

UPJA


The Undergraduate Philosophy
Journal of Australasia

Vol. 1


December, 2019

UPJA is the first undergraduate philosophy journal run by students from the Australasian region. We aim to be an inclusive and diverse journal that will provide a platform for undergraduate philosophy students to showcase their work and to gain their first experience of the publication process. We welcome submissions on any philosophical topic and are interested in publishing work that attempts to make a substantive contribution to contemporary debate.

 Undergraduate Philosophy Journal of Australasia – UPJA

 <https://forms.gle/we6ZeiHAx1hDKf6E6>

 contact@upja.online

 www.upja.online

Editors-in-Chief

Kida Lin (Australian National University)

Matthew Wiseman (Australian National University)

Associate Editor

Bendix Kemmann (Munich Center for Mathematical Philosophy, LMU)

Faculty Advisors

Dr Stephanie Collins (Australian Catholic University)

Dr Sandra Field (Yale-NUS College)

Dr Carolyn Mason (University of Canterbury)

Graduate Student Advisors

Susan Pennings (Australian National University)

Shang Long Yeo (Australian National University)

Referees

Rory Collins: Rory is an undergraduate at the University of Canterbury triple-majoring in philosophy, English, and psychology. His research is diverse and interdisciplinary, having previously published on topics from performance-enhancing drugs to 16th century French literature to logical paradoxes.

James Eder: James is a philosophy Honours student at the University of Canterbury, New Zealand. He has a keen interest in bioethics and trans-human ethics.

Sam Giles: Sam is an Honours student in philosophy at UNSW. His main areas of interest are aesthetics and the history of philosophy, particularly in relation to Deleuze and Kant.

Mariska Harlie: Mariska is a final-year Bachelor of Philosophy student at University of Notre Dame, Fremantle WA. She is interested in practical application of philosophical theories, particularly in relation to history and religion.

Chia Zhi Heng: Zhi Heng is a philosophy student at Yale-NUS College. His main interests in philosophy are philosophy of mind, grounding and meta-ethics.

Kelly Herbison: Kelly is an Honours student at The University of Melbourne. She is interested in social philosophy; in particular, how to understand and challenge hierarchy.

Daniel McKay: Daniel recently completed a PhD at the University of Canterbury. He is particularly interested in consequentialism, freedom, and free will.

Rouein Momen: Rouein is a philosophy Honours student at the University of New South Wales. He is particularly interested in social critical theory and how it could help dissect the pathologies of the modern life-world.

Ella Moon: Ella is a second-year philosophy student in the Bachelor of Liberal Arts and Science at the University of Sydney. Their main interest is in ethical theory, specifically utilitarianism, as well as ethics' intersections with political philosophy and various areas of ancient philosophy.

Tristan Moore: Tristan is a third-year philosophy major at the University of Sydney. He is principally interested in philosophy of language and epistemology.

Campbell Rider: Campbell is completing his Bachelor of Arts at the University of Melbourne. His main interests include philosophical aesthetics and the history of early modern metaphysics.

Thomas Santilli: Thomas is a philosophy Honours graduate from the Australian National University. His area of study is the metaphysics of time and causation.

Jake Stone: Jake graduated from Honours in philosophy at the Australian National University in 2018. His main interest is in the application of formal philosophy to artificial intelligence.

Gerald Teng: Gerald is a philosophy Honours student at the Nanyang Technological University in Singapore. He is particularly interested in the philosophy of science and the intersections of Taoism and Western philosophies.

Joel Towell: Joel graduated from Honours in philosophy at Monash University in late 2018. He is particularly interested in non-classical logics and the normative status of logic.

Danny Wardle: Danny is a Master of Philosophy candidate at the University of Adelaide. He is primarily interested in the metaphysics of persistence and time.

Patrick Williamson: Patrick is a philosophy Honours student at the Australian National University. He is currently working on population axiology and the Repugnant Conclusion.

Editors' Note

We are pleased to present the first volume of The Undergraduate Philosophy Journal of Australasia (UPJA). We started this journal with the hope that it will provide aspiring philosophers in this region the chance to showcase their talent and gain experience of the publication process. And we hope that we've achieved our goal.

The response to our first call for papers was very encouraging. We received over thirty submissions, with authors drawing from sixteen institutions spanning six different countries. We are proud of the fact that roughly 50% of authors who submitted papers were members of an underrepresented group in philosophy. Thanks very much to all who submitted papers and/or applied to referee. In particular, a big thank you to the referees listed on the previous page; many authors commented on the acuity of your feedback.

The four papers in this volume cover metaphysics, social philosophy, epistemology and aesthetics. In *Fragility, Influence, and Intrinsicity*, Patrick Williamson interrogates David Lewis's writings on causation. In *Epistemic Vices*, Kelly Herbison proposes a new way to understand the practice of 'vice-charging'. Jake Stone examines the nature of high-order evidence in *Models of Rational Inference*. And Campbell Rider critiques Kendall Walton's views on the visual arts in *Seeing Double*.

We are happy to announce the winners of our two prizes, generously funded by the Australasian Association of Philosophy. Best Paper goes to Patrick Williamson of the Australian National University, and Best Paper (Member of an Underrepresented Group in Philosophy) goes to Kelly Herbison of the University of Melbourne. Congratulations to you both!

There are a number of organisations and people whose support we would like to acknowledge. First and foremost, we have greatly benefited from the assistance of the Australasian Association of Philosophy. Thank you also to Minorities and Philosophy for promoting the journal. We'd also like to thank our faculty advisors, Dr Stephanie Collins (Australian Catholic University), Dr Sandra Field (Yale-NUS) and Dr Carolyn Mason (University of Canterbury), as well as our graduate student advisors, Susan Pennings and Shang Long Yeo. Finally, special thanks to our Associate Editor Bendix Kemmann.

We hope that this journal will continue to be a platform for undergraduate students from the region going forward. If you are interested in joining us, or hearing about our upcoming 2020 call for papers, please keep an eye on our Facebook page and website, and sign up to our newsletter.

Kida Lin and Matthew Wiseman

December 2019

Table of Contents

Fragility, Influence, and Intrinsicity	[1]
PATRICK WILLIAMSON, AUSTRALIAN NATIONAL UNIVERSITY	
Epistemic Vices: Should Members of Oppressed Groups Vice-Charge?	[13]
KELLY HERBISON, UNIVERSITY OF MELBOURNE	
Models of Rational Inference: Incorporating Higher-Order Evidence	[24]
JAKE STONE, AUSTRALIAN NATIONAL UNIVERSITY	
Seeing Double: Assessing Kendal Walton's Views on Painting and Photography	[37]
CAMPBELL RIDER, UNIVERSITY OF MELBOURNE	

Fragility, Influence, and Intrinsicity

PATRICK WILLIAMSON

AUSTRALIAN NATIONAL UNIVERSITY

Abstract

On David Lewis' original theory of causation, one event c causes another event e iff the right kind of counterfactuals are true. I outline the original counterfactual analysis before considering the challenge from redundant causation. I note that we might respond to redundant causation by adopting extreme standards of fragility, and that Lewis' criticisms of fragility are not persuasive. Lewis himself responds to redundant causation by developing an influence account of causation. I argue that influence is too imprecise as an analysis of causation, and that causation and causal influence in any case seem to be two distinct phenomena. I conclude by emphasising that the correct account of causation must allow for the intrinsicity of causal processes. This gives us good grounds for adopting Lewis' early account of causation as quasi-dependence, regardless of whether or not we view events as fragile. Lewis levelled trenchant criticisms against quasi-dependence, and I conclude by pinpointing the concessions that will, I expect, follow for advocates of quasi-dependence like myself.

1. Possible Worlds and Counterfactuals

A *possible world* is a causally-closed, maximal description of the way things could have been.¹ There is the set of all possible worlds, and this set includes our actual world, since our actual world is also a possible world.

Our world resembles some possible worlds more than others. We represent this fact with an *overall comparative similarity* relation across possible worlds. We measure similarity along two dimensions, both perfect match in particular matters of spatiotemporal fact and commonality in the laws of nature, but the relation will be inevitably vague inasmuch as it relies on our own subjective and context-dependent notions of resemblance.² By definition, every possible world falls somewhere in the similarity order and our own world most resembles actuality.³

This account of possible worlds proves pivotal in our analysis of counterfactual statements. Counterfactuals are subjunctive conditional statements, statements like *if I had been a banker I would have been rich*. The conversational implicature of these

¹ I assume this definition of possible worlds, putting aside complicating factors like modal realism.

² Lewis 1973, p. 560.

³ Ibid.

statements is a false antecedent – that I am not in fact a banker, even if I would like to be rich. We write these counterfactuals $A \Box \rightarrow B$, *if A then would have B*.

Lewis gives the truth conditions for counterfactuals by employing possible worlds and the overall comparative similarity relation.⁴ A counterfactual is true if either the antecedent A fails in every possible world, or if some possible A and B world is closer to actuality than any A and $\sim B$ world.⁵ So the counterfactual *if I had been a banker I would have been rich* is true either if there are no possible worlds in which I am a banker, or if some possible world in which I am a rich banker is closer to actuality than any possible world in which I am a poor banker.

2. Counterfactual Causation

Lewis uses this account of possible worlds and counterfactuals to provide a counterfactual analysis of causation: *e* causally depends on *c* iff *had not c, not e*.⁶ More precisely, *e* causally depends on *c* if the right kind of counterfactual statements are true, both *had O(c), then O(e)* and *had $\sim O(c)$, then $\sim O(e)$* , where *O* indicates an event-tracking proposition.⁷ All this is given that *c* and *e* are distinct, non-overlapping, determinate events.⁸

We can illustrate this with an example. I am a secret agent, going about my business. I follow the target to an abandoned warehouse, pull my gun, light up the empty building with the flash of gunfire. The target sighs before collapsing on the concrete. Now, according to Lewis, the target's death causally depends on my shot if two counterfactuals are true, both *had I shot target would have died* and *had I not shot target would not have died*. Since the first proposition is true in the actual world, we consider the second proposition. If the second proposition is true – namely, if some *not shoot not die* world is closer than any *not shoot still die* world – then target's death causally depends on my shot.

We have in fact described *causal dependence*, but Lewis draws a distinction between causation and mere causal dependence.⁹ Causation supervenes on chains of causal dependence: if *c* causally depends on *b*, and *b* causally depends on *a*, then *a* causes *c*. This allows us to maintain that causation is transitive, even though causal dependence may or may not be transitive.¹⁰

So, to revisit our example: I cause the target's death so long as his death causally depends on the bullet, which causally depends on my pistol barrel, which causally depends on my trigger finger. I cause the death, in other words, when there is a chain of causal dependence connecting my trigger finger to the death. This remains the case

⁴ Ibid.

⁵ Ibid.

⁶ Ibid., p. 562.

⁷ Ibid., p. 563.

⁸ For the purposes of this paper these are only stipulations. For a discussion of indeterminate counterfactual causation, see Lewis 1986.

⁹ Lewis 1973, p. 562.

¹⁰ Ibid., p. 563.

even if there is not *direct* causal dependence from my trigger finger to the target's death – after all, perhaps a second agent waits in the dark, ready to bring about the target's death in every possible world in which I fail to pull the trigger, hence destroying any direct counterfactual dependence from the target's death to my trigger finger.

3. Redundant Causation

I now turn to one of the strongest objections facing the original counterfactual analysis, the issue of redundancy. These are causal scenarios in which either of two events c_1 and c_2 would be sufficient to entail event e , but as a result, e counterfactually depends on neither.¹¹

I restrict myself to the most difficult type of redundancy, so-called late pre-emption cases. Here is an example from.¹² Billy and Suzy are out playing in the sun. They each chuck a rock at an old glass bottle. Suzy's rock soars through the air, crashes into the bottle, and explodes glass fragments across the dirt. Billy's rock passes through empty air where the bottle once stood and sails off into the distance.

Suzy is clearly the cause of the break, since it was her rock that crashed into the bottle. But the counterfactual analysis tells us that Suzy is *not* a cause of the break. After all, if Suzy's rock hadn't broken the bottle, then Billy's rock would have. So the counterfactual *had Suzy not thrown, the bottle would not have broken* is false. So the counterfactual analysis says that Suzy's throw is non-causal.

As a rule, we see that the plurality of potential causes in redundancy cases prevents the right counterfactuals from emerging. It is simply not true that *had not c_1 , not e* , since then c_2 would have entailed e . It is simply not true that *had not c_2 , not e* , since then c_1 would have entailed e . But without these counterfactuals, the counterfactual analysis cannot diagnose the cause of e . So we must either revise or abandon the counterfactual theory of causation.

4. The Fragility Response

We might respond to these late pre-emption cases by adopting extreme standards of fragility. When we say that an event e is *fragile*, we say that e would have been some different event altogether under some miniscule variation in spatiotemporal fact.

By adopting extreme standards of fragility, we rule out troublesome cases of redundant causation *a priori*. Why? It is simply not possible – in our world, at least – for any two events c_1 and c_2 to independently bring about an identical event e . Rather, they bring about two distinct fragile events e_1 and e_2 .

So the analysis goes as follows. Suzy and Billy each chuck a rock at the bottle. Each throw would have led to a different fragile event upon collision, *Suzy's Break* and

¹¹ Lewis 1986, p. 193.

¹² Lewis 2000, p. 82.

Billy's Break. Hence we have a perfectly respectable chain of counterfactual dependence, *Suzy's Throw* $\square \rightarrow$ *Suzy's Break*. Hence the standard counterfactual analysis gives the correct answer, that *Suzy's Break* counterfactually depends on *Suzy's Throw*. Hence *Suzy's Throw* causes *Suzy's Break*.

Lewis offers two arguments against adopting fragility. But in my mind fragility remains a viable response to the problem of pre-emption. In the following section I try to show why Lewis' criticisms are unpersuasive.

First, Lewis does not think extreme fragility fits with our folk intuitions regarding eventhood.¹³ We usually allow that an event might have varied in spatiotemporal minutiae without being instantly destroyed and replaced by another event altogether. For instance, the concert could have been postponed but remained the same concert, or the seminar could have been postponed while remaining the same seminar.¹⁴

This reason strikes me as insufficient. First, we can grant Lewis that our metaphysical analysis should (in broad and general terms) track our folk intuitions. But this fact alone cannot be used as a critique of any single philosophical doctrine without elevating folk intuitions to a status with which many of us will be uncomfortable: Lewis' reply gives the impression that folk intuitions are the primary grounds on which a metaphysical theory is to be accepted or rejected.¹⁵ Perhaps things would have been different if Lewis merely held up folk intuitions as an adjudicatory tool when assessing equally meritorious metaphysical theories.¹⁶ Second, all of this has assumed that there *is* some folk intuition when it comes to the fragility or non-fragility of events. But this point also seems open to dispute. It strikes me that 'the folk' refer to the delayed concert as the 'same' concert as a matter of convenience, without giving much thought to the underlying metaphysical status of events themselves. The conductor is interested in whether the musicians are ready, whether the house is full, whether his suit is straight and tidy – in all of these things – but *not* in whether the concert would have maintained its essential metaphysical identity if any of these individual components were to vary across space and time. After all, how do we know what the conductor says when cornered by a pack of philosophers? Perhaps the conductor gives us a nervous glance, admits that, *technically* speaking, the delayed concert would have been a different event, before adding that it is not very useful to get caught up in such philosophical technicalities. In sum: we should not assume that there is some established folk consensus on the fragility or non-fragility of eventhood.

Second, Lewis argues that extreme fragility allows too many events to qualify as spurious causes.¹⁷ There are hundreds of events that go into making *Suzy's Break* the perfectly fragile event that it is, including air pressure, rock-type, glass-thickness, wind-currents, and so on. If any of these factors had failed to obtain, some other fragile event would have obtained in the place of *Suzy's Break*. Hence, *Suzy's Break*

¹³ Ibid, p. 86.

¹⁴ Ibid.

¹⁵ Ibid.

¹⁶ Thank you to an anonymous referee.

¹⁷ Lewis 1986, p. 198.

counterfactually depends on all of the aforementioned background details. Hence, all these background details are spurious causes for *Suzy's Break*.

Lewis is right to raise the problem of spurious causes. But he is wrong to level it as a probative objection against fragility. Why? The standard counterfactual analysis was already infected by the problem of spurious causation. On the standard counterfactual analysis my smoking causes my lung cancer, since *had I not smoked I would not have had lung cancer*. But on the standard counterfactual analysis my lungs also cause lung cancer, since *had I not had lungs I would not have had lung cancer*.¹⁸ So in buying into the standard counterfactual analysis, we have already allowed for spurious causation.

Now, Lewis might object that the range of spurious causes permitted by the original counterfactual analysis is relatively slim, but the range of spurious causes permitted by the fragility account is virtually infinite. But this response seems incorrect so long as we maintain that causation is transitive. If causation is transitive then my parents cause my birth, and my birth causes my lungs, and my lungs cause my cancer; ergo, my parents cause my lung cancer. And the same might be said for my grandparents, great-grandparents, and so on, back down the causal chains of history. So it is not obvious that the original counterfactual analysis allows only a restricted range of spurious causes. What Lewis would have to argue is that some types of spurious causes are more repulsive than others – that fragility allows a certain *type* of spurious cause not previously allowed by the counterfactual analysis. And I maintain that all spurious causes are equally spurious – it does not matter whether my grandparents are a spurious (counterfactual) cause for my lung cancer, or whether the background breeze is a spurious (fragile) cause for *Suzy's Break*.

5. Influence

Lewis later responded to late pre-emption cases by recasting causation in terms of *influence*. On the influence account, one event *c* causes another event *e* if there is a "chain of stepwise influence from *c* to *e*".¹⁹ There is causal influence from *c* to *e* when a range of close alterations to *c* correspond to sufficient distinct alterations of *e*. So, influence emerges when there is counterfactual dependence of "whether, when and how on whether, when and how".²⁰ Since we are considering a wide range of possible alterations, influence is measured in degrees rather than absolutes.²¹

The influence account makes neat work of late pre-emption cases. Suzy's throw causes the break since it exercises a high degree of influence over the break. A little faster and the bottle falls on its side; a little harder and the hole takes a different shape; and so

¹⁸ Menzies 2004, p. 143.

¹⁹ Lewis 2000, p. 91.

²⁰ Ibid.

²¹ Some examples may prove useful. On the influence account, the following scenarios all qualify as causal: my waking up in the morning *whether* depends on *whether* my heart continues to function, *when* depends on *when* my alarm clock goes off, *how* depends on *how* many hours I slept, *how* depends on *whether* my friend throws a bucket of water on me, and so on. The original counterfactual analysis only adjudicated causation on the basis of *whether* / *whether* dependence, for instance, *whether* I wake up depends on *whether* my heart continues to function.

on.²² By contrast, Billy's throw fails to cause the break since he exercises a low degree of influence over the break. This is not to say Billy exercises no influence, but only that he exercises comparatively less than Suzy.

One other example, Lewis' trumping case, will prove useful in the following section.²³ Sergeant and Major both shout advance and the troops march forward. The troops follow Major's order, since he is the commanding officer. But either shout would have sufficed, giving that Sergeant and Major are both commanding officers. Now, who causes the advance? The influence account tells us, correctly, it seems, that Major is the cause. After all, the right range of alterations on his command c produce significant alterations in the troops' advance e . Since Sergeant is outranked, variations on his shout produce fairly little variation in the soldiers' conduct.

6. Against Influence: Imprecision

It seems to me that there are two good reasons for rejecting the influence account of causation. Here is one reason: influence admits of too much impreciseness and context-dependency to be a viable theory of causation.

First. How large is the set of possible alterations to c ? Unfortunately there is no precise answer. We want 'close' alterations to c , but 'close' is a subjective notion that varies from conversational context to conversational context. There is no clear answer, for instance, as to whether we consider worlds in which (a) the Major shouts the same word slightly differently, (b) the Major shouts different words, (c) does not shout at all, or (d) performs some other action instead of shouting.

Second. Suppose alteration c^* results in alteration e^* . When is e^* sufficiently 'different' from e for this pair $c^* \square \rightarrow e^*$ to contribute to the influence of c on e ? Is it enough if e and e^* are one atom apart? Or must e and e^* be thoroughly and unmistakably different? Again, there is no precise answer. We cannot simply count atoms, because we are also measuring *qualitative* differences, for example, differences in the manner of e . But the problem seems worse this time round. I don't know how to judge whether or not an alteration e^* is 'different' enough without relying on our preconceived notions of causation – in other words, without asking if the change from e to e^* is significant enough to qualify the previous event c as a *cause*. But what qualifies as causation is precisely the question at hand. So I don't know how to determine whether an alteration in e is sufficiently 'different' in any way that is non-circular.

Third. Whether or not c has enough causal influence to qualify as a cause of e depends on the other causal candidates for e under consideration. But the number of other causal candidates under consideration is, again, an imprecise matter determined by conversational context. I take this to be evident in the case of Sergeant versus Major. Sergeant *does* exercise a substantial amount of causal influence over the soldiers (had he shouted a second earlier the troops would have marched, had he shouted a little

²² Ibid., p. 92.

²³ Ibid., p. 81.

louder he would have blotted out Major's voice, and so on). But we deny that Sergeant is the cause because influence claims are comparative, and given that we are comparing him to Major in this particular conversational context, he does not exercise enough causal influence to qualify as a cause.

So the influence account is infected with at least three degrees of imprecision – how large the range of alterations to c , how to know whether a variation from e to e^* is significant enough to contribute to influence, and the range of other causal influencers under consideration. As I mentioned, some of these imprecisions hide a certain circularity. But in general, my concern is that a theory of causation with this degree of imprecision cannot tell us, in any objective sense, what is causation and what is not. As to whether any x causes e , there is no clear answer. We can only give an account of how we might talk about x causing e , conditional on vast quantities of imprecise information being filled in by the background conversational context. The original counterfactual analysis was not free from such imprecision, but limited imprecision to the overall comparative similarity relation between possible worlds.²⁴

7. Against Influence: Coming Apart

Here is another reason for rejecting influence. It seems to me that causation and causal influence are two distinct phenomena that might come apart. Genuinely causal relations might exhibit varying degrees of causal influence – a kind of causal hallmark – but we should not make the mistake of conflating the two.

Here is an initial thought to lead us in this direction. Suppose we reconsider the case of Sergeant and Major, but radically simplify the scenario such that Major is only capable of doing two things – barking ‘advance’ and barking ‘stop’. Perhaps this comes from a lifetime of mind-numbing drill training, and if Major attempts to say anything else he will simply collapse on the asphalt in a fit of cognitive dissonance, the troops letting out a familiar sigh. It seems that we should now say two things. First, that Major's command is still the cause of the advance – after all, the troops are listening to Major and not Sergeant. Second, that Major's command exercises *very little* causal influence over the troops. Alterations in his shout only result in a collapsed officer flat on the parade ground. So, by slightly modifying the scenario we find that causation is present but influence is absent.

This coming apart of causation and causal influence becomes clear in other less-contrived counterexamples. Take the following example adapted from Strevens.²⁵ Billy and Suzy have gone to the pub, and are now throwing darts at a dartboard. Billy has cheated and weighted Suzy's dart so that it flies low. Just as Suzy throws her dart, a wind blows through the window, and causes it to fly just as far right as it flies low:

²⁴ Lewis 1973, p. 560.

²⁵ Strevens 2003, pp. 403–04.

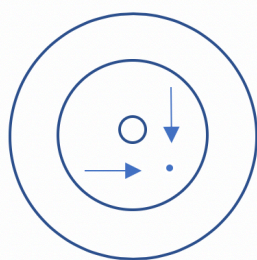


fig 1. weighted dart causes shot to fly low. wind causes shot to fly right.

Now, what causes Suzy's dart to fly low? A range of variations on Billy's weight b produce a range of fine-grained variations on the shot's landing low. But a range of variations on the wind w produce just as many fine-grained variations on the shot landing low. So, by influence, the horizontal wind is just as much a cause for flying low as the weighted dart is a cause for flying low.

Someone might object: *the wind doesn't influence the lowness, since the dart will always fly low no matter how hard the wind blows!* But this is to reject Lewis' conception of influence. On Lewis' view the wind must influence the lowness of the dart exactly as much as Billy's weight does, because the same degree of spatial alterations on 'lowness' can be produced by altering either cause.²⁶

In the dart case we are inclined to say that the wind influences the lowness but is not the cause of the lowness. In the modified Major case we are inclined to say the reverse, that Major is a cause of the advance but not an influencer of the advance. But if causation and influence can come apart in both directions, this is good evidence that causation and influence are two distinct phenomena. So distinct, in fact, that we should not ground our account of one in the other.

8. Intrinsic Processes

I have criticised the influence account of causation. But in light of these counterexamples we can say more about the nature of causation itself, a point that may not have been explicit in the mere fragility account. Suzy's bottle, the shouting Major, and the weighted dart all have something in common: in each case, the 'correct' cause is connected to the effect via some intrinsic series of spatiotemporal events. For instance, we can trace the rock as it leaves Suzy's fingers, slips through the air, and crashes into the glass. We can trace the Major's shout as it echoes across the training field, into the soldiers' eardrums, and causes certain neurons to fire. We can trace the weight as it angles the dart down towards the floor, right until the metal tip sinks into the corkboard.

In light of these examples we can posit the following: the correct analysis of causation will respect the intrinsicity of causation, that is, it will adjudicate causation based on the direct series of spatiotemporal events connecting cause and effect in and of themselves.

²⁶ Also see *ibid.*, p. 404.

To this end we can employ Lewis' early account of causation as *quasi-dependence*.²⁷ On the quasi-dependence account, *c* causes *e* if there is either a chain of causal dependence or quasi-dependence from *c* to *e*. *e* quasi-depends on *c* if *e* counterfactually depends on *c*, not in the actual world, but in some possible world in which both (1) the same intrinsic series of events connects *c* to *e*, but (2) the extrinsic background details of the causal scenario vary, extrinsic background details including pre-emptors.

So we take Suzy's throw; we go to some possible world where the same intrinsic process connects the throw to the break; we remove extraneous background details like Billy's throw; we find that the right type of counterfactual dependence arises; hence the break quasi-depends on her throw, and her throw is the cause of the break.

One key advantage of the quasi-dependence account is that it captures the intrinsicity of causation, while still allowing a certain agnosticism about whether events are fragile or non-fragile. We do not need to know whether the bottle break is fragile in order to determine its cause. We only need to know whether the right counterfactuals arise in intrinsically identical cases. For this reason, I now defend a quasi-dependence analysis of causation, putting aside the issue of fragility as a secondary theoretical consideration.

9. Concessions for Quasi-Dependence

Lewis later offered a range of reasons for rejecting quasi-dependence (QD) and the intrinsicity of causation. These are not trivial criticisms, and deserve an extended treatment of their own. But I want to close by getting clear about what type of concessions I think the defender of QD must be willing to make in light of Lewis' critique. I restrict myself to Lewis' two strongest criticisms.

QD cannot handle trumping cases, so says Lewis.²⁸ In trumping cases we have one cause, Major. And yet we have *two* completed intrinsic event chains leading to the effect, one from Major and one from Sergeant. An implicit consequence: if we attempt to analyse trumping cases via QD, we will be forced to admit that both trumper and trumpee are causes.

I suspect the defender of QD must maintain that all trumping cases are cutting cases.²⁹ That is, the defender of QD must maintain that all trumping cases involve one and only one completed causal chain – a chain which blocks any other potential causal chains from running to completion. The Major's command echoes across the yard and into the soldiers' eardrums, causes certain neural pathways to fire, and at some point these new pathways cut certain Sergeant-obedient neural pathways that would have otherwise run to completion. Hence, Major severs the intrinsic causal chain connecting Sergeant to the advance. So there is only one intrinsic cause-effect chain, and QD can still give a satisfactory analysis of the case.

²⁷ Lewis 1986, p. 206.

²⁸ Lewis 2000, p. 83.

²⁹ Ibid., p. 81.

Now, Lewis imagines alien possible worlds in which there *are* non-cutting trumping cases.³⁰ Perhaps Merlin trumps Morgana when it comes to casting a spell of transmutation. My concern with this answer is methodological: when constructing a theory of causation, we should first see whether our theory maps the actual world, nomologically accessible worlds, and nomologically approximate worlds, since these are the worlds in which our causal intuitions and our causal reasoning remain the sharpest. If our causal theory successfully maps these worlds, then we should see whether our theory maps alien worlds. The fact that a theory does not map alien worlds cannot be used as an objection to a causal theory until its competitors are shown to perform both better in the actual world, nomologically accessible, and nomologically approximate worlds, as well as in alien worlds containing, say, Merlin and Morgana.

QD cannot handle double prevention, so says Lewis.³¹ These are cases in which I cause an outcome by preventing a preventor – by preventing the President from firing nuclear missiles, I prevent Russia from retaliating, and by preventing Russia from retaliating I cause Joe to have his breakfast like normal.³²

Why do double prevention cases pose a challenge to QD? Because these cases are causal, but crucially, because they do not instantiate a chain of spatiotemporal events connecting cause to effect. Such a chain would be required for the intrinsicity of causation. So QD fails.³³

My first inclination was to try and find some intrinsic chain of events at work in double prevention cases. But how could there be, if omissions like *not pressing the button* are not concrete events in spatiotemporal reality? So I think the simplest solution is for defenders of QD to deny that omissions exist, and hence that omissions can be causes.

So Lewis presents a challenge to QD, an omissive causal string like *My Prevention* → *Omitted Button Press* → *Omitted Russian Missiles* → *Joe's Breakfast*. But the defender of QD denies that omissions exist. It follows that the string is some kind of misnomer or misdescription of reality. Hence it does not pose a challenge to the QD account of causation.

It seems to me that the defender of QD must deny the existence of absences altogether. They cannot maintain that absences merely exist in a non-causal way, because if absences exist then the absences will generate the right counterfactuals via their negations and hence qualify as causes on a counterfactual analysis.

³⁰ Ibid.

³¹ Ibid., pp. 83–84.

³² McDermott 1995, p. 529.

³³ Lewis 2000, p. 85. These cases do not instantiate a complete chain of spatiotemporal events connecting cause to effect, since omissions are serving as causes, and since omissions are non-spatiotemporally located.

I am not sure if it is radical to wholeheartedly deny the existence of omissions. The defender of QD will have to find other actual pieces of spatiotemporal reality to cause Joe's breakfast via the right chains of counterfactual dependence.

Instead of admitting absences into our ontology, perhaps we say something like the following: *the President omitted to press the button* is not the claim that a certain omission exists, but rather, a contrastive positive claim about reality, a shorthand for *the President was sitting down and combing their hair in the other room, far away from the button*. In this way, the defender of QD might redescribe omissive causal strings as non-ommissive causal strings, and hence, provide a causal analysis still grounded in unproblematic chains of intrinsic spatiotemporal process.^{34, 35}

So the defender of QD has been forced to make two concessions, that trumping cases are cutting cases and that omissions cannot be causes because they do not exist. I find both concessions plausible, and hence I find quasi-dependence a plausible theory of causation. I prefer it even to the fragility account, inasmuch as quasi-dependence foregrounds the essential intrinsicity of causation.

10. Conclusion

I began by considering Lewis' original counterfactual analysis of causation, arguing that we might reasonably respond to redundancy by adopting extreme standards of fragility. Lewis would rather resolve redundant late pre-emption by recasting his original counterfactual analysis in terms of influence, but I think this is mistaken, since influence is too imprecise as an analysis of causation and since there are cases in which influence and causation come apart. Regardless of whether or not events are fragile, the correct account of causation will respect intrinsicity, the fact that causation is somehow to do with chains of spatiotemporal events connecting cause to effect in and of themselves. To this end, Lewis' early account of causation as quasi-dependence is a profitable analysis of causation, although I suspect it requires certain non-trivial metaphysical concessions.³⁶

³⁴ Hart & Honore 1985, p. 38.

³⁵ If we deny omissions, do we undermine the entire counterfactual analysis so far, which depended on strange things like $\sim c$ and $\sim e$? Only if we think that $\sim c$ and $\sim e$ are claims about the existence of omissions: instead, as I have suggested, we might view $\sim c$ and $\sim e$ as contrastive positive claims about reality. c reads: c exists. $\sim c$ reads: x exists.

³⁶ Thank you to three anonymous referees for their helpful comments and suggestions.

References

- Hart, H. and Honore, A. (1985) *Causation in the law*. Oxford: Clarendon Press.
- Lewis, D. (1973) 'Causation', *The Journal of Philosophy*, 70:556–567.
- Lewis, D. (1986) 'Postscripts to causation', in *Philosophical Papers Vol. 2*, 172–213. Oxford University Press, Oxford.
- Lewis, D. (2000) 'Causation as influence', *The Journal of Philosophy*, 97:182–197.
- McDermott, M. (1995) 'Redundant causation', *British Journal for the Philosophy of Science*, 46:523–544.
- Menzies, P. (2004) 'Difference making in context', in J Collins, N Hall & LA Paul (eds) *Causation and Counterfactuals*, 139–180. MIT Press, Cambridge.
- Strevens, M (2003) 'Against Lewis's new theory of causation: a story with three morals', *Pacific Philosophical Quarterly*, 84:398–412

Epistemic Vices: Should Members of Oppressed Groups Vice-Charge?

KELLY HERBISON

UNIVERSITY OF MELBOURNE

Abstract

Epistemic vices, such as arrogance or closed-mindedness, are vices of an intellectual kind that can affect our belief-forming processes in a self-concealing manner. Such vices are sometimes undetectable to their possessor, meaning that other members of an epistemic community may be better positioned to address them. Ian Kidd outlines vice-charging as a process whereby individuals or groups can collaborate to address epistemically vicious behaviour. This paper discusses whether members of oppressed groups should vice-charge. I propose that (1) members of oppressed groups have good reasons not to engage in what I call perpetrator-focussed vice-charging, but (2) have good reasons to engage in what I call victim-focussed vice-charging. I label Kidd's theory as perpetrator-focussed, because on this approach, a charge is successful insofar as it ameliorates the perpetrator's vice. Members of oppressed groups have good reasons not to conduct perpetrator-focused charges, because by doing so, they risk suffering from what Nora Beranstein calls epistemic exploitation. Such agents do, however, have good reason to make victim-focussed charges. Victim-focussed charges involve a victim of a vice V signalling her condemnation of V to other potential victims of V for the purpose of collectivising. Vice-charging for this function sidesteps the problem that afflicts a perpetrator-focussed practice, while also yielding additional benefits.

1. Vice-charging & its Risks

1.1 Rhetorical and Robust Charges

Ian Kidd outlines two forms that vice-charges might take: *rhetorical* and *robust*. In this section, I define each of these forms of vice-charging and when they might be used, so that we can get clear on how vice-charging works.

Rhetorical charges are defined as an expression of disapproval or negative attitude. A charger who makes a rhetorical charge might do so through an eye-roll, sigh, head shake, terse remark, and so on. These sorts of charges communicate the charger's disapproval in a reactionary way that does not involve communicating the reason for

the charge to the charged.¹ So, if P has vice x, and V decides to charge P with x, the charge would be rhetorical so long as V does not communicate the reason for the charge to P. Because it does not involve a communicated justification, a rhetorical charge simply hints at “certain allegiances, sympathies, values, rather than to try to compel anyone else to adopt them”.² Their lack of communicated justification means that rhetorical charges require a low degree of effort to conduct.

The second kind of charge is a *robust charge*, which are based on robust justifications that are *communicated to the charged* and are therefore different from rhetorical charges.³ The requirement of communicated justification means that robust charges are costlier to carry out for the charger, but more likely to be effective in alleviating the vice. If we can give the vicious person a story about why we are charging them, Kidd thinks they will be more receptive to the charge we are making. As such, robust charges are the most effective kinds of charges, but also the most taxing charge for a charger to make.

1.2 Risks of Vice-Charging

Kidd identifies some risks that can arise in the practices of both rhetorical and robust charging. Recall that Kidd takes the sole purpose of vice-charging to be vice-amelioration. For this reason, the risks he outlines for vice-charging are considered as such because they impede the amelioration process: a charge is risky if it has a low chance of ameliorating the intended vice.⁴

Firstly, rhetorical charges are risky because the charger does not communicate reasons for the charge to the vicious person. Because they are non-justified, they have the potential to make no impact at all, or worse, inflame the vice. For instance, suppose I roll my eyes at someone’s arrogant remark. My eye-roll would probably have no effect if he saw it, given his arrogance. Alternatively, he might ask me to explain why I rolled my eyes. If I decide that it is not worth explaining myself and refuse to do so, he might believe that those who disagree with him are misinformed. If he does not think there was any reason behind the charge, the charge can rebound back onto me and the charged may become even more confident in his beliefs.⁵ While the lack of communicated justification makes rhetorical charges cheap to make, this feature also makes them prone to being ineffective in alleviating the targeted vice.

Recall that robust charges involve a communicated justification. Because of this, they do not run the risk of inflaming vices in the same way that rhetorical charges do. However, there are risks that arise when this justification process fails. Kidd outlines *the problem of consensus* as one way that this might happen. The problem of consensus arises when the vicious agent lacks the requisite shared meanings to make sense of what the charger is attempting to communicate. This can arise when agents have divergent hermeneutical resources. Hermeneutical resources refer to the interpretive

¹ Kidd 2016, p. 183.

² Ibid.

³ Ibid., p. 184.

⁴ Ibid., p. 192.

⁵ Ibid., p. 184.

tools, such as language or concepts, that are available to make sense of one's experiences.⁶ In the same way that hermeneutical resources are needed to make sense of experiences, the experiences available to a knower largely influence what hermeneutical resources they are likely to have.⁷ When a hearer and speaker have different hermeneutical resources, it is largely because they occupy different social positions and have different experiences relative to those positions.⁸ When someone from a marginalised social position tries to communicate something to a dominantly situated person, their efforts might be thwarted because the dominantly situated person lacks the hermeneutical resources to accurately interpret the oppressed person's testimony. To illustrate how different ways of knowing stem from different social groups in ways that impede communication, consider the following testimony

Paranoid

As a perpetual outsider, in virtue of my brown immigrant body, my accent, mannerisms, and the assumptions about my affinities and motivations, I have encountered what are termed as microaggressions both within the classroom and in context of presenting my research. There are countless such incidences, and they still occur every semester without fail. And even within these blatant instances of racism, there have been allies, who not only failed to understand the experience, but charged me with being overly-sensitive (paranoid). Thankfully, today's social media exposes me to the experiences of other women of color and I can receive validation of my reality from them.⁹

In this excerpt, Saba Fatima explains how even those whom she considered allies dismissed her experiences as nothing but a consequence of paranoia. Their failure to understand the experience might arise from their inability to have the experience, or properly listen to those who have the experience, seeing as they do not occupy the same social position that Fatima does as a woman of colour. The difference in opportunities and experiences that are relative to each group manifests in their different conceptual repertoire. Because of this, when people from marginalised groups explain their experiences of oppression to the dominant group, the dominant group often fail to understand. This can lead to injustices of an epistemic kind that occur on the basis of someone having their experiences rendered unintelligible, or having their testimony unfairly dismissed or given a low degree of credibility.¹⁰

In terms of vice-charging, vicious people might not have the hermeneutical resources available to receive the charge that is made against them. If a vicious person does not experience what it is like to be on the receiving end of the vice, they may not have the resources necessary to understand why their behaviour is being charged as vicious. If they do not share an understanding of the vice that the charger is invoking, it is unlikely that they will be receptive to the charge that is made against it.¹¹ The problem

⁶ Fricker 2007, pp. 147-69.

⁷ Ibid., 2007.

⁸ Pohlhaus 2011.

⁹ Fatima 2015.

¹⁰ For a discussion of epistemic injustice, see Fricker 2007.

¹¹ Kidd 2016, p. 192.

of consensus refers to the problems of interpretation that arise because of lacunas in shared interpretive resources. It is a problem for vice-charging because it creates divergent understandings of viciousness and hinders attempts to charge vicious behaviour, seeing as cooperative charging practices are unlikely to occur without some shared conceptions of viciousness to appeal to.

In the next section, I argue that Kidd's conception of risk is determined in terms of efficacy and so fails to capture potential harms that oppressed people face when they vice-charge. I argue for an understanding of the risks of vice-charging in a way that is sensitive to harms that are unique to members of oppressed groups, which persist *even when* the charge successfully ameliorates the vice it is aimed at ameliorating and sidesteps the aforementioned problems.

2. Perpetrator-Focussed Charges & Epistemic Exploitation

Because Kidd's model of vice-charging is committed to alleviating vices, the risks that he outlines for vice-charges are framed in terms of their efficacy: a charge is risky if it has a low chance of ameliorating the vice. As seen above, rhetorical charges might go wrong when the vice becomes inflamed, rather than ameliorated; and robust charges might go wrong when the justification for the charge cannot be communicated, meaning that the vice cannot be ameliorated. Due to amelioration of the perpetrator's vice being the intended goal, I will characterise Kidd's approach as *perpetrator-focussed*:

Perpetrator-focussed charges occur when someone distributes a vice-charge to the perpetrator of a vicious act with the intention of encouraging them to revise their vicious behaviour.

In this section, I argue that perpetrator-focussed charges can be too risky for members of marginalised groups to engage in. I argue that making the perpetrator the intended beneficiary and determining risks in terms of whether the perpetrator changes their vicious behaviour fails to capture risks that oppressed people face when they charge. I illuminate these risks with reference to the theory of *epistemic exploitation*, which occurs when "privileged persons compel marginalised persons to educate them about their oppression".¹² I argue that risks of epistemic exploitation persist in vice-charging practices regardless of whether the charge is ameliorative or not. This means that (a) judging the effectiveness of charges in terms of vice amelioration obscures some of the dangers oppressed people face when they vice-charge and (b) members of oppressed groups have particular reasons, other than those based on potential ineffectiveness, not to engage in perpetrator-focussed charges.

2.1 Epistemic Exploitation

Recall the earlier discussion about how different groups are conversant in different epistemic resources. Because of this, when someone is a member of an oppressed

¹² Berenstein 2016, p. 569.

group, they are uniquely positioned to speak about their oppression. Berenstain's theory of *epistemic exploitation* argues that when oppressed people are compelled to educate their oppressors about these experiences of oppression, this becomes epistemically exploitative.² Educating becomes harmful when the oppressed person educates their oppressor about their experiences of oppression because explaining one's experiences of oppression can be emotionally taxing, sometimes traumatic, and often undermined by the dominantly situated hearer's default scepticism.¹³ The educating process is exploitative because the oppressed person is *compelled* to directly confront and educate their oppressor, and hence run the risk of suffering the aforementioned harms, by virtue of the fact that they are in a double-bind. Namely, they are compelled to engage in risky practices because if they do not, they will continue to suffer the other harms that occur beneath existing social structures. To see how epistemic exploitation occurs, consider the following testimony cited by Berenstain:

Not My Job

Let me tell you what it feels like to stand in front of a white man and explain privilege to him. It hurts. It makes you tired. Sometimes it makes you want to cry. Sometimes it is exhilarating. Every single time it is hard. Every single time I get angry that I have to do this, that this is my job, that this shouldn't be my job. Every single time I am proud of myself that I've been able to say these things because I used to not be able to and because some days I just don't want to.¹⁴

In this excerpt, Manissa McCleave Maharawal, an oppressed person, educates her oppressors; she suffers emotional harms in doing so; and she is compelled to do so, even on days when she does not want to, because if she does not she will continue to suffer beneath existing structures. Maharawal does not comment on whether the white men she is educating are receptive or not to her message, because that is beside the point. The exploitation occurs regardless of the outcome of the educating process because the double-bind compels her to undertake this confrontation. Consequently, when members of oppressed groups are compelled to educate their oppressors, they face harms that persist even in instances where the educating process achieves its intended outcome.

Perpetrator-focussed vice-charging results in epistemic exploitation. By focussing on the outcome of the charge, a perpetrator-focussed theory overlooks the harms that are implicit in the practice. In *Not My Job*, epistemic exploitation occurs because an oppressed person is compelled to engage in emotionally laborious educative processes in order to alleviate other aspects of her oppression. When a victim of a vice is expected to call out the perpetrator of the vice, the same thing is going on. Being a victim of a vice compels the victim to challenge the vice in order to ameliorate the forces that are harming them. This is exploitative because oppressed peoples must either: ignore the compulsion to charge, and hence suffer the oppressive vices; or, act

¹³ Ibid., p. 571.

¹⁴ Maharawal 2011.

on the compulsion to charge, hence undergoing emotional labour and running the risk of their oppressor retaliating. Consequently, members of oppressed groups who are compelled to vice-charge by virtue of being a victim of the vice suffer particular kinds of harms that persist regardless of their charge's efficacy in alleviating the targeted vice. Kidd's perpetrator-focussed project of vice-charging obscures these harms, seeing as they endure in instances when the vice has been ameliorated. As such, members of oppressed groups have good reason not to engage in perpetrator-focussed vice-charges.

3. Redirecting Vice-Charging

So far, we have seen that the emphasis that perpetrator-focussed charges place on vice amelioration obscures the harms of epistemic exploitation that members of oppressed groups suffer. Seeing as these harms occur even in instances when the charge ameliorates the vice it intends to, I concluded that members of oppressed groups have good reason not to conduct perpetrator-focussed charges. Do these risks mean that the oppressed never have good reasons to engage in vice-charging? For the remainder of this paper, I argue that the answer is no. I argue this on the basis that vice-charging could be used as a signalling function for the oppressed to collectivise, which can then allow them to challenge the forces that oppress them in a less risky way. To reap these benefits, the intention of the charge must be *victim-focussed*, rather than perpetrator-focussed:

Victim-focussed charges occur when someone who is a victim of a vice distributes a vice-charge with the intention of communicating the charge to other potential victims of the vice.

Notice that this definition directs charges immediately towards victims of the vice as opposed to possessors of the vice. If P possesses vice *x* of which V is a victim, V should charge P with *x* via signalling to other victims of *x*, rather than for the purpose of communicating it to P. By shifting vice-charges so that they are victim-focussed, the victims become the intended beneficiaries of the initial charge. When the perpetrator is no longer the intended beneficiary, the risks of charging are no longer determined in terms of vice amelioration. Further, this kind of practice does not involve the kind of confrontation of one's oppressor that constitutes the morally objectionable double-bind in epistemically exploitative cases. As such, there are reasons to think that the aforementioned reasons not to vice-charge are absent from a victim-focussed practice.

It is worth distinguishing victim-focussed charges from both kinds of charging that Kidd outlined. Recall that rhetorical charges are not aimed at compelling others to adopt the reasons for the charge. Because victim-focussed charges are directed at common victims of the vice for the function of collectivising, they fail to meet this criterion. Further, victim-focussed charges are not robust, in that they do not aim to provide the justification of the charge to the charged. Accordingly, victim-focussed charging is a novel addition to Kidd's taxonomy.

3.1 Signalling for Collectivisation

Individuals can take steps to challenge cases of injustice by collectivising. Collectivising occurs when previously disorganised individuals come together to form a collective agent. Collectivising provides a way for disparate and disaffected individuals to come together and challenge the reason for their disaffection. The role of collectivising, and the definition I will make use of throughout this paper, is to make a collective agent out of individuals with the view that members of a collective can act to reach intended outcomes in more effective ways. The potential for change to occur is increased through collectivisation, as pressure against the injustice mounts by combining individual goals into a joint effort and distributing roles to members in order to achieve these goals.¹⁵ For example, suppose that there are a number of individuals who think that the large-scale enterprise of ‘fast fashion’ is unjust and should be challenged. If they remain as a disorganised set of individuals, they can make efforts to boycott particular brands and support ethical ones instead. These efforts might make some dint, but their potential is limited. Conversely, when individuals are brought together by an interest in securing common goals, they can form a collective agent by securing common goals and bearing obligations to pursue these goals.¹⁶ When they fulfil these roles, the likelihood of achieving their goals is increased.

Whilst collectivising is an effective way to challenge structural injustices, sometimes the injustices might obscure the tools necessary to collectivise in the first place. For example, if a city is totally inundated with ‘fast fashion’ brands, a consumer who feels uneasy with the enterprise might struggle to come into contact with those who share her reservations. If there are no pre-existing alternatives to locate like-minded individuals and collectivise with them, the agent can take a step towards collectivising through other means. For example, our concerned shopper might make her reservations known to others, with the hope of accruing support from like-minded individuals, through a process of *signalling*. Signalling acts as a way for individuals to communicate to others that they are not unwilling to challenge the injustice.¹⁷ In the face of seemingly insurmountable and pervasive injustices, such as those associated with large-scale enterprises, signalling might take the form of a boycott, consumption of more sustainably sourced products, or even a protest. The concerned shopper might ask a shop attendant at a fast fashion store what the ethical standards of the brand are. By asking this question in front of others, she signals that she is potentially willing to challenge the unethical enterprise of fast fashion. This acts as a step towards collectivisation as it lets others know that challenging the unjust structure is an option. Consequently, signalling is an effective way to challenge injustices because it begins to shift individuals’ ideas about what positions are available to them and which of these positions best aligns with their values.

¹⁵ Lawford-Smith 2015, p. 321.

¹⁶ Ibid.

¹⁷ Ibid., p. 322.

3.2 Charging as Signalling

As we have seen, signalling acts as a way for individuals to take steps towards collectivising against unjust practices or institutions. A vice-charging practice that is victim-focussed might take the form of signalling, whereby to signal means to signal willingness to collectivise. To illustrate, consider:

MAP

Ellie and Gen are the only women in their philosophy class. Ellie has recently come across the organisation Minorities and Philosophy (MAP). She saw that MAP chapters address issues of minority representation in university settings but was not sure if that was something that really affected her. In class, Ellie notices that Gen consistently gets interrupted by one of their male classmates. Ellie also notices that this does not happen when any of the males speak. When Gen is interrupted again, Ellie makes an effort to shake her head to let Gen know that what's going on is wrong. After class, Gen approaches Ellie to ask her why she shook her head. Ellie explains what she read about MAP and gender dynamics in philosophy, and that women's contributions are more likely to be met with closed-mindedness. Together, they collectivise to establish a MAP chapter at their university. Once collectivised, they create things like information sheets to inform others about classroom dynamics to address the combative elements that they have found harmful to them.

In this case, Ellie makes a kind of charge by shaking her head. Whilst her head shake is not likely or intended to influence the males who inspired it at the moment of charging, it signals to Gen that Ellie thinks that the male's sense of arrogance is wrong. Through making the signal, Ellie and Gen took a step towards collectivisation to challenge the structures that caused them harm.

Ellie did not engage in a dialogue with the vicious classmate, but her approach to dealing with it was effective for different reasons. Namely, her charge of the male's arrogance situated Gen as the intended audience and allowed Gen to feel as if the problems that she faced were indeed problems. By signalling that these dynamics were wrong, Gen and Ellie took a step towards collectivising, as they became aware of each other's potential willingness to address the injustice. Victim-focussed charges are more reliable for members of oppressed groups, because if Ellie had made her charge directly to the male, she risked being on the receiving end of retaliation.

Additionally, by focussing her charge to someone from the same social group, the problem of consensus was not a problem for Ellie. Recall that the problem of consensus arises when the vice-charge is being communicated to someone who does not possess the requisite hermeneutical resources to make sense of it. Because members of oppressed groups often share experiences, the likelihood of hermeneutic barriers existing between them is low. This was reflected in the MAP case, as Gen and Ellie's shared experiences meant that they could signal to each other easily, given they could both recognise the source of their harm. This was also reflected in *Paranoid*, as Fatima was able to find validation and understanding of her experiences with other

women of colour online. By finding others who shared her identity, she was better able to communicate her experiences without having them disregarded and misunderstood. If the chance of miscommunication is low, that means the chance of wasting energy on failed attempts at educating is also low. By avoiding situations where the oppressed must directly educate their oppressors, victim-focussed vice-charges avoid the harms associated with the perpetrator-focussed charges. They also have the additional benefits of helping distribute the labour of resistance in a way that is more likely to succeed in challenging the oppressor. As such, the problems associated with perpetrator-focussed charges are absent from victim-focussed charges, and benefits that were not available in perpetrator-focussed charges become available in victim-focussed charges through the prospect of collectivisation.

4. Do Victim-Focussed Charges Still Exploit?

Above, I made the case that epistemic exploitation occurs when oppressed people are compelled to educate their oppressors. I demonstrated this through *Not My Job*, which reflected the asymmetrical and persistent harms that are involved when oppressed people are compelled to educate their oppressors. The alternative vision of vice-charging that I proposed, victim-focussed charging, might be thought to run into the same problems. In taking steps to challenge dominant social narratives, it still seems as if oppressed people are compelled to do the legwork to alleviate their own oppression. An important part of epistemic exploitation is that oppressed people feel compelled to do it on the basis that if they do not, they continue to suffer the oppression. In *MAP*, members of a marginalised group are compelled to mobilise their knowledge of oppression in order to educate their oppressors because they are under pressure to alleviate their own oppression. It seems like vice-charging for signalling will therefore still involve an exploitative element.

I spend the remainder of this section pushing back against the worry that victim-focussed charges still oppress on two fronts: (1) In perpetrator-focussed charges, the agent does not act within the same double-bind that we see in paradigm cases of epistemic exploitation; and a large part of this is because (2) signalling for collectivisation is safer for oppressed people because it increases the chance of them realising their intended outcomes. In paradigm cases, individuals put in a high degree of effort for very little or no pay-off, and they do so because they are compelled by virtue of a double-bind. It is these components of a marginalised person putting herself in a risky position by confronting her oppressor and doing so because she will be continuing to suffer if she does not, that strikes us as so morally objectionable. Both of these components are less severe in a victim-focussed practice.

First, the double-bind that oppressed individuals find themselves in when making perpetrator-focussed charges does not motivate victim-focussed charges. Recall the earlier discussion of how the double-bind compels a marginalised person to either: stay silent and continue suffering harms relating to the vice of which they are a victim; or, charge the vicious person, and hence undergo emotional labour and risk retaliation. For victim-focussed charges, the same kind of emotional labour will not be involved,

in that fellow victims of the vice will be much more likely to share a conceptualisation of the vice on the basis of their shared experience. Further, the risk of retaliation is absent, in that the charge is not directed at the perpetrator, and so there is no confrontation. As such, the double-bind that is characteristic of epistemic exploitation is weakened significantly in a perpetrator-focussed practice, seeing as the element of being forced to explain one's oppression to her oppressor, and run the risk of retaliation for doing so, is absent.

In addition to weakening the double-bind, perpetrator-focussed charges have a better chance of achieving the intended goal of a vice-charge. This is because, as shown in the previous section, collectivisation helps to tackle injustice by distributing roles in order to achieve intended goals with greater probability. Additionally, collectivisation amplifies the voices of marginalised people. This itself constitutes a kind of alleviation of injustice, as it challenges the existing structures that silence marginalised people and inhibit their goals from being realised.¹⁸ This helps to further reduce the double-bind that may compel a marginalised person to enter into an epistemically exploitative exchange, seeing as it provides an alternative way to have her interests met that does not involve confronting her oppressor.

Subsequently, even if it is not guaranteed that victim-focussed charges have no exploitative features, they will be markedly less risky than perpetrator-focussed ones, meaning that they are a better option for members of oppressed groups.

5. Conclusion

I addressed the question of whether members of marginalised groups should vice-charge. The framework offered by Kidd acted as a useful way to consider how vice-charging works, and Berenstain's notion of epistemic exploitation helped to identify why his approach may involve risks involved for marginalised people. I made the case that Kidd's theory is unable to make sense of these risks because it is perpetrator-focussed. From there, I moved on to consider whether there could be a victim-focussed practice of charging that would be beneficial to marginalised people. Signalling for the purpose of collectivisation is one such way that this could occur. Signalling avoids the problems associated with perpetrator-focussed vice-charging, as well as reaping additional benefits that allow marginalised people to challenge the forces that oppress them. Whilst the threat of epistemic exploitation may still arise in victim-focussed charges, the double-bind is not a compelling feature of this practice and the chance of pay-off is much higher, which therefore distinguishes the practice from paradigm cases of morally objectionable exploitation. As such, members of marginalised groups have good reason to conduct victim-focussed vice-charges.

¹⁸ Kolers 2014.

References

- Berenstain, N. (2016). Epistemic Exploitation. *Ergo: An Open Access Journal of Philosophy*, 3(22), pp.569-590.
- Fatima, S. (2015). *Am I Being Paranoid? Being a Woman of Color in Academia*. [online] SIUE Women's Studies Program. Available at: <https://siuewmst.wordpress.com/2015/04/20/am-i-being-paranoid-being-a-woman-of-color-in-academia/> [Accessed 23 Jun. 2019].
- Fricker, M. (2007). *Epistemic injustice*. 1st ed. Oxford: Oxford University Press, pp.147-169.
- Kidd, I. (2016). Charging Others with Epistemic Vice. *The Monist*, 99(2), pp.181-197.
- Kolers, A. (2014). The Priority of Solidarity to Justice. *Journal of Applied Philosophy*, 31(4), pp.420-433.
- Lawford-Smith, H. (2015). Unethical Consumption and Obligations to Signal. *Ethics & International Affairs*, 29(3), pp.315-330.
- Maharawal, M. (2019). *So real it hurts: notes on Occupy Wall Street*. [online] Overland Literary Journal. Available at: <https://overland.org.au/2011/10/so-real-it-hurts-notes-on-occupy-wall-street/> [Accessed 23 Jun. 2019].
- Pohlhaus, G. (2011). Relational Knowing and Epistemic Injustice: Toward a Theory of Willful Hermeneutical Ignorance. *Hypatia*, 27(4), pp.715-735.

Models of Rational Inference: Incorporating Higher-order Evidence

JAKE STONE

AUSTRALIAN NATIONAL UNIVERSITY

Abstract

The aim of this paper is to develop a model of rational inference incorporating higher-order evidence (HOE). I conduct this analysis in three stages. First, I show that HOE is evidence of a distinct type as it is agent-relative and undermines an agent's rationality. Here I largely follow the position adopted by David Christensen, which I take to give strong *prima facie* plausibility to the idea that there is HOE and that it should modify our initially-formed beliefs by forcing us to bracket our first-order evidence (FOE). Secondly, I discuss some possible models of rational inference incorporating HOE. In particular, steadfasting, in which one prefers FOE, and calibrating, which favours HOE. Finally, I mediate between steadfasting and calibration in order to determine whether Christensen's position can be sustained. The key question is: how exactly should first-order and higher-order evidence interact? If we cannot give a good account of the interaction, then Christensen's position looks less tenable. I argue that if we keep in mind the different perspectives from which we can evaluate rationality – internally and externally – a defensible model emerges.

1. Introduction

To understand higher-order evidence (HOE), let's first consider an example of such evidence:

Fatigue

You are a highly trained army commando conducting operations behind enemy lines. After three sleepless days and nights, you have located the enemy's secret headquarters. Unfortunately, it is near a school. Using a map and your geographical knowledge, you determine (correctly) the precise coordinates of the headquarters. Having become highly confident of the location you prepare to radio in an airstrike. Before you do, your second-in-command reminds you that you haven't slept for days and that fatigued officers often make errors of calculation.

In this scenario it seems prudent to re-evaluate the relationship you perceive to hold between your evidence and the hypothesised location of the base. This holds even where you have determined the location correctly. The higher-order evidence about being fatigued appears to undermine your ability to utilise your first-order evidence, E, in support of a hypothesis, H, about the location of the enemy base. To be precise, it isn't evidence about the actual relationship between E and H. Rather, it is evidence that causes you to re-evaluate whether you have perceived the relationship between E and H correctly. The prevailing consensus is that higher-order evidence raises the possibility that you may have made a cognitive error in your calculations utilising E in support of H.¹ As a result, the higher-order evidence of being fatigued seems to be evidence that is relevant to the hypotheses under consideration.

But then again, there is some doubt over whether HOE is even evidence at all. That is, one might doubt that higher-order evidence is the type of thing that has any bearing on your beliefs in the hypotheses under consideration. Why should being fatigued affect the location of the enemy headquarters? If HOE is evidence, it is at the very least evidence of a unique kind. Consequently, the distinction between HOE and FOE gives rise to two interesting questions. First, is HOE really a unique kind of evidence? And secondly, if it is unique, what relationship does HOE bear to FOE? This essay will answer these questions in three sections. First, Section 2 explains why there is a strong *prima facie* case to conclude that HOE really is evidence of a unique kind. Section 3 then considers the different models which have been proposed to deal with HOE. The original contribution of this paper occurs in Section 4 where I mediate between these proposed models. I argue that, if we keep in mind the different perspectives from which we can evaluate rationality, a defensible model emerges. More specifically, I argue for a model I call J*-calibration which allows us to vindicate our desire to judge in accordance with our first-order evidence while taking the presence of higher-order evidence seriously.

2. A *Prima Facie* Case for Higher-order Evidence

I am largely in agreement with Christensen, and other authors, that HOE is a unique kind of evidence which is epistemically significant.² Consequently, the following discussion is expository and relies upon Christensen's arguments as *prima facie* evidence that HOE is unique in that it is (1) agent-relative, and (2) appears to undermine an agent's rationality by requiring us to "put aside or bracket" our FOE.³

2.1 Agent Relativity

¹ Christensen 2010, p. 186; Lasonen-Aarnio 2014, p. 315. It is also worth noting that some HOE might raise the possibility that you are thinking especially accurately. For example, an agent who has taken a neurotropic.

² Christensen 2010. For other authors who also argue that HOE can cause us to bracket our FOE see Feldman 2005, Lasonen-Aarnio 2014, Leydon-Hardy 2016, and Schechter 2013.

³ Christensen 2010, p. 195.

HOE is unique as it is agent-relative. In the ‘Fatigue’ case, the higher-order evidence undermines your belief that H. However, for any other agent (who is well rested) with belief H and evidence E, the higher-order evidence will have no such effect. For why should someone else being fatigued affect their belief in H? This, argues Christensen, is contrary to our normal conception of evidence where the effect does not depend on who the agent is (though it may differ depending on an agent’s prior beliefs or background evidence).

2.2 HOE Appears to Undermine an Agent’s Rationality

2.2.1 Bracketing of FOE

HOE is also unique for its ability to undermine an agent’s rationality by preventing them from giving first-order evidence its full due, by undermining their belief in H based on E. Once aware that you are fatigued, you can no longer update your beliefs on the basis of E even though it seems rational to do so. That is, you can’t give E the full consideration that rationality seems to require. This is so whether the agent has committed an error or not. Recall that in the ‘Fatigue’ case, despite your lack of sleep, you calculated the coordinates of the enemy headquarters correctly. However, as you are unaware of your correctness, it still seems prudent to re-evaluate your credence in the coordinates (especially considering the innocent lives at stake). Nevertheless, this does not necessarily require you to reduce your credence. Reduction is problematic as it requires you to increase your credence in the alternative hypotheses and the HOE is silent on these alternatives.⁴ So, HOE prevents you from giving your FOE its full due, even where that FOE appears to support your original conclusion, and where it may actually do so. In this way, HOE requires an agent “to at least to some extent, *put aside* or *bracket*” their evidence.⁵

2.2.2 Contradicts Conditionalisation

Conditionalisation is the process by which an agent updates their credences in light of new evidence. For example, suppose you have a coin which you assume to be fair. Suppose you then learn that the coin was been flipped 100 times and never landed on heads. In light of this new evidence, E, it is rational to revise your credence in the hypothesis that the coin is fair. In particular, an agent’s new credence should match the credence she would have assigned to her hypothesis had she initially supposed that E was the case. More formally, the principle of conditionalisation requires that: a rational agent possessing an initial credence in some hypothesis $\text{Pr}_{\text{initial}}(\text{H})$ who subsequently gains new evidence E, should update their beliefs so that their new credence $\text{Pr}_{\text{new}} = \text{Pr}_{\text{initial}}(\text{H} \mid \text{E})$.

So how does the principle of conditionalisation relate to HOE? Well, once we come to possess HOE, we must re-evaluate our credence in the coordinates of the enemy

⁴ My thanks to Katie Steele for bringing this point to my attention.

⁵ Christensen 2010, p. 195.

headquarters we determined on the basis of E. However, we also know that prior to possessing the HOE, conditionalisation would recommend something different. Suppose that prior to departing on your mission you learnt the same evidence, E, used in the original 'Fatigue' case. Let us also suppose that E entails the coordinates H of the enemy headquarters. If E entails those coordinates, then conditionalising on E should lead us to increase our credence in H. Similarly, if in three days' time, you come to learn E plus evidence, F, that you are fatigued, conditionalisation recommends you adopt the same credence in H, as F has no impact on the relationship between E and H (E entails H whether you are fatigued or not). But we have already said that in three days' time when you learn E and F, rationality requires you to re-evaluate your confidence in H. Schoenfield and Christensen agree that this result contradicts conditionalisation, which says that the credence we should adopt in H, supposing we learn E & F, should be the same as the credence we adopt when we actually learn E & F.⁶

3. Possible Models for Dealing with HOE

Before considering whether Christensen's characterisation is defensible, I first want to canvas the possible models of rational inference incorporating HOE. When considering how first-order and higher-order evidence interact, we have two models to choose from: *steadfasting* and *calibrationism*. In what follows, I adopt the definitions of steadfasting and calibration utilised by Schoenfield.⁷ Steadfasters maintain that an agent's total evidence (E+HOE) supports the same credence that the agent's first-order evidence (E) alone supports. Conversely, calibrationist theories are those theories which require your credence to match your expected degree of reliability. Specifically, calibrationists recommend that

if, independently of the first order reasoning in question, your expected degree of reliability concerning whether P at time t is r, r is the credence that it is rational for you to adopt at t.⁸

This requires an agent to align their credences with their expected degree of reliability. In much of the literature on higher-order evidence, it isn't entirely clear what one's expected degree of reliability is. However, I think it makes the most sense if we take r to be spelled out in terms of a higher-order probabilistic belief function⁹ which assigns a probability to the hypothesis under consideration supposing that your first-order reasoning supports that hypothesis. This might be spelt out more formally as: $\text{Pr}_{\text{HOE}}(\text{H})$

⁶ Schoenfield 2016; Christensen 2010.

⁷ Schoenfield 2015, 2018.

⁸ Schoenfield 2015, p. 428.

⁹ A function is a mathematical device that takes inputs and assigns to each input a single output. A belief function is a kind of function which takes propositions about the world as inputs and assigns to each of those propositions a probability that it is true (epistemologists call these probabilities 'credences' or 'degrees of belief'). A higher-order belief function is just a function that takes your degrees of belief as inputs and assigns to them a probability of being correct.

$\mid \Pr_{\text{FOE}}(H) > 0.5) = r$.¹⁰ Or more simply, the probability that H really is true given the calculations you conducted using your FOE led you to conclude that H was true. In 'Fatigue', this would be the probability that the target is at x supposing you have calculated that the target is at x: $\Pr_{\text{HOE}}(\text{target at } x \mid \Pr_{\text{FOE}}(x) > 0.5) = r$.

Calibrationist views can be further divided into what Schoenfield calls evidential calibration ('E-calibration') and judgment calibration ('J-calibration'). J-calibration requires an agent in possession of HOE to adopt a credence equal to r (the agent's expected degree of reliability) in the judgment they have made, even if the evidence does not actually support that judgment. E-calibration, however, requires an agent with HOE to take on a credence of r in the hypothesis that is supported by their FOE. And what does it mean to say that an agent makes a judgment which is supported by the evidence? Well, a judgment is supported by the evidence if it is the judgment an agent would make upon reflection when in circumstances conducive to good reasoning.¹¹

These proposals may be elucidated by considering some examples of the beliefs J-calibration and E-calibration recommend. Let's assume that having an interest in knowing such things, the military has conducted extensive research into the influence of fatigue on calculating capacity. Consequently, we know that soldiers in situations like 'Fatigue' tend to reach the correct conclusion 60% of the time. Therefore, you know that your expected reliability, r, in 'Fatigue' is 0.6. That is, $\Pr_{\text{HOE}}(\text{target at } x \mid \Pr_{\text{FOE}}(x) > 0.5) = 0.6$. Assuming that your calculation of the coordinates is correct, J-calibration and E-calibration deliver the same verdict. J-calibration recommends you adopt a credence equal to r in the judgment you made on the basis of your evidence, which would mean adopting a credence of 0.6 in H. E-calibration recommends you adopt a credence equal to r in the proposition which is supported by your FOE. In this case, that would mean adopting a credence of 0.6 in H. Conversely, the two models give different recommendations when the agent has committed a cognitive error concerning the relationship between E and H. If you incorrectly judge $\sim H$ to be supported by E, then J-calibration recommends 0.6 credence in $\sim H$, while E-calibration's recommendation is unchanged. Having canvassed the models of rational inference incorporating HOE, we can now determine whether Christensen's position is sustainable.

4. Models of Rationality

¹⁰ $\Pr_{\text{FOE}}(H)$ is the probability that your first-order belief function assigns to the proposition H. In this case, $\Pr_{\text{FOE}}(H)$ must be greater than 0.5 as any probability lower than 0.5 (50%) indicates that your first-order reasoning does not support the conclusion that H is true.

¹¹ My thanks to an anonymous reviewer for pointing out that a pedant might object that an agent may reason poorly even if in circumstances conducive to good reasoning, in which case the judgment would not necessarily be supported by the evidence. In response to such a pendant I would argue that to be in circumstances conducive to good reasoning an agent must be free of cognitive defects. Therefore, the fact that an agent has made a cognitive error indicates they are not in circumstances conducive to good reasoning.

When determining whether Christensen's position can be sustained, the key question is: how exactly should first-order and higher-order evidence interact? If we cannot give a good account of this interaction, then Christensen's position looks less tenable. When answering this question, we should keep in mind that there are different perspectives from which we can assess rationality. In particular, we can employ internalist or externalist standards. I take an internalist standard of rationality to be one which requires consistency only between an agent's doxastic states and evidence to which an agent has access, where doxastic states and evidence will be sufficiently accessible to count as internal if they can be accessed through reflection alone. Conversely, an externalist standard of rationality requires consistency between an agent's doxastic states and evidence to which that agent does not necessarily have access. As the externalist versus internalist debate in epistemology is a long and storied one I don't intend to mediate between the positions.¹² Instead I utilise this distinction as a means of clarifying the debate in the higher-order evidence literature. Making these perspectives explicit demonstrates why the models of rational inference incorporating HOE differ as well as the benefits and shortfalls of each of the proposed solutions. As a result, a discussion of the internal and external requirements of rationality helps clarify how to incorporate HOE into our model of rational inference.

4.1 *An Internalist Standard*

Christensen's model of HOE gives clear guidance regarding what behaviour is rational from the perspective of an agent possessing HOE. We said that possessing HOE means you can't give your evidence the full consideration that rationality requires. Once you become aware of the fact that you are intoxicated, fatigued, etc., you can no longer update your beliefs on the basis of E even though it seems rational to do so. That is, HOE denies an agent the ability to determine through reflection alone the true relationship between their evidence and the hypotheses under consideration. Consequently, the agent no longer has sufficient access (by internalist standards) to the true relationship between E and H. Conversely, the relationship an agent *judges* to hold between E and H and the fact that the agent possess HOE remains accessible. Consequently, calibrating (which type I'll consider shortly) is the behaviour which is rational from the perspective of an agent with HOE. As put by Christensen:

After all, if I could give all my evidence its due, it would be rational for me to be extremely confident of my answer, even knowing that I'd been drugged...Yet it seems intuitively that it would be highly irrational for me to be confident in this case..., I must (at least to some extent) bracket the reasons this evidence provides, if I am to react reasonably to the evidence that I've been drugged.¹³

So, it is rather puzzling to say that in the face of higher-order evidence one should maintain the same credence in a proposition about which your reasoning is called into

¹² For readers who would like to know more about this debate both Pappas (2017) and Poston (2008) provide accessible introductions.

¹³ Christensen 2010, p. 195.

doubt. Consequently, a denial of steadfasting seems a modest claim. However, steadfasting has some intuitive appeal of its own: why should my being drugged, delusional, fatigued etc., have any bearing on the likelihood that some unrelated proposition is true? And it isn't clear which one of these intuitions should prevail. Consequently, both sides need to justify their models.

So if the internalist wishes to advocate for calibrationism they must provide a plausible account of this model of rational inference. From an agent's perspective, E-calibration appears to be a non-starter for at least two reasons. First, E-calibration commits agents to Moore-paradoxical epistemic states.¹⁴ That is, E-calibration requires agents to contradict their own judgments so that they must think to themselves 'I judged that H but I believe that \sim H.' Many authors have considered it unlikely that such a state could ever be rational.¹⁵ And secondly, from an agent's perspective, E-calibration can't be a guide to their behaviour. The presence of HOE means that an agent can no longer be certain which proposition is supported by the FOE; remember that HOE causes us to question whether we have correctly assessed the relationship between our FOE and hypothesis under consideration. That is, in the presence of HOE you should reconsider whether you have used your FOE appropriately; maybe the hypothesized location of the enemy base is wrong. But then again, it is also possible that you calculated the location correctly. In the presence of HOE you can't be sure what conclusion is actually licensed by your FOE. And since E-calibration requires you to adopt a credence which is supported by your FOE (not merely what appears to be supported by the FOE) agents with HOE have no way of knowing whether they have complied with E-calibration. Therefore, any agent with HOE will be epistemically stranded, with no guidance on to how to proceed. Consequently, E-calibration isn't an acceptable standard of internalist rationality.

On the other hand, J-calibration is a reasonable standard from the perspective of an agent with HOE. J-calibrating requires agents to adopt beliefs which are consistent with their judgments. Judgments which, unlike the relation of evidential support, are accessible. Furthermore, J-calibration also avoids the Moore-paradoxical epistemic states which make E-calibration seem untenable. Finally, by assigning credence r to the proposition an agent takes their FOE to support J-calibration accommodates the requirement that we bracket our first-order reasoning. Thought of in this way, J-calibration is an acceptable model of internalist rationality.

However, as put by Schoenfield J-calibration "makes rationality too cheap."¹⁶ This cheapness arises in cases where an agent's judgment regarding the relationship between E and H is incorrect. In the case that E supports H but an agent incorrectly judges that E supports \sim H, J-calibration recommends one adopt a credence r in \sim H. But this requires approval of credences which our evidence does not support. Consequently, if rationality only requires coherence between internal criteria, it appears to allow agents to adopt absurd yet rational beliefs. Agents will be deemed rational no matter how clearly their beliefs contradict their first-order evidence,

¹⁴ For a discussion of these issues see Sliwa and Horowitz 2015.

¹⁵ See for example: Feldman 2005, Silins 2012, and Titelbaum 2015.

¹⁶ Schoenfield 2015, p. 431.

provided their beliefs are based on genuinely held judgments. Surely this is too permissive: we don't want any old judgment to count as sufficient for rational belief.

One could argue that internal coherence constraints are insufficient for rationality regardless of any appeal to an agent's HOE. I think this is correct, but I also think that J-calibration exacerbates the permissiveness of internalism. J-calibration designates agents as 'rational' simply because they recognise that they are in an impaired state.¹⁷ Merely recognising that one is impaired, however, doesn't negate epistemic errors in judging the relationship between E and H.

So even where an agent is hindered by HOE, we feel that their judgments shouldn't be divorced from the degree of evidential support provided by their FOE. But what the FOE supports isn't accessible to an agent with HOE. So, intuitively, we feel that rationality should have some externalist requirements. Therefore, I'd now like to turn to a consideration of externalist evaluations of rationality.

4.2 An Externalist Standard

Evaluating rationality from a purely external third-person point of view is problematic for its inability to take HOE into account. Presumably, an external third-person perspective is unaffected by HOE given that the effects of such evidence are agent-relative. Consequently, on an externalist view, it is correct to adopt a credence in the hypothesis which is supported by the evidence unadulterated by higher-order considerations. So if an agent is affected by HOE and wrongly concludes that E supports $\sim H$, the agent is irrational and should change their credence to favour H.¹⁸ Where an agent correctly concludes that E supports H, the agent should steadfast. Therefore, all that counts for a determination of rationality is if our agent is correct in their evaluation of E and H. In this way, an external model of rationality fails to deal with HOE at all. I echo the opinion of Feldman that such a model is deficient.¹⁹ The fact that fatigued agents often make mistakes is a genuine source of doubt concerning whether the relationship between E and H had been perceived correctly. That is, an agent who moderated their credence in the hypotheses under consideration as the result of possessing HOE would be regarded as more rational than one who did not. Consequently, a theory of rational inference should be developed in a way that acknowledges the existence of HOE. Therefore, a model that can't incorporate these doubts is lacking.

Furthermore, as with E-calibration, such a model leaves agents epistemically stranded. When an agent has higher-order evidence, they can't be certain that they have perceived the relationship between E and H correctly. We said that if an agent wrongly concludes that E supports $\sim H$, the agent should change their credence to favour H. Or, in the event that an agent is correct about the relationship of evidential support, they should steadfast. However, In the presence of HOE an agent won't know which

¹⁷ Again, my thanks to Katie Steele for raising this point.

¹⁸ Note that this is neither steadfasting nor calibrating.

¹⁹ Feldman 2005, p. 107.

situation they find themselves in. Consequently, they won't be able to determine which course of action to take. Even worse, agents will often find themselves epistemically stranded. Christensen begins his paper by acknowledging that: "spending a few hours reading a book like Kahneman reveals a depressing plethora of unexpected and unobvious factors that subtly but significantly disrupt reliable cognition."²⁰ So, if we adopt a purely externalist evaluation of rationality, we should expect to find agents upstream without the proverbial epistemic paddle most of the time.

4.3 *A Hybrid Standard*

It is important to note that internalist and externalist models are simply different viewpoints from which we can evaluate rationality. However, this does not necessitate that these viewpoints are in competition. Consequently, it is possible to propose a standard which incorporates both internal and external constraints on rationality. This is particularly important in the case of HOE as a hybrid model can overcome the difficulties we have identified so far. We said that internalism rendered rationality too cheap and that externalism's failure to deal with HOE was a deficiency. Interestingly, the advantages of each model appear to complement the disadvantages of the other—internalism gives weight to one's HOE and externalism isn't cheap as it requires agents to accurately assess the relationship between E and H.

As discussed, our intuitions point in favour of taking the presence of HOE into account. Consequently, we should calibrate in one form or another. We also noted that E-calibration is a non-starter given it requires Moore-paradoxical epistemic states and leaves agents epistemically stranded. J-calibration then seems to be our only option. But then again, we said that J-calibration cheapened rationality. However, this objection isn't fatal and Schoenfield recognises that it is possible to respond by arguing that this only shows that J-calibration is necessary but not sufficient for rationality.²¹ For why, says the J-calibrationist, should we expect J-calibration to remedy an agent's initial irrational judgment? J-calibration only requires coherency between our judgments and our belief states. Consequently, it only speaks to that coherence and is silent on whether one's initial judgment involved rational errors.

While acknowledging that this reply is open to the J-calibrationist, Schoenfield (2015: 438) argues that it also fails to preserve a robust version of rationality. Even if J-calibration is only a necessary requirement, it still obliges agents to believe hypotheses not supported by their FOE. If J-calibration is necessary for rationality, adopting a credence which doesn't conform to your expected degree of reliability is irrational. So, to be rational an agent must adopt a credence equal to r in H. Consequently, if a rational credence does exist it must assign r to H even if H isn't supported by one's FOE. So, argues Schoenfield, the retreat to necessity rather than sufficiency doesn't avoid our original concerns with J-calibration.

²⁰ Christensen 2016, p. 397; Kahneman 2011.

²¹ Schoenfield 2015, p. 437.

Nevertheless, all hope is not lost. If the J-calibrationist wishes to accommodate FOE they can argue that where an agent has incorrectly judged the relationship between E and H there is no rational belief state that can be adopted. J-calibration is only concerned with coherence between our judgments and our beliefs, it does not require that the initial judgments are rational. If, then, the judgments made by an agent are irrational, it shouldn't come as a surprise that complying with J-calibration fails to make the agent rational. Rationality on this model would require coherence between the agent's judgments and their beliefs (J-calibration) plus rational judgments to begin with!²²

For these reasons, J-calibration won't secure rationality by itself. However, it is still a necessary principle; surely an agent isn't rational if they don't abide by their own judgments. We concluded that in order to be rational we require compliance with J-calibration plus judgments which are initially rational. Consequently, our next question arises: what would make our initial judgments rational? The answer is indicated by our preceding analysis. We said that where an agent has misjudged the relationship between E and H there is no rational belief state they can adopt. Therefore, in order to be rational, an agent must judge the relationship correctly. More precisely, an agent must make judgments which are supported by their first-order evidence. So the two conditions required for rationality are:

J-FOE – One should make judgments that are supported by one's first order evidence.

J-calibration – One should assign a credence r (equal to one's expected degree of reliability) to the hypothesis H they judge to be correct.

J-FOE then serves as our external requirement – an agent must judge in accordance with the evidence, even where HOE denies the access needed to determine what the evidence supports. While J-calibration requires coherence between internally-accessible doxastic states – the agent's judgments and beliefs. These conditions taken together I call J*-calibration. If J*-calibration is correct it places tight constraints on what will count as a rational belief. In order to be rational, an agent is required to both judge correctly and form beliefs which are consistent with those judgments. Therefore, rather than cheapening rationality, J*-calibration makes rationality more difficult to satisfy.

One could argue that J-FOE is an odd requirement given that HOE prevents an agent from assessing whether they have complied with J-FOE. However, this requirement isn't odd if we remember that we are allowing a partially external evaluation of rationality. Even though it may not be possible to know whether one has complied with J-FOE, all things considered, an agent who judges in accordance with their FOE is more rational than one who does not. So, from an external perspective, compliance with FOE is a requirement of rationality. As a result, an agent's inability to determine whether they have satisfied J-FOE is consistent with J-FOE being a necessary requirement of rationality.

²² Christensen 2016, p. 409.

One might also worry that J^* -calibration, in particular J -FOE, leaves agents epistemically stranded in the same way that the purely external evaluation of rationality and E -calibration left agents stranded. I disagree. Unlike E -calibration or purely external rationality, J^* -calibration still provides some guidance to agents with HOE. Specifically, an agent with HOE should assign r to the hypothesis they judged to be best supported by their evidence, whereas an agent in the other scenarios is unable to determine what to do.²³

Another worry regarding the ability to follow this model concerns cases in which an agent receives HOE before forming a credence in H . In such a case the required inputs for J^* -calibration do not appear to be available. According to Christensen such cases make it impossible to form a credence in H independently of your HOE.²⁴ However, I argue that an independent assessment is still possible. To see why recall that HOE isn't evidence about the actual relationship between E and H . Rather, it is evidence that causes you to re-evaluate whether you have perceived the relationship between E and H correctly. Consequently, the presence of HOE does not make it seem as though the FOE supports some other hypothesis. Instead your HOE motivates you to re-examine the hypothesis which seems to be supported by your FOE. That is, the way the FOE appears to you is unchanged by the presence of HOE. For example, if you looked at the sky and concluded it was orange and I then told you that I had secretly drugged you with a chemical that makes blue things look orange, the sky would still appear orange to you. Accordingly, the fact that your FOE appears to support a particular hypothesis is not affected by the presence of HOE. And this is true whether you receive the HOE before or after making an appraisal of your FOE. Therefore, it is still the case that you can make an appraisal of the FOE independently of your HOE. Consequently, in cases where you receive HOE before appraising your FOE an application of J^* -calibration is straightforward – determine which hypothesis your FOE appears to support, then bracket that hypothesis as required by your HOE.

So J^* -calibration doesn't leave agents epistemically stranded, but it is still impossible to consciously comply with when one has HOE. An agent might J -calibrate, but still fail to satisfy J -FOE and therefore be irrational. Furthermore, because of the presence of HOE, the agent would be unaware that they had behaved irrationally. So, in this sense, J^* -calibration is not a useful guide to rectifying rational errors in our judgement of the relationship between our FOE and H . In response to this objection I am in agreement with Christensen who, when defending a similar objection against his own model of rational inference incorporating HOE, had the following to say:

...acknowledging that we can't always get evidential relations right or be certain that we have gotten them right is compatible with holding that failure to get evidential relations right is a rational failure. In epistemology, as in ethics,

²³ Of course, from an agent's point of view J^* -calibration prescribes the same course of action as J -calibration; that is, an agent in possession of HOE should adopt a credence in the hypothesis they judged to be best supported by their evidence which is equal to their expected degree of reliability. Consequently, from an agent's perspective J^* -calibration doesn't provide different guidance to J -calibration. Nevertheless, J^* -calibration is an improvement on J -calibration as it acknowledges that we should adopt a credence equal to our expected degree of reliability *and* assess our FOE correctly.

²⁴ Christensen 2016, p. 408.

the correct norms may put perfection beyond what ordinary humans can generally achieve.²⁵

What I have provided is an ideal theory of rational inference incorporating HOE. It is unreasonable to expect that such a theory also ensures we no longer make rational errors when assessing our FOE. By way of analogy, consider a diet that contains all necessary nutrients and has been shown to substantially improve one's health. It is lamentable that some people may find it difficult to stick to this diet. However, it is no objection to the desirability of such a diet that some agents prefer junk food to healthy meals. Rather than showing the diet is theoretically deficient this highlights a separate problem concerning an agent's willpower or preferences. Similarly, ensuring that agents correctly utilise their FOE is a separate problem to developing a theoretically sound model incorporating HOE. For this reason, it is no objection to J*-calibration that agents cannot always comply with this standard. Of course, it is also desirable to have some means of ensuring we evaluate our FOE correctly, however, this is a task for another theory.

5. Conclusion

The aim of this paper was to defend a model of rational inference which could accommodate HOE. The preceding analysis has done just this. This analysis helps clarify why HOE presents a problem for our models of rational inference and why agents often fall short of ideal rationality. Once these issues are made clearer a defensible model that explains how HOE and FOE interreact becomes apparent. In particular, J*-calibration allows us to vindicate both internal and external constraints on rationality – we can accommodate our desire to judge in accordance with our first-order evidence while taking the presence of higher-order evidence seriously. Of course, the theory is not a panacea for human irrationality as it cannot rectify human fallibility in the evaluation of first-order evidence, but this is not something such a theory is required to do.

²⁵ Christensen 2016, pp. 411-2. Christensen defends what he calls the Idealised Thermometer Model which is related to, but different from, my own model of rational inference incorporating HOE.

References

- Christensen, David. (2010) 'Higher-Order Evidence.' *Philosophy and Phenomenological Research* 81: 185-215.
- Christensen, David. (2016) 'Disagreement, Drugs, etc.: from Accuracy to Akrasia.' *Episteme* 13 (4): 397-422.
- Feldman, Richard. (2005) 'Respecting the Evidence.' *Philosophical Perspectives* 19 (1): 95-119.
- Kahneman, Daniel. 2011. *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.
- Lasonen-Aarnio, Maria. (2014) 'Higher-Order Evidence and the Limits of Defeat.' *Philosophy and Phenomenological Research* 88, no. 2: 314-345.
- Leydon-Hardy, Lauren. (2016) 'Some Resistance to the Idealized Thermometer Model.' *Episteme* 13 (4): 423-426.
- Pappas, George. (2017) 'Internalist vs. Externalist Conceptions of Epistemic Justification.' <<https://plato.stanford.edu/archives/fall2017/entries/justep-intext/>> accessed 29 August 2019.
- Poston, Ted. (2008) 'Internalism and Externalism in Epistemology.' Internet Encyclopedia of Philosophy. <<https://www.iep.utm.edu/int-ext/>> accessed 29 August 2019.
- Schechter, Joshua. (2013) 'Rational Self-Doubt and the Failure of Closure' *Philosophical Studies* 163 (2): 428-452.
- Schoenfield, Miriam. (2015) 'A Dilemma for Calibrationism.' *Philosophy and Phenomenological Research* 91 (2): 425-455.
- Schoenfield, Miriam. (2018) 'An Accuracy Based Approach to Higher Order Evidence.' *Philosophy and Phenomenological Research* 96 (3): 690-715.
- Silins, Nicholas. (2012) 'Judgment as a Guide to Belief.' In *Introspection and Consciousness*, edited by Declan Smithies and Daniel Stoljar. Oxford University Press.
- Sliwa, Paulina, and Sophie Horowitz. (2015) 'Respecting All the Evidence.' *Philosophical Studies* 172 (11): 2835-2858.
- Titelbaum, Michael. (2015) 'Rationality's Fixed Point (or: In Defense of Right Reason).' *Oxford Studies in Epistemology*, edited by Tamar Szabó Gendler and John Hawthorne Oxford University Press.

Seeing Double: Assessing Kendall Walton's Views on Painting and Photography

CAMPBELL RIDER

UNIVERSITY OF MELBOURNE

Abstract

In this paper I consider Kendall Walton's provocative views on the visual arts, including his approaches to understanding both figurative and non-figurative painting. I introduce his central notion of fictionality, illustrating its advantages in explaining the phenomenon of 'perceptual twofoldness'. I argue that Walton's position treats abstract artwork reductively, and I outline two essential components of our aesthetic encounters with the non-figurative that Walton excludes. I then offer some criticisms of his commitment to photographic realism, emphasising its theoretical inconsistencies with his account of representation. My own proposal is that in our apprehension of non-figurative artworks, our attention is drawn to the underlying structures of both emotive and perceptual experience. In this way, paintings, particularly abstract ones, disclose human cognition in a manner that makes fictionality an inappropriate tool for their analysis.

1. Fictionality

Kendall Walton's position on the representational arts—a field encompassing painting, drawing sculpture and collage—is grounded in his theory of fictionality. He proposes that we should understand the experience of such artworks as analogous to engagement in a make-believe game. Just as we assign the status 'knight' and 'queen' to pieces on a chessboard, converting the game of chess into an imagined world of medieval combat, the materials of representational works operate as props in the production of fictional worlds. But these props create a fictional world that is distinct from mere imagining—they *prescribe* certain imaginings and consequently '*generate fictional truths*' that are normative in the sorts of imaginings they prescribe.¹ Just as chess pieces *authorise* a certain kind of imagined battle scenario, props in make-believe games can authorise certain kinds of games. If a painting depicts a horse, then to perceive it as though it were some other animal would be to incorrectly engage in the fictional world the arrangement of paint on the canvas invites us to imagine. Walton says that "what makes it fictional in *La Grande Jatte* that a couple is strolling in a park

¹ Walton 1990, p. 21.

is the painting itself, the pattern of paint splotches on the surface of the canvas.”² The paint splotches, brush strokes, textures and colours behave as props, which prescriptively authorise us to imagine a world where the described scene is fictionally true. When I focus on one feature of a painting, such as a horse standing in a field, I may be using this section as a specific prop in a particular make-believe imagining. I may also step back, and view the painting as a whole, in which case the painting in its entirety constitutes a single prop in the fictional imagining of whatever it represents. For the sake of simplicity, I tend to use ‘prop’ in the latter sense.

One advantage of Walton’s view is that ‘fictionality’ offers a solution to the problem, described by Richard Wollheim, of a depiction’s perceptual ‘twofoldness’—the fact that when viewing an artwork, we have “two simultaneous perceptions: one of the pictorial surface, the other of what it represents.”³ Looking at a canvas, I am aware of the fact that it is a physical object, a painting, that belongs to the world in front of me. But I am at the same moment having a perceptual experience of the things the painting represents. In Wollheim’s language, these are the ‘configurational’ and ‘recognitional’ aspects of a given pictorial representation. By ‘configurational’, we mean the elements of a painting’s construction—brush strokes on the canvas, the different pigments used, the volume and thickness of paint. The ‘recognitional’ aspects are what a viewer sees depicted—a landscape, a tragic scene, or a person strolling in a park. The fact that we are able to simultaneously cognise both aspects of a painting is a perceptual feature central to the success of the visual arts, but explaining how we are able to do this requires an act of cognition that unifies two apparently contradictory attitudes:

1. When viewing a painting, I have a conscious awareness that what I see is not real. Indeed, I may even have an acute awareness of the painting’s physicality: its gilded frame or firm brushstrokes, for example.
2. I am aware of what the painting actually represents, so that even when I closely inspect the thickness or texture of the paint I vividly perceive it as a mountainous landscape or person strolling in the park.

Using the concept of fictionality we can describe this duality in terms of the painting’s role ‘configurationally’ as a prescription to imagine what is seen ‘recognitionally.’ A painting’s physical elements, the strokes of paint, pigments and so on, act as props in imaginative games of make-believe, as prescriptions to imagine something that is not actually there. This notion explains how we can have both a recognitional imagining and a configurational awareness of a work in one perceptual act. To imagine is to experience *fictionally*— we are prompted to form a propositional attitude that prescribes an imagining without committing the subject to a belief that what is imagined is *actual*. Through a single cognitive step, such as ‘it is fictionally true that this woman, Liberty, is leading the people’, the subject acknowledges that Delacroix’s painting is a prop, and that what it depicts is not actual, while still recognising the make-believe imagining it prescribes. ‘Fictionality’ as a concept describes an

² Ibid, p. 38.

³ Wollheim 1998, p. 221.

imaginative encounter that unifies recognitional experience of both what a painting depicts, as well as awareness of its physical makeup, without contradiction.

2. Non-Figurative Art

In this section I wish to describe Walton's novel approach to understanding abstract, or non-figurative, artworks. I then elaborate on the limiting nature of this approach, in particular focussing on how it conceives of abstract art in terms that fail to acknowledge the role played by pre-conscious structures of human receptivity to colour, form and space.

'Figurative' artworks, which depict familiar objects such as people, places and things, are ordinarily distinguished from 'non-figurative' artworks, which depict shapes, colours and forms in an abstract, conceptual fashion. The salient difference might appear to be that only figurative artworks are representational—they *represent* existing people, places or things. It is an interesting feature of Walton's theory, then, that he considers both kinds to be representational. Non-figurative artworks 'represent' in the specific sense that although the figures they depict are abstract (they are not worldly objects) they are experienced by the viewer in a manner that is not true of their actual configuration. Walton offers the example of an abstract Malevich work, featuring coloured shapes overlapping on the canvas. Although these elements actually lie on the same plane, we nonetheless perceive a yellow rectangle sitting *in front of* a green one because of its illusory (or fictional) appearance of depth. According to Walton, "this is how we see the painting, not how it actually is. Actually the yellow, green, and black are all on (virtually) the same plane... To see the painting this way is, in part, to imagine [depth]".⁴ In short, even a highly abstract piece can produce an imagined appearance of three-dimensionality through careful arrangement of paint on a flat, two-dimensional canvas, making the Malevich painting a 'fiction' since it prompts make-believe imaginings that are not reflective of the work's physical configuration. Fictionality, according to Walton, is therefore central to understanding all kinds of representational artwork, including both figurative and non-figurative kinds.

My principle objection to this account is that it demands interpretation of non-figurative artworks in terms that diminish their artistic value by ignoring their potential to disclose critical elements of human experience. Given Walton's contention that make-believe games "have a profound role in our efforts to cope with our environment" by allowing us to practice social behaviours and improve our self-understanding and capacity for empathy, we can see how Walton might regard figurative artworks as the superior form of creative expression, since only they can allow us to enact games that are grounded in the real world.⁵ A figurative neoclassical painting such as David's *The Oath of the Horatii* might enable the viewer to engage in an imagining that provokes a better understanding of honour, duty and sacrifice. On the other hand, a Pollock artwork, which features abstract paint drippings splattered

⁴ Walton 1990, pp. 54-5.

⁵ Ibid, p. 12.

across the canvas, can never produce an imagining rich in the same kind of moral content. If fictionality is constitutive of our encounters with artworks, then in Walton's view, abstract, non-figurative splotches of paint on a canvas must fail to achieve a level of artistic value comparable to figurative works.

In my view, this definition of the non-figurative ignores abstract artworks' capacity to elicit profound emotional responses and thus prompt self-reflection on our innate receptivity to colour and form. Mark Rothko's abstract expressionist paintings, for example, provoke emotional responses through the use, arrangement and texture of colour, but they do not achieve this because the colours, as props, induce an imagining or create a scene which in turn causes us to feel an emotion. This is the mechanism at work in a painting such as the *Death of Socrates*, where the viewer imagines the depicted tragic scene, and then responds emotionally. While the use of colour will contribute to the tone of the scene, the viewer's response is in large part mediated by empathetic acknowledgement of the portrayal's tragic nature. In Rothko's work, by contrast, the emotional response is *immediate*. A viewer witnesses the painting and is provoked by its sensitive configuration of texture, tone and colour, without any intervening step involving an imagined scenario.

This activity is suggestive of what Heidegger calls the 'primordial' fore-structures of understanding: the implicit understanding of the world, others, and art that precedes all conception in conscious, reflective or logical cognition.⁶ The fore-structure most relevant to understanding Rothko is the phenomenon of 'mood'. Heidegger argues that moods, emotional states, emerge from our primordial orientation towards the world as an environment charged by temperament and affect.⁷ On this reading, the



Figure 1. Rothko, Mark. 'No. 14, 1960.' San Francisco: Collection SFMOMA, 1960.

⁶ Gorner 2000, pp. 77-8.

⁷ Heidegger 2006, p. 179.

viewer of a Rothko painting *directly* submits to whatever affective inflection it ‘assails’ them with—whether it be grief, melancholy or an uplifting sense of elation.⁸

Walton might respond that this is still explained by the representational nature of the artwork, in its capacity to make it fictionally true that these emotions are being ‘played out’ on the canvas. But by his own admission, whereas a figurative painting “induces and prescribes imaginings about things external to the canvas”, a non-figurative one “calls merely for imaginative rearrangement of the marks on its surface.”⁹ The fictionality of an abstract artwork extends only to the recognition of the forms and their orientation suggested by the paint—the appearance of depth or contrasting size, for example. In this respect, “figurative paintings ‘point beyond’ themselves in a way that [non-figurative ones do] not.”¹⁰ Only the former type can represent something external, or beyond, what is contained in the painting such as mountains or people strolling in a park. Non-figurative paintings are necessarily confined to the fictional representation of their own formal elements. Rothko’s own views on the emotive nature of his abstraction articulate what this reductive account fails to capture:

I’m not interested in relationships of color or form or anything else... I’m interested only in expressing basic human emotions—tragedy, ecstasy, doom and so on—and the fact that lots of people break down and cry when confronted with my pictures shows that I communicate those basic human emotions... The people who weep before my pictures are having the same religious experience I had when I painted them. And if you... are moved only by their color relationships, then you miss the point!¹¹

The way Rothko’s artworks elicit emotional responses that are ‘beyond’ the canvas suggests there is something more fundamental to the experience of non-figurative artworks than the representational fictions that relate the formal elements of the work. As a mode of aesthetic criticism and interpretation, Walton’s ‘fictionality’ offers no grounds for the evaluation of whether an abstract artwork succeeds or fails aesthetically in its attempt to capture a mood or feeling. For Walton, the role of art criticism is instead functional—it assesses the relationships of the ‘props’ on a canvas and their function in generating worlds: “We are interested, sometimes, in seeing what contributions it is [props’] function to make to games of make-believe, what fictional truths it is their functions to generate, and what sorts of games would accord with their function.”¹² Since there is no intermediate, make-believe game in the emotional experience of a Rothko work, Walton has nothing to offer by way of interpretive explanation. Art criticism, we might object, ought to capture just what elements of a painting like Rothko’s are able to produce such vivid emotional responses, and what this reveals about the temperamental orientations Heidegger claims enable them.

It may be replied that Walton simply offers an explanation for how artworks are perceived and recognised as such, in order to tackle problems like Wollheim’s.

⁸ Ibid, p. 176.

⁹ Walton 1990, p. 55.

¹⁰ Ibid, p. 57.

¹¹ Doss 2011, p. 1.

¹² Walton 1990, p. 53.

Aesthetic judgement is a *hermeneutical* question, a matter of literary and critical interpretation; he is concerned merely with perceptual ones. However, if he thinks fictionality is constitutive of our imaginative encounters with artworks and the primary feature operative in our cognition of them, then the emotional responses prompted by abstract artworks must be in some way determined by their fictional characteristics. Walton's theory fails to account for this, and this shortcoming exposes the extent to which fictionality as a theoretical proposal is unable to grasp those responses to artworks that are most revealing of our emotive comportment to the world.

A further shortcoming of Walton's view is that by so rigidly distinguishing figurative and non-figurative art, he is forced to categorise a given work as either one kind or the other. For this reason, he is unable to make sense of artworks that blur this distinction. Consider Mondrian's *Pier and Ocean*, which is considered one of his intermediate works: while anticipating his later phase of total abstraction, it is still, in an important way, figurative. Although the work is dominated by abstract, non-figurative features, the concentration of longer lines at the bottom of the canvas, which gradually shorten towards the middle, create an impression of depth. Using Walton's theory, we might say these lines 'point beyond themselves' to prompt an imagining of a pier extending into the ocean—it is *fictionally* true that there is a pier, so, given the rigid distinction, the painting must be categorised as figurative.

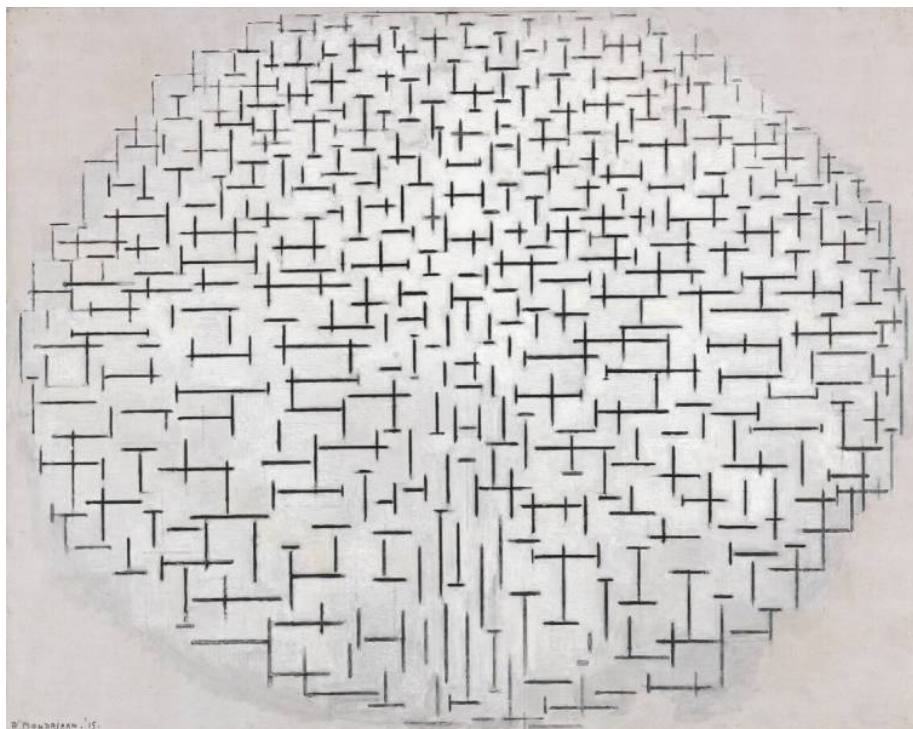


Figure 2. Piet Mondrian, *Composition No. 10: Pier and Ocean*, 1915. Kröller-Müller Museum.

However, conceiving of the painting as fictional in this way fails to recognise the effect produced by its gestures towards abstraction. By portraying a seascape using a network of lines arranged horizontally and vertically, Mondrian prompts the viewer

to reflect on the spatial geometry that underpins our perceptions. The painting does not prompt us to simply imagine what it represents—a pier extending into the ocean—but to realise through its unusual manner of representation a new way of seeing and understanding the subject. The spatial determinations that orient and enable perception are made explicit by this artwork in a manner that would be obscured by filling in its geometrical structure with the textures of an ordinary seascape.

Walton's sharp differentiation between figurative and non-figurative works simply reduces artworks like this to an imagining of what they represent figuratively. And since the use of abstraction impairs the viewer's engagement in a make-believe game, Walton would consider the work less artistically valuable than a more conventional, realist representation. When viewing the Mondrian, we are prompted to imagine the ordinary blue sky, horizon, and waves that a figurative painting of the same subject would also contain. Both the figurative and non-figurative depictions fictionally portray the same scene, and thus prompt the same kind of imagining. So, Walton might ask, why not enrich this imagining by filling in the canvas with the appropriate colour and texture? He does not manage to explain how the use of non-figurative techniques to abstractly conceptualise this scene makes explicit the preconscious structures of spatial orientation, represented by Mondrian as criss-crossing lines, that enable cognition of distance and depth.

These examples from Rothko and Mondrian show two distinct features of non-figurative artworks that remain unaccounted for by Walton's theory:

1. Non-figurative artworks have emotional resonances that expose our own receptivity to abstract colour, texture and form.
2. Non-figurative artworks draw attention to underlying structures of perceptual experience—geometrical, spatial or otherwise—through abstraction of familiar shapes and scenes.

In both cases, our responses to such artworks have a reflexive component, where what is understood is not simply what is *in* the artwork (the associations it produces, or imaginings it prescribes), but in our own activity of apprehension. Non-figurative artworks make features of our own cognition visible—its affective comportments to colour, for example, or its dependence on horizontal lines to orient perception.

3. Photographic Transparency

In this section, I explore Walton's commitment to a realist understanding of photography. I argue that Walton's emphasis on the reality of photographs is dependent on a notion of representability that is in tension with his above theory of fictionality. Given that we might expect of a theorist a consistent approach to interpreting the visual arts, encompassing painting as well as photography, I consider this tension an important feature of his work deserving of critical appraisal.

I wish to problematise the relationship between photography and painting in Walton's theory by way of comparison with another philosopher of perception and

aesthetics, Maurice Merleau-Ponty. In the *Phenomenology of Perception*, Merleau-Ponty posits that the structures unique to human perception grasp the world in ways that are distorted by photography, but captured by painting. He points out that the perspective of a photograph is fundamentally artificial, and has little in common with ordinary human perception. Painters, on the other hand, are able to reproduce the peculiarities of ordinary vision, as Cézanne does with his subtly warped representations of bowls and plates that do not correspond to the photographic perspective, but do reflect their actual appearance to human eyes. The fact that our visual apprehension of the world does not correspond to the geometrical perspective, he argues, is reflected in the way my subjective perceptions of an approaching object change more quickly than the change in size of the physical image projected onto my retina would suggest:

This is why the train that approaches us in a film gets larger much more quickly than it would in reality. This is why a hill that seemed quite elevated becomes insignificant in a photograph. Finally, this is why a disc placed diagonally in relation to our face resists the geometrical perspective as Cézanne and other painters have shown in representing a soup plate in profile with the inside remaining visible.¹³

This is the precise opposite of what Walton defends. Whereas Merleau-Ponty's position, consistent with my own, is that painting captures pre-conscious structures of cognition, Walton claims that paintings are best understood as fictions, and require interpretation in terms of how they elicit unreal imaginings. Moreover, Merleau-Ponty sets up painting in opposition to photography, which he believes *distorts* human perceptual experience. Walton, by contrast, defends the provocative claim that 'photographs are transparent', since when looking at one the viewer 'sees, literally, the scene that was photographed'. Whereas "photographs are counterfactually dependent on the photographed scene", paintings are also dependent on the psychological states of the person producing them, meaning a painter necessarily misrepresents their subject.¹⁴

The basis for this claim is the direct causality involved in the production of a photograph: "to see something is to have visual experiences which are caused, in a certain manner, by what is seen" and photographs meet this criterion—something has produced an image via mechanical means, whether on film or digitally, without the mediation of a fallible human subject.¹⁵ Unlike the subject of a painting, which must be interpreted and thus distorted by a painter, the subject of a photograph is involved in a causal relationship that can be equated with ordinary perception. The crucial point here is that photographs, according to Walton, are 'transparent' and offer direct perception of their subjects, with the result that they are not *representations*—they do not admit of interpretation using the language of fictionality.

¹³ Merleau-Ponty 2013, p. 271.

¹⁴ Walton 2008, p. 86.

¹⁵ *Ibid.*, pp. 97-100.

There is an interesting resemblance between Walton's formulation of photographic realism and Locke's defence of realism in the *Essay Concerning Human Understanding*. Here, Locke argues that an object has primary 'powers' to produce *secondary* sensations which represent it: although we do not see the primary qualities of a snowball, for example, it has 'powers' to produce secondary sensations of whiteness and coldness which give us the perceptual experience of seeing one.¹⁶ True, says Walton, in looking at a photograph we are not experiencing the depicted object as it is in itself, but what we do see, the photograph, is caused by it in a manner that is enough to say the visual experience we have *is* of the object itself.

As Berkeley explains in his critique of Lockean realism, such a distinction between an object's real qualities, and how they are perceived, is ultimately untenable if we want to claim that the resulting perception is *directly of that object*.¹⁷ While the 'secondary', qualities of a snowball (coldness, and whiteness) may give us good reason to infer the constitution of a real, external object with certain 'primary' qualities (physical ones such as low temperature and high reflectivity), there is still a logical gap connecting the one to the other—they are not the same thing. The snowball is ultimately only known through its representation, which cannot be directly identified with the object it represents. In much the same way, Walton admits that there is a difference between a photograph and what it depicts, but insists that when viewing the photograph, we really 'see' the subject in the depiction. This distinction relies on an unacceptable identification of *representation* with the *represented* that is inconsistent with his claim that photographs are "transparent" and enable *direct* perception of their subjects.

More problematically for Walton's theory, the representative nature of photographs means 'twofoldness' is inherent to the viewing of them. Recall that twofoldness refers to the fact that a composition is not literally its physical elements (brushstrokes, pigments and so on), but also what it depicts (a person, a mountain). Like paintings, photographs have 'configurational' elements—the ink on the photograph, or pixels on a screen—which provoke 'recognitional' experiences of what they represent—people, animals, and so on. By representing their subjects in this way, photographs behave more like props which prompt us to imagine what they depict. As Walton himself says, when viewing photographs, "fictionally, one is in the presence of what one sees."¹⁸ The problem for Walton is that when the experience of a work is principally representational, his theory of fictionality ought to be deployed to explain it. When I look at a photograph from the American Civil War, I imagine or *fictionally experience* seeing Lincoln. But if photographs are better explained like this, in *fictional* terms, then we are no longer committed to a theory of photographic 'realism', since the interpretive emphasis moves from the photograph's authentic reproduction to the manner in which it prescribes an imagining of its subject. If Walton's theory of the visual arts is to be internally coherent, then photographs must be subject to the same method of interpretation as paintings, where fictionality in representation takes priority. Photographic realism is therefore an anomalous and contradictory position

¹⁶ Locke 1979, sec. 2.8.8.

¹⁷ See Berkeley 2012, pp. 168-70.

¹⁸ Walton 2008, p. 89.

for Walton to defend in light of his commitment to the value of fictionality as a theory of representation.

According to Walton, photography accurately captures the world in a way that is perfectly analogous to seeing, while the role of painting is only to direct imagination through fictional games. This prioritises photography as the aesthetic medium with the best claim to faithful communication of perceptual experience. What I have shown, however, is that both painting and photography are ultimately representational, and thus deserving of criticism according to a single aesthetic theory of representation. Merleau-Ponty's example of the way photography misrepresents the human perspective on the world, while impressionist painting accurately reproduces it, suggests that whatever that theory may be, it must be prepared to acknowledge what insights painting offers for our understanding of perception.

4. Conclusion

In this essay I have attempted to examine the overall strength of Walton's views on the visual arts. While his theory is an attractive tool for examining figurative artworks, and offers a compelling answer to the problem of perceptual twofoldness, it fails to adequately describe aesthetic encounters with non-figurative works. I have also argued that Walton's account of photographic realism collapses into a representational theory, which theoretical consistency requires him to explain in terms that contradict his realism.

Through painting, whether figurative or not, artworks communicate truths about human experience and perception. Photographs, by contrast, distort them. We might say that the artworks Walton supposes most exhibit fictionality (paintings) disclose the affective and spatial elements of perceptual apprehension *better* than the artworks he claims most explicitly communicate vision (photographs). A crucial function of paintings is therefore to expose the mechanisms of perception unique to the human perspective. In appreciating art, particularly non-figurative art, the cognitive activities undertaken by the viewer make themselves visible; abstract works are aesthetic provocations to reflect on our own subjectivity. Walton's mistake is to conceive of artworks in terms that emphasise their unreality—their *fictionality*—ignoring how non-figurative works direct our experiences towards a better understanding of our emotive and perceptual faculties.

References

- Berkeley G. (2012) *Philosophical Writings*. Cambridge: CUP.
- Doss, E (2011) 'Makes Me Laugh, Makes Me Cry: Feelings and American Art,' *American Art*, 25: 2-8.
- Gorner P. (2000) *Twentieth Century German Philosophy*. Oxford: OUP.
- Heidegger M. (2006) *Being and Time*. Malden: Blackwell.
- Locke, J. (1979) *An Essay Concerning Human Understanding*. Nidditch PH, editor. Oxford: OUP.
- Merleau-Ponty M. (2013) *Phenomenology of Perception*. London: Routledge.
- Mondrian, P (1915) *Composition 10. Pier and Ocean*. Kröller-Müller Museum. Available at: <https://krollermuller.nl/en/piet-mondriaan-composition-10-in-black-and-white>
- Rothko M (1960) No. 14. San Francisco Museum of Modern Art. Available at: <https://www.sfmoma.org/artwork/97.524>
- Walton K (1990) *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Cambridge: Harvard University Press.
- Walton K. (2008) *Marvellous Images: On Values and the Arts*. Oxford: OUP.
- Wollheim R. (1998) 'On Pictorial Representation,' *The Journal of Aesthetics and Art Criticism*, 56: 217-26.



Based in Australasia.

Publishing quality ideas
from undergraduates worldwide



Undergraduate Philosophy Journal of Australasia – UPJA



<https://forms.gle/we6ZeihAx1hDKf6E6>



contact@upja.online



www.upja.online