatio. Philosophical Research Bulletin* 10 (17): 19–49.

¹ In the preparation of this paper, I received help from Samuel Cuello Muñoz, Daniel García Simón, Peter Kingston and *Proof-Reading-Service.com*.

DEFENDING JOINT ACCEPTANCE ACCOUNTS OF GROUP BELIEF AGAINST THE CHALLENGE FROM GROUP LIES

Lukas SCHWENGERER

ABSTRACT: Joint acceptance accounts of group belief hold that groups can form a belief in virtue of the group members jointly accepting a proposition. Recently, Jennifer Lackey (2020, 2021) proposed a challenge to these accounts. If group beliefs can be based on joint acceptance, then it seems difficult to account for all instances of a group telling a lie. Given that groups can and do lie, our accounts of group belief better not result in us misidentifying some group lies as normal assertions. I argue that Lackey's argument is not decisive. The cases she proposes as challenges for joint acceptance accounts can be dealt with in the joint acceptance framework. I present two different readings of Lackey's central case, showing that in both readings Lackey's example of a problematic group lie should not be identified as a lie, but rather as an epistemic mistake by the group. What kind of mistake the group makes depends on the exact reading of Lackey's case, but either way the group is not telling a lie.

KEYWORDS: group lies, group belief, joint acceptance, Jennifer Lackey

Introduction

Joint acceptance accounts of group belief (e.g. Gilbert (1989, 1994, 2014)) hold that groups can form a belief in virtue of the group members jointly accepting a proposition. These accounts are well equipped to explain why and how group beliefs can differ from the beliefs that individual members have. According to joint acceptance accounts a group might have a belief that p , even if no single individual member has the belief that p . Recently, however, Jennifer Lackey (2020, 2021) proposed a challenge to these accounts. If group beliefs can be based on joint acceptance, then it seems difficult to account for all instances of a group telling a lie. Given that groups can and do lie, our accounts of group belief better not result in us misidentifying some group lies as normal assertions. Not only for purely theoretical reasons, but also because our theoretical framework ought to help us with social, moral and practical issues. We want to hold groups accountable for their lies, so we better identify group lies correctly. Hence, if Lackey is right, we should abandon joint acceptance accounts of group belief.

My aim is to argue that Lackey's argument is not decisive. The cases she proposes as challenges for joint acceptance accounts can be dealt with in the joint acceptance framework. The paper is structured as follows: I start with the joint acceptance account and Lackey's argument against it. I then present two different readings of Lackey's central case, showing that in both readings Lackey's example of a problematic group lie should not be identified as a lie, but rather as an epistemic mistake by the group. What kind of mistake the group makes depends on the exact reading of Lackey's case, but either way the group is not telling a lie.

Joint Acceptance Accounts of Group Belief

The guiding idea of joint acceptance accounts of group belief is that groups form beliefs by their members deciding together what to believe. And they can decide to believe that p , even when no individual member believes p . A board of directors might jointly accept that Maggie is the best candidate for a job, even though no single member believes that to be the case. Perhaps some members have ranked Maggie as the second-best candidate, and others ranked her as the third. But nevertheless, Maggie might be the best compromise candidate for the group, so the members jointly accept that Maggie is the best candidate for the job.¹ This sort of case can be captured nicely by identifying the group belief with something that the group members have agreed on – something that they have *jointly accepted*. This is the basis for the conception of joint acceptance account I am working with. Of course, these accounts are not always spelt out in terms of 'acceptance.' Gilbert speaks of joint commitments (Gilbert 1989, 1994, 2014) rather than joint acceptances. But for my purpose I bundle theories that follow this guiding idea as joint acceptance theories. The bundle includes accounts by Gilbert (1989, 1994, 2014), Tuomela (1992) and Tollefsen (2009), who are the primary targets for Lackey's criticism. The details of the accounts do not matter much for my purpose. The important part is merely the role of jointly accepting that p as the cornerstone in forming a group belief. However, not every single group member has to be part of the joint acceptance. Only *operative* members are required. In many groups not everyone is part of the decision-making process. Some members have a say and some do not. The workers at a local Apple store are part of Apple, but they do not decide what Apple intends or believes. Only a select few people at the top of the company do. 'Operative members' is therefore introduced as a technical term picking out those members of the group that are relevant for the group's decision-

¹ For similar arguments see Gilbert (1989), Schmitt (1994), Tollefsen (2009).

making. These are also the members that can determine group beliefs by joint acceptance.

With this picture of a joint acceptance account in place I can proceed to Lackey's challenge for joint acceptance proponents. This challenge is based on cases of group lies.

The Challenge from Group Lies

Groups lie. There is not only a theoretical option for groups to lie, but groups have lied in the past. Of course, not always and all the time, but sometimes with large and unwelcome consequences. Perdue Pharma claimed that less than 1% of patients become addicted to their opioid painkiller as part of their marketing campaign (Meier 2018). This number was not only false, but Perdue Pharma knew that it was false. Perdue Pharma lied and as a result those painkillers were widely prescribed and lead to many people's addiction to painkillers.

Given that groups can lie, good accounts of group belief have to be suitable to identify group lies as group lies. Accounts of belief play this role because belief is part of a plausible account of lying that Lackey works with.²

A lies to *B* if and only if (1) *A* states that *p* to *B*, (2) *A* believes that *p* is false, and (3) *A* intends to be deceptive to *B* with respect to whether *p* in stating that *p*.

(2) is a belief condition for lies. Hence, the account of group belief influences whether (2) is satisfied or not in case of a potential group lie. Only if the group believes that *p*, the group can lie by claiming that not-*p*. Lackey's strategy is to use this connection to show that joint acceptance accounts of group beliefs identify some cases as normal assertions, even though we intuitively take the cases to be group lies. The paradigmatic case is the following:

TOBACCO COMPANY Philip Morris, one of the largest tobacco companies in the world, is aware of the massive amounts of scientific evidence revealing not only the addictiveness of smoking, but also the links it has with lung cancer and heart disease. While the members of the board of directors of the company believe this conclusion, they all jointly agree that, because of what is at stake financially, the official position of Philip Morris is that smoking is neither highly addictive nor detrimental to one's health, which is then published in all of their advertising materials. (Lackey 2020, 195)

Intuitively Philip Morris lies, says Lackey. But the joint acceptance account of group belief entails that the group is not lying at all. Hence, the joint acceptance account has to be false.

² And has independently argued for in Lackey (2013).

To see why the joint acceptance account gives us this result let me consider the conditions for lying again. Philip Morris lies here if and only if (1) the group states that smoking is not detrimental to one's health to its consumers, (2) believes that 'smoking is not detrimental to one's health' is false, and (3) intends to be deceptive to the consumers with respect to whether smoking is not detrimental to one's health in stating that smoking is not detrimental to one's health. But if the joint acceptance account is true, then (2) is not satisfied. The board of directors – the operative members of Philip Morris – jointly accept that smoking is not detrimental to one's health. And if joint acceptance determines group belief, then Philip Morris believes that smoking is not detrimental to one's health. Philip Morris just asserts what it believes. And asserting what one believes is not a lie. What we end up with is an intuition that Philip Morris lies and the joint acceptance based result that Philip Morris does not lie. Only one of these can be right and the other has to go. Hence, we should drop the joint acceptance account of group belief (Lackey 2020, 196-197).

There is little room to resist that joint acceptance accounts entail that the group is not lying in TOBACCO COMPANY. The case stipulates joint acceptance in a way that results in group belief under the joint acceptance accounts. Nevertheless, there is room to argue that the joint acceptance based result is correct. TOBACCO COMPANY is not a case of a group lie. To go this route, one needs to propose a different explanation of the intuition that Philip Morris is doing something blameworthy that we want to hold them accountable for.

Defending Joint Acceptance

A defence of joint acceptance accounts against Lackey's argument cannot merely claim that Lackey's proposed intuition is wrong. I need to explain why one wants to blame Philip Morris in TOBACCO COMPANY, if not for lying. The intuition that Philip Morris is doing something improper is hard to deny, so I need to provide a different story of what exactly is going wrong. My suggestion is that Philip Morris does something epistemically improper. The reason why we want to blame Philip Morris is that the group commits an epistemic mistake – and does so intentionally. This epistemic mistake is what we want to hold Philip Morris accountable for. Identifying the epistemic mistake involved depends on how exactly the case is understood. Hence, I discuss two different readings that lead to two different kinds of epistemic mistakes. Both are ways of forming epistemically improper beliefs that explain why we have the intuition that something bad is going on in TOBACCO COMPANY.

To distinguish the two kinds of improper belief I use the concept of epistemic expectations from Goldberg (2018). These are epistemic expectations one has towards other agents in a community. Goldberg distinguishes two kinds of these expectations: basic epistemic expectations and non-basic epistemic expectations. The former are based on an entitlement to expect other people to use reliable belief-forming processes and an entitlement to expect that other people update their beliefs appropriately given newly acquired beliefs or evidence. I can expect other members in my community to form their beliefs on reliable methods rather than, say, wild guessing. And I can also expect other members in my community to be at least minimally coherent.

Non-basic epistemic expectations are primarily about the evidence we expect an agent to have in a particular situation. This is best illustrated by pointing to the phenomenon of normative defeat. Take Kornblith's (1983) example of a physicist who believes his pet theory. Suppose that physicist could easily come across counterevidence to this theory, but whenever there is a chance for counterevidence he refuses to engage with the source of that potential counterevidence. When there is a talk that might contain counterevidence he does not attend. If a journal article might contain counterevidence he does not read that article. There is a clear sense in which the physicist is doing something epistemically improper. His way of gathering evidence is flawed, such that he does not have evidence that he should have. The community expects from a physicist that they look for available evidence, but this physicist violates our expectations. He does so to a degree at which he loses justification for his belief. He is not justified, because the evidence he should have constitutes a normative defeater. This is exactly what Goldberg has in mind when he talks about non-basic epistemic expectations: expectations about the evidence that someone should have (Goldberg 2016, 2018).

To my knowledge Lackey has not explicitly endorsed these two kinds of expectations. However, she does accept normative defeat in other contexts (e.g. Lackey (2005)), so the general idea of epistemic expectations that are relevant for evaluating epistemic agents is something that Lackey should accept. With epistemic expectations in my toolset I can now proceed with the two different readings of TOBACCO COMPANY. The first reading will involve basic epistemic expectations, and the second reading will involve non-basic epistemic expectations. In both interpretations the group fails to satisfy a relevant epistemic expectation. Therefore, Philip Morris holds an improper belief, but does not lie.

The First Reading

TOBACCO COMPANY includes the stipulation that Philip Morris is aware of the massive amounts of scientific evidence about the health effects of smoking. How exactly we read the case depends on the interpretation of Philip Morris being aware of that scientific evidence. The first option is to accept that Philip Morris has this scientific evidence as part of the group's evidence. Nevertheless, the group forms the belief that smoking is safe by joint acceptance.

Looking at the basic epistemic expectations of the group it is easy to see that they are violated. Even if we stipulate that Philip Morris is generally reliable, the coherence requirement is violated. The beliefs of agents are expected to be appropriately updated based on the evidence agents have. The group has evidence about the detrimental health effects of smoking, but does not update the group's belief accordingly. Hence, the basic epistemic expectation is not satisfied. The belief is epistemically improper. This is the source of the intuition that Philip Morris is doing something wrong and blameworthy in TOBACCO COMPANY, according to the first reading. The community expects agents to have a certain degree of coherence between evidence and beliefs. Philip Morris does not have that coherence, so the community should hold Philip Morris accountable for the improper epistemic practices. Even though the group is not lying, the group is still acting in a way that it ought not to. Moreover, Philip Morris acts in a way that might be bad for the community overall and therefore should be minimized and sanctioned. We are entitled to hold them accountable to a standard set by the basic expectation partially because that is required for our practice of testimony. Testimony cannot function well if we cannot expect other people to be minimally coherent regarding their beliefs and evidence.

I have now argued that the first reading – that the group has the scientific evidence as part of their body of evidence – leads to a violation of basic epistemic expectations by Philip Morris. This violation is blameworthy and the group should be held accountable for it. Hence, the intuition that Philip Morris is doing something wrong is explained, but now identified as an intuition caused by the group acting epistemically improper, not by lying.

The Second Reading

The second reading understands Philip Morris being aware of scientific evidence differently. One can also read it as the individual members of Philip Morris having the scientific evidence as part of their individual evidence, without Philip Morris as a group having that piece of evidence. This option is only available if the group evidence is not determined by the evidence the individual members have. A joint

acceptance account of group evidence as proposed by Schmitt (1994), Hakli (2011) or Schwengerer (2021) is an option that seems a good fit here. Joint acceptance for group belief goes well with joint acceptance for group evidence. The reading then goes as follows: the group jointly accepts that there either are no relevant scientific studies about tobacco's health effects, or that the studies are unreliable. They do so for financial reasons, but that is no obstacle to group evidence under a joint acceptance account. The group evidence is now compatible with the group's belief that smoking is safe, because they have no reason to believe otherwise. Hence, there is no internal inconsistency in the group in this second reading. The group fulfils its basic epistemic expectations. However, the group can still be criticized with regard to non-basic expectations. This is exactly what Schwengerer (2021) suggests to deal with problematic consequences of arbitrary justification in the joint acceptance accounts of group evidence. Just like individual agents, groups are under normative expectations about the evidence they should have in a particular situation. Groups can fail to satisfy these expectations when their evidence does not match the evidence the group ought to have. The group can lack evidence it should have, or have evidence it should not have. In the second reading of TOBACCO COMPANY the group lacks evidence it should have. The group should have these scientific studies as part of their evidence. It should have that evidence partly because we expect tobacco companies to know about the safety of their products, and partly because the individual group members know about the studies. The studies are easily accessible for the group, but nevertheless the group does not jointly accept the studies as evidence. Hence, the group fails to fulfil its non-basic epistemic expectations. This is what we blame Philip Morris for. It is not a lie, it is a failure to fulfil the non-basic epistemic expectations the community has towards the group.

Conclusion

I have shown in two different interpretations of Lackey's case against joint acceptance accounts of belief that her argument is not decisive. Proponents of joint acceptance accounts can make a reasonable case that TOBACCO COMPANY is not a group lie, but a form of an epistemic mistake. The group does not fulfil its epistemic expectations. In the first reading the group fails to satisfy basic epistemic expectations, in the second reading non-basic epistemic expectations. Both are failures that we want to hold the group accountable for. But they are not lies. This way the joint acceptance accounts can capture why we intuitively think there is something wrong about the group's actions in TOBACCO COMPANY, but can explain that intuition in a way that is compatible with joint acceptance proposals

Lukas Schwengerer

for group belief. This does not entail that groups cannot lie, but merely that cases that are put forward by Lackey against joint acceptance accounts can be dealt with. Other forms of group lies in which groups jointly agree that *p* and then claim that non-*p* were no problem to begin with.

References

- Gilbert, Margaret. 2014. *Joint Commitment*. Oxford: Oxford University Press.
- . 1989. *On Social Facts*. London: Routledge.
- Gilbert, Margaret. 1994. "Remarks on Collective Belief." In *Socializing Epistemology: The Social Dimensions of Knowledge*, edited by Frederik F. Schmitt, 111-134. Lanham, MD: Rowman and Littlefield.
- Goldberg, Sanford. 2016. "On the epistemic significance of evidence you should have had." *Episteme*, 449–470.
- . 2018. *To the best of our knowledge: social expectations and epistemic normativity*. Oxford: OUP.
- Hakli, Raul. 2011. "On Dialectical Justification of Group Beliefs." In *Collective Epistemology*, edited by Hans Bernhard Schmid, Daniel Sirtes and Marcel Weber, 119-153. Frankfurt: Ontos.
- Kornblith, Hilary. 1983. "Justified Belief and Epistemically Responsible Action." *Philosophical Review*, 33-48.
- Lackey, Jennifer. 2013. "Lies and Deception: An Unhappy Divorce." *Analysis*, 236-248.
- . 2020. "Group Belief: Lessons from Lies and Bullshit." *Aristotelian Society Supplementary Volume*, 185-208.
- . 2005. "Memory as a Generative Epistemic Source." *Philosophy and Phenomenological Research*, 636-658.
- . 2021. *The Epistemology of Groups*. Oxford: Oxford University Press.
- Meier, Barry. 2018. *Pain Killer: An Empire of Deceit and the Origin of America's Opioid Epidemic*. New York: Random House.
- Schmitt, Frederick F. 1994. "The Justification of Group Beliefs." In *Socializing Epistemology: The Social Dimensions of Knowledge*, edited by Frederick F. Schmitt, 257-287. Lanham, MD: Rowman and Littlefield.
- Schwengerer, Lukas. 2021. "Defending Joint Acceptance Accounts of Justification." *Episteme*. doi:10.1017/epi.2020.55.
- Tollefsen, Deborah. 2009. "Wikipedia and the epistemology of testimony." *Episteme*, 8-24.
- Tuomela, Raimo. 1992. "Group Belief." *Synthese*, 285–318.

NOTES ON THE CONTRIBUTORS

Timothy Kirschenheiter is a Visiting Assistant Professor of Philosophy at Oakland University. He recently successfully defended his dissertation in order to receive his doctorate from Wayne State University. His research interests include ethics, bioethics, epistemology, and the philosophy of religion. Contact: kirschenheiter@oakland.edu.

Ryan Miller is a Swiss National Science Foundation Doc.CH Fellow at the University of Geneva. His work focuses on the metaphysics of science, especially the fundamental mereology and interpretation of quantum mechanics, but also issues related to computation. Ryan did previous graduate work at the University of Saint Andrews, where he developed an interest in formal and social epistemology. Contact: Ryan.Miller@unige.ch.

Rogelio Miranda Vilchis is a Ph.D. in Philosophy of Science from the National Autonomous University of Mexico (2019) and a research candidate for the National System of Researchers of Mexico (SNI). His main areas of specialization are the methodology and epistemology of science and philosophy. His recent work focuses on metaphilosophy, conceptual engineering, the epistemology of disagreement, and how cognitive biases influence philosophical expert performance. Amongst most recent publications are „Holding Points of View Does Not Amount to Knowledge” (*Ratio*, 2022), „The Place of Discourse in Philosophy as a Way of Life” (*Metaphilosophy*, 2022), and „More than Merely Verbal Disputes” (*Metaphilosophy*, 2021). Contact: rogeliovmv0101@gmail.com.

Gustavo Picazo obtained his PhD at the London School of Economics and is currently Associate Professor of Logic and Philosophy of Science at the University of Murcia (Spain). His research interests include philosophy of logic, philosophy of language, metaphysics and Zen philosophy. His publications include “Five observations concerning the intended meaning of the intuitionistic logical constants” (*Journal of Philosophical Logic*, 2000), “Truths and Processes” (*Philosophia*, 2014), “The Long Shadow of Semantic Platonism (I), (II)” (*Philosophia*, 2021) and *Kairos zen: el poder de mirar y hacer* (Urano, 2018, in Spanish). More information is available on his website: <http://webs.um.es/picazo/>. Contact: picazo@um.es.

Lukas Schwengerer is currently the primary investigator of the *Collective Self-Knowledge* project funded by the Deutsche Forschungsgemeinschaft (project number: 462399384) at the University of Duisburg-Essen. He received his PhD at the University of Edinburgh in 2018 for work on a transparency account of self-knowledge. His research is located at the intersection of philosophy of mind and epistemology, with a particular interest in self-knowledge and how anti-individualistic approaches in the philosophy of mind impact epistemological question. His current focus includes epistemological questions in the context of social groups, and research on the transformation of our epistemic practices due to modern technology. His work has appeared in journals such as *Erkenntnis*, *Episteme* and *Review of Philosophy and Psychology*. Contact: lukas.schwengerer@uni-due.de.

Ragnar van der Merwe is a PhD student at the University of Johannesburg. His research interests are in the philosophy of science, pragmatism, and complexity science. He is particularly concerned with the nature of scientific progress and how it relates to the notion of truth. The paper in this volume was produced as part of the John Templeton Foundation funded project ‘Increasing complexity: The first rule of evolution?’ Contact: ragnarvdm@gmail.com.

NOTES TO CONTRIBUTORS

1. Accepted Submissions

The journal accepts for publication articles, discussion notes and book reviews.

Please submit your manuscripts electronically at: logosandepisteme@yahoo.com. Authors will receive an e-mail confirming the submission. All subsequent correspondence with the authors will be carried via e-mail. When a paper is co-written, only one author should be identified as the corresponding author.

There are no submission fees or page charges for our journal.

2. Publication Ethics

The journal accepts for publication papers submitted exclusively to *Logos & Episteme* and not published, in whole or substantial part, elsewhere. The submitted papers should be the author's own work. All (and only) persons who have a reasonable claim to authorship must be named as co-authors.

The papers suspected of plagiarism, self-plagiarism, redundant publications, unwarranted ('honorary') authorship, unwarranted citations, omitting relevant citations, citing sources that were not read, participation in citation groups (and/or other forms of scholarly misconduct) or the papers containing racist and sexist (or any other kind of offensive, abusive, defamatory, obscene or fraudulent) opinions will be rejected. The authors will be informed about the reasons of the rejection. The editors of *Logos & Episteme* reserve the right to take any other legitimate sanctions against the authors proven of scholarly misconduct (such as refusing all future submissions belonging to these authors).

3. Paper Size

The articles should normally not exceed 12000 words in length, including footnotes and references. Articles exceeding 12000 words will be accepted only occasionally and upon a reasonable justification from their authors. The discussion notes must be no longer than 3000 words and the book reviews must not exceed 4000 words, including footnotes and references. The editors reserve the right to ask the authors to shorten their texts when necessary.

4. Manuscript Format

Manuscripts should be formatted in Rich Text Format file (*.rtf) or Microsoft Word document (*.docx) and must be double-spaced, including quotes and footnotes, in 12 point Times New Roman font. Where manuscripts contain special symbols, characters and diagrams, the authors are advised to also submit their paper in PDF format. Each page must be numbered and footnotes should be numbered consecutively in the main body of the text and appear at footer of page. For all references authors must use the Humanities style, as it is presented in The Chicago Manual of Style, 15th edition. Large quotations should be set off clearly, by indenting the left margin of the manuscript or by using a smaller font size. Double quotation marks should be used for direct quotations and single quotation marks should be used for quotations within quotations and for words or phrases used in a special sense.

5. Official Languages

The official languages of the journal are: English, French and German. Authors who submit papers not written in their native language are advised to have the article checked for style and grammar by a native speaker. Articles which are not linguistically acceptable may be rejected.

6. Abstract

All submitted articles must have a short abstract not exceeding 200 words in English and 3 to 6 keywords. The abstract must not contain any undefined abbreviations or unspecified references. Authors are asked to compile their manuscripts in the following order: title; abstract; keywords; main text; appendices (as appropriate); references.

7. Author's CV

A short CV including the author's affiliation and professional postal and email address must be sent in a separate file. All special acknowledgements on behalf of the authors must not appear in the submitted text and should be sent in the separate file. When the manuscript is accepted for publication in the journal, the special acknowledgement will be included in a footnote on the first page of the paper.

8. Review Process

The reason for these requests is that all articles which pass the editorial review, with the exception of articles from the invited contributors, will be subject to a strict double anonymous-review process. Therefore the authors should avoid in their manuscripts any mention to their previous work or use an impersonal or neutral form when referring to it.

The submissions will be sent to at least two reviewers recognized as specialists in their topics. The editors will take the necessary measures to assure that no conflict of interest is involved in the review process.

The review process is intended to be as quick as possible and to take no more than three months. Authors not receiving any answer during the mentioned period are kindly asked to get in contact with the editors.

The authors will be notified by the editors via e-mail about the acceptance or rejection of their papers.

The editors reserve their right to ask the authors to revise their papers and the right to require reformatting of accepted manuscripts if they do not meet the norms of the journal.

9. Acceptance of the Papers

The editorial committee has the final decision on the acceptance of the papers. Papers accepted will be published, as far as possible, in the order in which they are received and they will appear in the journal in the alphabetical order of their authors.

10. Responsibilities

Authors bear full responsibility for the contents of their own contributions. The opinions expressed in the texts published do not necessarily express the views of the editors. It is the responsibility of the author to obtain written permission for quotations from unpublished material, or for all quotations that exceed the limits provided in the copyright regulations.


11. Checking Proofs

Authors should retain a copy of their paper against which to check proofs. The final proofs will be sent to the corresponding author in PDF format. The author must send an answer within 3 days. Only minor corrections are accepted and should be sent in a separate file as an e-mail attachment.

12. Reviews

Authors who wish to have their books reviewed in the journal should send them at the following address: Institutul de Cercetări Economice și Sociale „Gh. Zane” Academia Română, Filiala Iași, Str. Teodor Codrescu, Nr. 2, 700481, Iași, România. The authors of the books are asked to give a valid e-mail address where they will be notified concerning the publishing of a review of their book in our journal. The editors do not guarantee that all the books sent will be reviewed in the journal. The books sent for reviews will not be returned.

13. Copyright & Publishing Rights

The journal holds copyright and publishing rights under the terms listed by the CC BY-NC License (). Authors have the right to use, reuse and build upon their papers for non-commercial purposes. They do not need to ask permission to re-publish their papers but they are kindly asked to inform the Editorial Board of their intention and to provide acknowledgement of the original publication in *Logos & Episteme*, including the title of the article, the journal name, volume, issue number, page number and year of publication. All articles are free for anybody to read and download. They can also be distributed, copied and transmitted on the web, but only for non-commercial purposes, and provided that the journal copyright is acknowledged.

No manuscripts will be returned to their authors. The journal does not pay royalties.

14. Electronic Archives

The journal is archived on the Romanian Academy, Iasi Branch web page. The electronic archives of *Logos & Episteme* are also freely available on Philosophy Documentation Center web page.

LOGOS & EPISTEME: AIMS & SCOPE

Logos & Episteme is a quarterly open-access international journal of epistemology that appears at the end of March, June, September, and December. Its fundamental mission is to support philosophical research on human knowledge in all its aspects, forms, types, dimensions or practices.

For this purpose, the journal publishes articles, reviews or discussion notes focused as well on problems concerning the general theory of knowledge, as on problems specific to the philosophy, methodology and ethics of science, philosophical logic, metaphilosophy, moral epistemology, epistemology of art, epistemology of religion, social or political epistemology, epistemology of communication. Studies in the history of science and of the philosophy of knowledge, or studies in the sociology of knowledge, cognitive psychology, and cognitive science are also welcome.

The journal promotes all methods, perspectives and traditions in the philosophical analysis of knowledge, from the normative to the naturalistic and experimental, and from the Anglo-American to the Continental or Eastern.

The journal accepts for publication texts in English, French and German, which satisfy the norms of clarity and rigour in exposition and argumentation.

Logos & Episteme is published and financed by the "Gheorghe Zane" Institute for Economic and Social Research of The Romanian Academy, Iasi Branch. The publication is free of any fees or charges.

For further information, please see the Notes to Contributors.

Contact: logosandepisteme@yahoo.com.