dialectica

International Journal of Philosophy

Contents

EUROPEAN SOCIETY FOR ANALYTIC PHILOSOPHY

Vol 77 : 3
September 2023

dialectica

International Journal of Philosophy
Official Organ of the European Society of Analytic Philosophy

founded in 1947 by Gaston Bachelard, Paul Bernays and Ferdinand Gonseth

Editorial Board

Jérôme Dokic, EHESS, Paris, France
Pascal Engel, EHESS, Paris, France
Manuel García-Carpintero, Universitat de Barcelona, Spain
Diego Marconi, Università di Torino, Italy
Carlos Moya, Universitat de Valéncia, Spain
Martine Nida-Rümelin, Université de Fribourg, Switzerland
François Recanati, Collège de France, Paris
Marco Santambrogio, Università degli Studi di Parma, Italy
Peter Simons, Trinity College Dublin, Ireland
Gianfranco Soldati, Université de Fribourg, Switzerland
Marcel Weber, Université de Genève, Switzerland

Editors

Fabrice Correia, University of Geneva
Philipp Blum, University of Lucerne
Marco R. Schori, University of Bern (managing editor)

Review Editors

Stephan Leuenberger and Philipp Blum

Editorial Committee

Sara Amighetti, Joshua Babic, Davood Bahjat, Philipp Blum (né Keller), Claudio Calosi, Alessandro Cecconi, Zoé Christoff, Fabrice Correia, Matthias Egg, Andrea Giananti, Martin Glazier, Aleks Knoks, Arturs Logins, Jörg Löschke, Giovanni Merlo, Robert Michels, Ryan Miller, Michael Müller, Paolo Natali, Donnchadh O'Conaill, Simone Olivadoti, Edgar Phillips, Stephanie Rennick, Sebastian Schmidt, Marco R. Schori, David Schroeren, Mike Stuart, Daniel Vanello.

Consulting Board

Johannes Brandl (Salzburg), João Branquinho (Lisboa), Elke Brendel (Bonn), Ingar Brinck (Lunds), Eros Corazza (Ikerbasque and Carleton), Josep Corbi (Valéncia), Michael Esfeld (Lausanne), Dagfinn Føllesdal (Stanford and Oslo), Frank Jackson (Australian National University, Canberra), Max Kistler (Paris I), Max Kölbel (Wien), Jan Lacki (Genève), Karel Lambert (Irvine), Paolo Leonardi (Bologna), Fraser Macbride (Manchester), Josep Macià (Barcelona), Genoveva Martí (Barcelona), Élisabeth Pacherie (Institut Jean Nicod, Paris), David Piñeda (Girona), Wlodek Rabinowicz (Lund), Barry Smith (Buffalo), Christine Tappolet (Montréal), Neil Tennant (Ohio State), Mark Textor (King's College London), Achille Varzi (Columbia University), Alberto Voltolini (Torino), Timothy Williamson (Oxford).

dialectica

vol 77
issue 3

September 2023

Contents

Avoid Avoiding the Wishful Thinking Problem

Adam Patterson

The wishful thinking problem purported to be a new problem for pure non-cognitivist expressivist views in metaethics in addition to the similar, yet distinct, Frege-Geach problem. After a smattering of initial responses, discussion of the problem has faded. One might think this is because the responses were fatal, and the problem is not really a problem. I do not think so. I aim to re-start discussion of the wishful thinking problem. I do so by recasting it in terms of the distinction between propositional and doxastic justification. Doing so is instructive, for it shows some of the initial, prominent responses to the problem fail. The problem is thus not as dead as one might otherwise think.

Consider two things. First, consider the distinction between *propositional* and *doxastic* justification for some subject, S's, belief that p. On one hand, S is *propositionally* justified in believing that p when S has sufficient reasons, R, to believe that p. S's belief that p is justifi*able*, in other words. On the other hand, S's belief that p is *doxastically* justified when S believes that p, *given* (or *on the basis of*) R (Silva and Oliveira 2024). That is, S's belief that p is justifi*ed* (Korcz 2000).[1] The crucial difference is the basing relation: there is a difference between *having* (available, or at hand) R to believe that p and actually believing p *based on* R (Alston 1985). Second, consider the younger cousin of the popular Frege-Geach problem for non-cognitivist, expressivist meta-ethical views: the oft-neglected wishful thinking problem (Dorr 2002). While the latter is about *validity*, the former is about *justification*. How are these two things related?

The above distinction plays an important role in both characterizing the problem and evaluating its proposed solutions. Yet few (if any) explicitly acknowledge this. I remedy that here. Doing so is instructive, for as we will see,

1 Henceforth, "S justifi*edly* believes that p" (and similar expressions) means the belief is doxastically justified, and "S justifi*ably* believes that p" means the belief is "propositionally justified."

understanding both the problem and its solutions with the above distinction in mind reveals the ways that the responses fail. This is an interesting result, for the problem is thus not as dead as it seems.

In section 1, I recast the wishful thinking problem in terms of two kinds of justification. In section 2, I do the same for several prominent responses. I also argue that they fail. In section 3, I conclude.

1 The Wishful Thinking Problem: Recast

Consider the following moral-descriptive[2] *modus ponens* (Dorr 2002). Call it the *Liar Argument* (Long 2016).[3]

LIAR ARGUMENT

(P1) If lying is wrong, the souls of liars will be punished in the afterlife.
(P2) Lying is wrong.
(C) So, the souls of liars will be punished in the afterlife.

As we can see, moral-descriptive *modi ponentes* have as their major premise a conditional claim. That conditional's antecedent is a moral claim, whereas the conclusion is a non-moral (descriptive) claim.

Now consider Edgar. Edgar is reasoning himself through the LIAR ARGUMENT. The states of affairs as he does can be represented as follows (Dorr 2002, 98):

T_1 Edgar's belief that (P1) and ?(C)[4] are both doxastically justified. He believes ¬(P2).

At T_1, it seems *irrational* for Edgar to believe (C) for two reasons. First, it is incoherent to believe that (C) given that he already justifiably believes (P1)—on the basis of reliable testimony—and also believes that ¬(P2). Second, any belief that (C) at this time lacks *propositional* justification. Edgar, in fact, right now has *good reason* to be ambivalent about (C) and is ambivalent precisely because of those reasons (Dorr 2002, 98).

2 The "moral-descriptive" label is from Schroeder (2011).
3 Mabrito (2013) calls this "the damnation argument."
4 The "?C" denotes ambivalence about C. I borrow this from Guan (2014).

Now suppose that Edgar then reads some moral philosophy. As a result, he reconsiders his moral beliefs and thereby comes to immediately, justifiably believe (P2). Thus:

T_2 Edgar has doxastic justification for (P1), ?(C), and (P2).

He now does as attitude-coherence demands: He revises. Given that Edgar has doxastic justification for (P1) and (P2), he jettisons his previous ambivalence about (C). So, he comes to believe (C) on the basis of (P1) and (P2). Hence:

T_3 Edgar has doxastic justification for (C).

According to Dorr (2002), pure non-cognitivist expressivism[5] struggles to explain cases like this. Why is that?

At T_1, it is irrational for Edgar to believe (C). This seems plausible. After all, Edgar has no justification to believe that (C). Moreover, he cannot justifiably believe that (C) on the basis of what he believes at T_1 on pains of incoherence. So, any belief that (C) of Edgar's is neither propositionally nor doxastically justified.

Now, from T_1–T_2, Edgar's mental states changed. In particular, it seems like he underwent a change in beliefs; he gained a new one. He *came to believe that* (P2) for the first time at T_2, and justifiably so. Now, a diachronic change in mental states in general is compatible with non-cognitive expressivism. The trouble is that on non-cognitive expressivism, Edgar did not gain a belief. Rather, his *non-cognitive states changed*. That is, Edgar only gained a new *non-cognitive* state. This is because (P2), remember, is a moral claim. And that the state of accepting a moral claim is a *non-cognitive state* is part and parcel of non-cognitivist expressivism.

This is bad for non-cognitive expressivism. It means that on that view, Edgar *still* cannot justifiably believe that (C) at T_2. That is, at T_2, Edgar's belief that (C) still lacks doxastic justification for him—just as did for him at T_1. This is because a new non-cognitive state cannot be that on the basis of which Edgar justifiedly believes that (C). So, for the non-cognitive expressivist, nothing changed from T_1–T_2 that explains why it seems intuitively rational for Edgar to believe (C) *on the basis of* (P1) and (P2), or why Edgar seems rational to justifiedly believe that (C) on the basis of only accepting the premises.

5 I henceforth drop the "pure."

doi: 10.48106/dial.v77.i3.01

Said differently, it is intuitively plausible that the following two states can obtain over time in the Edgar case:[6]

(i) At T_1, it is *irrational* for Edgar to believe (C). For he justifiably believes (P1), ?(C), and believes ¬(P2). So, the belief that (C) is not propositionally justified and cannot be doxastically justified (C) given what he believes;

and

(ii) It is *rational* for Edgar at T_3 to believe (C). He has doxastic justification for both (P1) and (P2), and he accepts (C) on their basis. So, he has doxastic justification for (C).

Yet on non-cognitivist expressivism, both cannot obtain. On that view, (ii) is not possible. This is because Edgar's belief that (C) at T_3 is neither justified nor justifiable. Remember: this was also true at T_1. The only diachronic change in Edgar's mental states was his attaining a new *non-cognitive state* at T_2. But non-cognitive states cannot be justifiers, things on which it is rational to base our beliefs; they are just the wrong kinds of things.[7]

Call this the *wishful thinking problem* for pure, non-cognitivist expressivism.[8] It concerns justification. More precisely, given the *doxastic/propositional justification* distinction, one can see that it concerns whether Edgar's coming to believe (C) over time is rational because (and insofar as) said belief

6 This is how Mabrito (2013) perspicuously frames the problem.

7 Dorr (2002, 99) seems to implicitly rely on this to explain why "only a change in one's cognitive states, or in one's evidence, can make the difference between a case in which it would be irrational to believe something and one in which it would be rational to do so." But why think that non-cognitive states cannot be justifiers? Because revising your views about the world is rational when the change coheres with your *belief set*. It is, he says, "irrational to modify your views about the world so that they cohere with your desires and feelings" (Dorr 2002, 99). But again, one might ask: Why? Pryor's (2005) discussion of Davidson might be relevant. According to Pryor (2005), for Davidson, if some state, x, is a justifier, then x has propositional content, content that is expressible with "that-clauses" and which functions to assertively represent the world as being-such-and-such-a-way. This is because only by standing in *logical* relation to a belief can some state doxastically justify a belief. And since the state of accepting (P2) is, by dint of non-cognitive expressivism, a *non-cognitive state*, and these do not function to assertively indicate anything about the world, it cannot stand in a logical relation to Edgar's belief that (C). Hence, Edgar's coming to believe (P2), on pure non-cognitivist expressivism, cannot doxastically justify (C).

8 On non-cognitive expressivism, moral evaluations seem like danglers qua justifiers.

becomes *doxastically justified* after having been not previously even propositionally justified, i.e., justifi*able*.

With this in mind, let us now re-evaluate some proposed solutions. As we will see, I find them all wanting, given this understanding of the problem.

2 Re-Evaluating Some Proposed Solutions

2.1 *The Decalogue Proposal*

Consider the following argument (Lenman 2003). Call it the *decalogue argument* (Schroeder 2011). Suppose that S is reasoning through it over time, and we can represent how his beliefs seemingly rationally change over time like this:

T_1 (P3) S never contravenes the Decalogue.
(P4) All and only contraventions of the Decalogue are wrong.
(P5) S never does anything wrong.

Currently, for S to believe the descriptive claim that "S never looks at a woman with lustful intent" would be neither justifi*ed* nor *justifiable*. It is irrational for S to believe that at T_1. Now, suppose that on the basis of (P3) and (P4), S comes to *justifiably believe* the following:

T_2 (P6) If looking at a woman with lustful intent is wrong, then S never looks at a woman with lustful intent.

Thus, at T_2, the belief that (P6) is doxastically justified for S; it is justifiedly believed on the basis of a pair of claims, (P3) + (P4). However, S's belief in the descriptive claim "S never looks at a woman with lustful intent" is still neither justified nor justifiable at T_2 and is irrational to believe.

Moving on, S subsequently comes to believe two *more* things:

T_3 (P7) Looking at a woman with lustful intent contravenes the Decalogue.
(P8) Looking at a woman with lustful intent is wrong.

More precisely, here S gains at least one new belief: (P8). S believes it, given (or on the basis of) his belief in (P4) + (P7). So, the belief that (P8) is doxastically justified for S. Also, at T_3, notice that it is justifi*able* for S to believe that "S

doi: 10.48106/dial.v77.i3.01

never looks at a woman with lustful intent." This is because (P3) + (P7)[9] is sufficient reason to believe it. Finally, after all this, S comes to justifiedly, rationally believe that (C) on the basis of (P6) + (P8).

T_4 (C1) S never looks at a woman with lustful intent.

What is the point of all of this? In general, Lenman's (2003) goal seems to be to show that it can be rational to infer the conclusion of a moral-descriptive *modus ponens* without wishful thinking. This is achieved by showing that S is guaranteed to have evidence for (C1) that can justify S's coming to believe (C1) without wishful thinking; S's justification for believing (C1) is guaranteed to be overdetermined, in other words.[10] For in the very act of accepting a moral claim like (P8), one is guaranteed to accept beliefs that support it and also support (C1) independently of (P6) and (P8).

I have said nothing yet about propositional/doxastic justification. But now I ask: how can one understand the *decalogue argument* and this proposed solution in general, given the propositional/doxastic justification distinction? Asked differently: how can this solution be recast with the distinction—with it, how can one explain the way in which this proposal vindicates non-cognitive expressivism from the wishful thinking problem?

To see, recall (C1):

(C1) S never looks at a woman with lustful intent.

Also, recall that the solution works, in part, by guaranteeing overdetermined justification for S's belief that (C1) by the very process of coming to believe it in the first place. Further, recall: this overdetermined justification is secured by finding two pairs of claims. One is (P6) + (P8):

(P6) If looking at a woman with lustful intent is wrong, then S never looks at a woman with lustful intent.
(P8) Looking at a woman with lustful intent is wrong.

The other is (P3) + (P7):

(P3) S never contravenes the Decalogue.

9 And remember: both (P3) and (P7) were each part of the arguments for claims that make S's belief that (C1) doxastically justified, namely (P6) and (P8).

10 I borrow characterizing this proposal in terms of guarantees from an anonymous referee.

(P7) Looking at a woman with lustful intent contravenes the Decalogue.

Both sets justify S's coming to believe (C1). Hence, with the relevant distinction between kinds of justification in mind, the proposal works by establishing one of two things. S's belief is guaranteed to have, for S,

(iii) overdetermined doxastic justification;

(In which case, the *decalogue case* shows that S's belief that (C1) is always partly based on (P3) + (P7). So, said belief is doxastically justified, i.e., based on two sets of claims, one set of which lacks a moral claim. And thus, basing the belief on (C1) is always, in part, not wishful thinking on S's part.) or,

(iv) overdetermined propositional justification.

(In which case, two sets of claims always support (C1), and thus S's coming to believe (C1) is rational. This is because (and insofar as) while S's believe that (C1) is justifi*edly inferred* on the basis of a non-cognitive claim—(P8)—the inference is guaranteed to be justifi*able* given (P3) + (P7).) This puts the non-cognitivist in a dilemma.

If the proposal establishes (iii), then it relies on the following assumption: Namely, for all of S's doxastically-justified moral beliefs, whenever a moral belief is doxastically justified for S on the basis of R, and that moral belief entails some non-moral claim, then R doxastically justifies S's belief in the non-moral claim. This is dubious, though. S can be unaware of what *propositionally justifies* the non-moral belief, which is the fact that it is entailed by the moral belief. This means that S need not necessarily form the non-moral belief *on the basis of* the moral belief, in which case that non-moral belief is not doxastically justified.

Moreover, if the proposal establishes (iii), then S is still basing their belief that (C1) *in part* on the basis of (P8). There is still wishful thinking present; there is just less of it on this proposal since it is also based on (P3) and (P7)—neither of which are non-cognitive claims given non-cognitive expressivism. The proposal would be better to establish the following: in coming to believe (C1) at T_4, S's belief is *only* based on that other pair of claims—the pair that lacks a moral claim, namely, (P3) + (P7)—and thereby makes the justified belief in (C1) not a case of wishful thinking.

doi: 10.48106/dial.v77.i3.01

If the proposal establishes (iv), then it does not help with the wishful thinking problem. It needs to be shown that S is rational because (and insofar as) S went from, at one time, having no justification for believing (C1) to having doxastic justification for it (and crucially without believing (C1) on the basis of a non-cognitive attitude). But if the proposal establishes (iv), this still happens: S still comes to accept (C1) on the basis of (P8). So, S's belief is based on, problematically, a non-cognitive attitude. That S's belief is guaranteed some bonus propositional justification may lessen the sting of a charge of irrationality. But it stings nonetheless.

So, if the proposal showed that S's belief in (C1) is rational insofar as the belief is guaranteed to be always justi*fied* or justifi*able* for S without wishful thinking, then it either relies on a false assumption or fails to address the problem. Either way, the wishful thinking problem remains.

2.2 *The Modified Proposal*

Another way to deal with the *wishful thinking problem* is to attempt to argue that, in reasoning through the LIAR ARGUMENT, S is guaranteed to be *propositionally justified* in believing (C). So, for example, whenever S argues through the LIAR ARGUMENT, S has available to them at any time via introspection the following modified, companion argument (Enoch 2003):

MODIFIED ARGUMENT[11]

(P9) If I accept that lying is wrong, the souls of liars will be punished in the afterlife.
(P10) I accept that lying is wrong.
(C) So, the souls of liars will be punished in the afterlife.

This MODIFIED ARGUMENT is always available through introspection. The idea is thus that the LIAR ARGUMENT will never lead Edgar to irrationally believe (C) in the sense that S believes it without that which sufficiently justifies it.

11 Here, I continue the sequential numbering of premises from the decalogue argument. This does not mean that the MODIFIED ARGUMENT is a part of (or some extension of) the *decalogue argument*. I use this numbering convention to avoid referring to one premise in a distinct argument with the same expression.

This line is also problematic. It fails to guarantee that S's belief that (C) is justifiable for S. Why is that? The MODIFIED ARGUMENT is "available" to S in a weak sense. It is guaranteed to be possible that S can come to possess a MODIFIED ARGUMENT. But this does not entail that S actually *has available* (or is in the possession of) the relevant argument. The mere *presence* of the MODIFIED ARGUMENT only propositionally justifies Edgar's acceptance of (C) only if Edgar actually *has accepted* (P9) and (P10).

Suppose that I am wrong. Suppose that, on this strategy, any time S accepts (P1) and (P2), S will always possess (in some suitably strong sense) propositionally justification for (C) since there will always be other things that propositionally justify (C).

This proposal does not address the problem, either, for familiar reasons. In particular, the problem is that when S forms the belief that (C) *on the basis of* both (P1) and (P2),

(P1) If lying is wrong, the souls of liars will be punished in the afterlife;
(P2) Lying is wrong;

then S's subsequent belief that (C) is doxastically justified. And that means that S's belief that (C) is doxastically justified whether S is also justified in having some *other, auxiliary belief(s)*. Said differently: what generates the problem in the original case is that it seems rational to believe that (C) because the belief is doxastically justified for S regardless of whether S has some other, auxiliary claims available that are themselves doxastically justified.

2.3 *The Entailment Proposal*

Another way to think about the *wishful thinking problem* is that there is a condition that needs to be met for the states (i) and (ii) (section 1) to simultaneously obtain, and the pure non-cognitivist expressivist cannot satisfy it. As we saw, the constraint seems to be something like this: S's belief that *p* goes from not doxastically justified to being just that only if S acquires a new *cognitive state*. Call this *Dorr's constraint* (Mabrito 2013, 1072). Given this, the *decalogue problem* and *modified proposals* can be thought of as attempts to show that the non-cognitivist expressivist can meet *Dorr's constraint*.

This is not the only way to vindicate non-cognitive expressivism, though. One can also attempt to argue that S can meet the constraint and still be rational (Mabrito 2013). How might one show this? One idea is to say that

doi: 10.48106/dial.v77.i3.01

while the Edgar case violates *Dorr's constraint*, it is compatible with another, independently motivated constraint (Mabrito 2013). The obeyance of *this* constraint vindicates the intuitive rationality of Edgar coming to believe (C) in the LIAR ARGUMENT. Here is the argument again:

LIAR ARGUMENT

(P1) If lying is wrong, the souls of liars will be punished in the afterlife.
(P2) Lying is wrong.
(C) So, the souls of liars will be punished in the afterlife.

This other constraint is called the *entailment constraint* (EC). With respect to the *wishful thinking argument*:

> ENTAILMENT CONSTRAINT (EC). S moves from T_1 (during which S lacks justification for believing that p) to T_2 (during which S has justification for believing that p) only if S comes to accept claims that entail p or acquires evidence that supports p (Mabrito 2013, 1074).

Edgar's case seems compatible with this because the *wishful thinking problem* assumes that the *Frege-Geach problem* is solved: it concedes that moral-descriptive *modi ponentes* are valid. So, while Edgar initially lacks justification for believing (C), he later comes to accept two claims that entail (C)—securing the obeyance of the ENTAILMENT CONSTRAINT—and is thereby rational for believing (C) on the basis of them.

Again, though, we should think of how this solution goes in terms of the doxastic/propositional justification distinction. To illustrate, consider the ENTAILMENT CONSTRAINT itself. As a necessary condition, with respect to changes in what kinds of justification is the condition plausible? And obeying which of these various formulations also helps with the *wishful thinking problem*?

Suppose that the relevant change concerns a belief's status with respect to being propositionally justified:

> ENTAILMENT CONSTRAINT 2 (EC2). S moves from T_1 (during which S lacks propositional justification for believing that p) to T_2 (during which S has propositional justification for believing that p) only

> if S comes to accept claims that entail p or acquires evidence that supports p.

(EC2) is plausible but irrelevant. No one denies that Edgar's belief that (C) of the Liar Argument is justifi*able*—i.e., denies that there are reasons to accept it. The worry is that Edgar's belief seems justifi*edly inferred* from his belief in the premises. So, a case that obeys *this* constraint is explanatorily moot with respect to the relevant intuitive rationality that needs preservation on the *wishful thinking problem*.

Suppose that the salient change is from a belief being not justifi*able* to actually being justifi*ed*. Thus:

> Entailment Constraint 3 (EC3). S moves from T_1 (during which S lacks propositional justification for believing that p) to T_2 (during which S has doxastic justification for believing that p) only if S comes to accept claims that entail p or acquires evidence that supports p.

(EC3) is false. One can both believe that p for no reason and then only later base one's belief that p on very good reasons, and yet neither come to believe that which entails p nor acquire evidence for p. This happens in cases where one dogmatically believes that p but only later comes to accept p on the basis of good reasons that one already had. This is because one can fail to recognize good reasons for beliefs when they have them. Hence, one can fail to base that which they already believe on the basis of those good reasons.

2.4 *The Hopeful Proposal Proposal*

Another proposed solution to show that wishful thinking is sometimes rational is accepting that the premises of arguments akin to moral-descriptive *modi ponentes* do make the belief in the conclusion justifi*able* (Long 2016). As an example, consider the following argument that S is reasoning through. Call it the *hopeful proposal argument*.

> Hopeful Proposal Argument
>
> (P11) If I hope that my proposal will be accepted, then my proposal will be accepted.
> (P12) I hope that my proposal will be accepted.
> (C2) So, my proposal will be accepted.

doi: 10.48106/dial.v77.i3.01

This argument is *wishful thinking* as it is normally understood outside the seminar room: forming *beliefs* about how the world *is*, given one's wants, desires, hopes, dreams, etc. (which tell us how the world is *not*). Now, how does one argue that it is rational to infer (C2)? The idea is that "accepting of the [...] premises is often a reason to accept its conclusion, since paradigm cases of wishful thinking are often valid" (Long 2016, 3).

Once again, the propositional/doxastic justification distinction renders this proposal ambiguous. So, for example, either S's belief that (C2) is rational because (and to the extent) that it is *doxastically* or *propositionally* justified given the fact that *modus ponens* is valid.[12]

No one would deny that S coming to believe (C2) is justifi*able*, given the validity of *modus ponens*. The issue, recall, is whether S is justifiably rational in going on to believe (C2) *on the basis of* the premises—as it intuitively seems. So, if this proposal is to work, the very fact of the argument's validity must be a part of that set of things on the basis of which S believes that (C2).

Here is how this works with respect to the LIAR ARGUMENT. S rationally, justifiedly infers (C),

(C) So, the souls of liars will be punished in the afterlife;

when the belief that (C) is based on the following:

(P1) If lying is wrong, the souls of liars will be punished in the afterlife;
(P2) Lying is wrong; and
(P13) (P1) and (P2) entail that the souls of liars will be punished in the afterlife.

The trouble is that this proposal misses the mark. The mere availability of (P13) does not mean that S is doxastically justified in believing (C). S can fail to base their belief in (C) on (P13). S would need to be shown to, in every case, in fact base their belief in (C) on the basis of (P1) + (P2) + (P13).

Moreover, the *wishful thinking* problem remains. The non-cognitivist expressivist still cannot make sense of how S's belief that (C)—even in that case—is doxastically justified. For (P2), on their view, is still a non-cognitive state. And S cannot, it seems, justifiedly believe (C) *on the basis of* (P1) + (P2) + (P13) for that very reason: (P2) is the wrong kind of thing for one to justifiedly believe something else on the basis of. It would help if (P3) was itself sufficient for S to justifiedly believe (C) on the basis of. We saw attempts at this. But the problem

12 This assumes a solution to the Frege-Geach problem.

with those attempts remains. It would have to be shown that in every case, S in fact believes the conclusion on the basis of the stuff that is both guaranteed to be present; would, in fact, make the belief doxastically justified if the belief was based on them; and are the right kinds of things to base beliefs on.

3 Conclusion

The wishful thinking problem seemed dead, the recipient of several fatal blows. I hope to have shown that the distinction between propositional/doxastic justification helps clarify the nature of the problem, the nature of proposed solutions, and why those solutions are dubious. Perhaps, then, it is premature to ignore the wishful thinking problem.*

Adam Patterson
XXXX-XXXX-XXXX-XXXX
Syracuse University
apatters@syr.edu

References

Alston, William P. 1985. "Concepts of Epistemic Justification." *The Monist* 68(1): 57–89, doi:10.5840/monist198568116.

Dorr, Cian. 2002. "Non-Cognitivism and Wishful Thinking." *Noûs* 36(1): 97–103, doi:10.1111/1468-0068.00362.

Enoch, David. 2003. "How Noncognitivists Can Avoid Wishful Thinking." *The Southern Journal of Philosophy* 41(4): 527–545, doi:10.1111/j.2041-6962.2003.tb00965.x.

Guan, Chengying. 2014. "The Wishful Thinking Problem for Non-cognitivism: Does It Really Make Sense?" *Kriterion – Journal of Philosophy* 28(1): 30–44, doi:10.1515/krt-2014-012804.

Korcz, Keith Allen. 2000. "The Causal-Doxastic Theory of the Basing Relation." *Canadian Journal of Philosophy* 30(4): 525–550, doi:10.1080/00455091.2000.10717542.

* I received many great comments on this paper. I'd like to thank the audiences at the Philosophy in Progress conference at the University of Nottingham and the Syracuse Working Papers Workshop for helpful comments and discussion. In particular, I'd also like to thank the following individuals for their useful feedback: Neil Sinclair, Preston Werner, Kris McDaniel, Hille Paakkunainen, Nate Sharadin, and Evelyn Hudson. Finally, I would also like to thank anonymous referees from the following journals for their thorough comments: *Dialectica*, *Analysis*, *Thought*, and the *Journal of Ethics & Social Philosophy*.

LENMAN, James. 2003. "Noncognitivism and Wishfulness." *Ethical Theory and Moral Practice* 6(3): 265–274, doi:10.1023/a:1026037820411.

LONG, Joseph. 2016. "Non-Cognitivism and the Problem of Moral-Based Epistemic Reasons: A Sympathetic Reply to Cian Dorr (2002)." *Journal of Ethics and Social Philosophy* 10(3), doi:10.26556/jesp.v10i3.186.

MABRITO, Robert. 2013. "Are Expressivists Guilty of Wishful Thinking?" *Philosophical Studies* 165(3): 1069–1081, doi:10.1007/s11098-012-0003-8.

PRYOR, James. 2005. "There Is Immediate Justification." in *Contemporary Debates in Epistemology*, edited by Ernest SOSA and Matthias STEUP, 1st ed., pp. 181–201. Contemporary Debates in Philosophy n. 3. Malden, Massachusetts: Basil Blackwell Publishers. Second edition: Steup, Turri and Sosa (2014, 202–221).

SCHROEDER, Mark. 2011. "How Not to Avoid Wishful Thinking." in *New Waves in Metaethics*, edited by Michael Sean BRADY, pp. 126–140. New Waves in Philosophy. London: Palgrave Macmillan, doi:10.1057/9780230294899_7.

SILVA, Paul, Jr. and OLIVEIRA, Luis R. G. 2024. "Propositional Justification and Doxastic Justification." in *The Routledge Handbook of Evidence*, edited by Maria LASONEN-AARNIO and Clayton LITTLEJOHN, pp. 395–407. Routledge Handbooks. New York: Routledge, doi:10.4324/9781315672687-36.

SOSA, Ernest and STEUP, Matthias, eds. 2005. *Contemporary Debates in Epistemology*. 1st ed. Contemporary Debates in Philosophy n. 3. Malden, Massachusetts: Basil Blackwell Publishers. Second edition: Steup, Turri and Sosa (2014).

STEUP, Matthias, TURRI, John and SOSA, Ernest, eds. 2014. *Contemporary Debates in Epistemology*. 2nd ed. Contemporary Debates in Philosophy n. 3. Oxford: Wiley-Blackwell. First edition: Sosa and Steup (2005).

Whittle vs. Cantor on the Size of Infinite Sets

Ran Lanzet

I examine several arguments by Bruno Whittle against the Cantorian conception of the size of infinite sets. I find that none of them succeed.

Sets A and B are the same size—A has exactly as many elements as B does—iff there is a bijection between them; A is at least as large as B—has at least as many elements as B does—iff there is an injection from B to A. This conception of set size has become standard following Cantor. What Cantor's theorem means, on this conception, is that the powerset of every set A is larger than A (see, e.g., Enderton 1977, 132; Hrbacek and Jech 1999, 65; Smullyan and Fitting 1996, 7–8). Thus interpreted, the theorem entails that there are different sizes of infinity (assuming, of course, that there is at least one infinite set and that that set has a powerset).

Whittle (2015a, 2015b, 2018) objects to this standard conception. He holds that we do not in fact have good reason for believing either of the following two principles, and that we are thus not in a position to know either of them or to know that there are infinite sets of different sizes:

> Size→Function
> For any sets A and B, if A is the same size as B, then there is a bijection from A to B.

> Size*→Function
> There is some "size-like" property—a property *similar* to size—size* such that, for any sets A and B, if A is the same size* as B, then there is a bijection from A to B.

Whittle puts forward a series of arguments that aim to establish this. Five of his arguments purport to refute the following widely accepted theses:

THESIS 1. There being a bijection between A and B is what it is for A and B to be the same size.[1,2]

THESIS 2. The notion of cardinality defined in set theory is at least "size-*like*": it is at least similar to our ordinary notion of size (of sets) and perhaps a natural generalization of the notion of finite size.

THESIS 3. Cantor's theorem establishes that there are infinite sets of different sizes.[3]

Here, I examine Whittle's arguments against theses 1 and 2; I argue that none of them succeed as refutations of either of those theses. The arguments, appearing in their fully developed form in Whittle (2018), are:

(i) an argument against THESIS 1 based on its interpretation as stating that "c is the same size as d" and "There is a bijection from c to d" express the same structured proposition;
(ii) an argument against the same thesis, based on what would be true in a mathematically-impossible situation in which there are no functions from certain sets;
(iii) a Benacerraf-style challenge for accounts of size in terms of functions—accounts that may be offered in support of THESIS 1;
(iv) an objection to THESIS 2 based on (ii).

Whittle also argues directly against THESIS 3 in his (2015a) and (2015b). I examine and reject his argument against this thesis in (manuscript).

1 Whittle targets this thesis and THESIS 2 below since they can be offered as grounds for SIZE→FUNCTION and SIZE*→FUNCTION.

2 Some may consider THESIS 1 too strong and prefer a more tolerant approach, on which the identification of size with cardinality-as-defined-in-set-theory is just one of several legitimate options. However, if this identification is indeed untenable, as some of Whittle's arguments purport to establish, then it is unclear how one could legitimately adopt it—even tolerantly. I thank an anonymous referee for drawing my attention to this point.

3 Whittle also attacks the following theses, which, I believe, are not as popular as the ones I mentioned in the main text: We can infer that SIZE→FUNCTION holds for sets in general from the fact that it holds for finite sets; we can infer SIZE→FUNCTION by Inference to the Best Explanation, since size differences are the only explanation of the absence of a bijection between two given sets; we are entitled to consider SIZE→FUNCTION as a basic mathematical truth; we can argue for SIZE→FUNCTION inductively, based on the consequences that it allows us to derive. I will not discuss Whittle's arguments against those additional theses here.

If I am right in rejecting the arguments against theses 1–3, then Whittle fails to establish his claim that we are not in a position to know that there are different sizes of infinity. For these three theses, if correct, provide routes to such knowledge (this is, indeed, why Whittle attacks them); and this is so whether or not some additional routes are successfully blocked by those of Whittle's arguments that I will not consider here.[4]

Although my treatment of Whittle's arguments can be considered a defense of the Cantorian view of size, I will not offer positive arguments for this view or argue that other views, incompatible with it, are incorrect or unjustified; my point is only that Whittle fails to establish this for Cantor's view.

Sections 1–4 below are each dedicated to the examination of one of the arguments (i)–(iv). I briefly conclude in section 5.

1 Structured Propositions

Consider the following two sentences:

> SIZE
> c is the same size as d.
>
> FUNCTION
> There is a bijection from c to d.

According to Whittle, the most straightforward interpretation of THESIS 1 is this:

(*) SIZE and FUNCTION express the same proposition.[5]

Whittle thinks of propositions here as structured, Russellian propositions. The propositions that SIZE and FUNCTION seem to express are, according to him,

$$(p_S)\ \langle \exists!P[\text{Size}(P) \wedge c \text{ has } P] \wedge \exists!Q[\text{Size}(Q) \wedge d \text{ has } Q] \wedge \exists R[\text{Size}(R) \wedge c \text{ has } R \wedge d \text{ has } R]\rangle$$

and

4 See footnote 3.

5 An alternative interpretation of THESIS 1 would be that SIZE and FUNCTION describe the same feature of reality. Whittle's argument against THESIS 1 on *this* reading is the one considered in the next section.

doi: 10.48106/dial.v77.i3.02

(p_F) $\langle \exists f \ f$ is a bijection from c to $d\rangle$,

respectively.

Whittle now suggests that each of the steps in the following argument against (*) is at least very plausible, or prima facie correct (Whittle 2018, 855–856):

(1) (p_S) and (p_F) are about very different sorts of things: (p_S) is about a certain sort of property—a size—that c and d have in common; (p_F) is about a function.
(2) If (*) is true, then either both SIZE and FUNCTION express (p_S), or both of them express (p_F).
(3) Most sentences about functions express propositions that are genuinely about functions.
(4) There is no plausible account on which the following sentence expresses a proposition about properties:

> CONSTANT
> There is a constant function from c to d.

(5) The sentences FUNCTION and CONSTANT seem to be very similar.
(6) From (5): FUNCTION and CONSTANT should express similar propositions—not ones about completely different sorts of things.
(7) From (1), (4), and (6); and also from (1) and (3): FUNCTION cannot express (p_S).
(8) SIZE has the same general form as the following:

> COLOR
> e is the same color as g.
>
> HEIGHT
> e is the same height as g.
>
> SEX
> e is the same sex as g.

(9) From (8): SIZE and sentences like COLOR, HEIGHT, and SEX should express similar propositions—propositions of the same general form.

(10) The propositions expressed by sentences such as COLOR, HEIGHT, and SEX are not about functions.
(11) From (1), (9), and (10): SIZE cannot express (p_F).
(12) From (7) and (11): It is not the case that both SIZE and FUNCTION express (p_S), and it is not the case that both of them express (p_F).
(13) From (2) and (12): (*) is false.

Whittle's argument here relies on the following view of structured propositions:

> PROP
> For each declarative sentence such as SIZE and FUNCTION, there is one unique structured proposition that the sentence can be taken to express; the structure of that proposition, as well as what it is about, is given by the most straightforward formalization of the sentence in higher-order predicate logic.

Without PROP, we seem to have no reason to accept the inference from (5) to (6) or the one from (8) to (9). As I will now argue, however, it is illegitimate to rely on PROP in this context. This is so, since PROP is incompatible with Whittle's reading of "what it is"—a reading on which

- There being a bijection between A and B is what it is for A and B to be the same size

is correctly interpreted as

- SIZE and FUNCTION express the same structured proposition.

PROP, together with this reading, yields absurd consequences, as illustrated by the following example.

Being a man who has never been married is just what it is to be a bachelor. But PROP, together with the aforementioned reading of "what it is," entails that this is not so. To see this, note first that the most straightforward formalizations in predicate logic of the following differ in structure:

(i) x is a bachelor.
(ii) x is a man who has never been married.

(i) and (ii) are most straightforwardly formalized as, respectively:

doi: 10.48106/dial.v77.i3.02

(iii) Bx
(iv) $M^1x \wedge \neg\exists y(M^2xy)$.

This, by PROP, means that (i) and (ii) express different propositions. Hence, on Whittle's interpretation of "what it is," being a never-married man is *not* what it is to be a bachelor.

One can, of course, hold that, in formalizing (i), we should take into account the analysis of "bachelor" as "never-married man" and correspondingly allow the formalization of both (i) and (ii) as (iv). This is, I think, a reasonable approach, but note that it deviates from PROP. Moreover, if a similar deviation is allowed in the case of SIZE and FUNCTION, then Whittle's argument does not go through, as there is no longer reason to accept, e.g., the inference from (8) to (9).

Whittle's approach entails, then, that being a never-married-man is *not* what it is to be a bachelor. Similarly, it entails that:

- There being an isometry that transforms a figure c in the Euclidean plane into a figure d is *not* what it is for c and d to be congruent.
- There being a formal proof of a first-order sentence ϕ from a first-order theory T is *not* what it is for ϕ to be a theorem of T.

Additional, similar examples are not hard to produce.

I conclude that Whittle's argument from structured propositions fails by relying on two incompatible principles: PROP and a reading of "what it is," on which THESIS 1 is correctly interpreted as (*). If PROP is given up, then, as noted above, Whittle's argument seems not to go through; if the aforementioned reading of "what it is" is given up, then the argument does go through, but its conclusion, (13), no longer means that THESIS 1 is false.

2 Sets in an Impossible Situation

THESIS 1, recall, was this: There being a bijection between A and B is what it is for A and B to be the same size. This thesis can be supported by what Whittle calls a *functional account of size*: an account of size properties of sets in terms of functions. According to Whittle, however, no such account can be correct (Whittle 2018, 860–861).

Whittle specifically considers a functional account on which complex properties are "Russellian propositions with gaps" and sizes are properties of the form

⟨There is a bijection from __ to κ⟩,

where κ is a von Neumann cardinal number (i.e., a von Neumann ordinal equipollent with none of its elements). This, of course, is not the only way to characterize size properties in terms of functions. A more natural functional account—and one that avoids commitment to a specific metaphysical account of properties—is, perhaps, this: A size-property is the property of belonging to a given bijection-type[6] (cf. Whittle 2018, fn.15). Whittle explains, however, that his argument can be adapted so as to refute *any* functional account of size, and, in fact, any account of size in terms of things other than those that constitute the set, such as its elements or those elements' parts or elements.

Whittle's argument involves the sets $A = \{0, 1\}$ and $B = \{2, 3\}$ in a (mathematically impossible) situation S that is "exactly like the actual world, except that there are no functions from either set" (Whittle 2018, 860). The argument is this:

(1) In S, there are no bijections from A or from B.
(2) In S, A and B have the same size.
(3) From (1): In S, neither A nor B has any property of the form ⟨There is a bijection from __ to κ⟩.
(4) From (3): If the functional account under consideration is correct, then, in S, neither A nor B has a size.
(5) From (4): If the functional account is correct, then, in S, A and B do *not* have the same size.
(6) From (2) and (5): The functional account is incorrect.

There is a good reason to suspect this argument, if not to reject it outright: Arguments very similar to it, and ones that are not any less appealing, lead to absurd conclusions. Consider, for example, the following account of what a *circle* is:

> Circle
> To be a circle is to be a set of all the points in the Euclidean plane that are at the same given distance r from a given point o.

I submit that this is a *correct* account of what a circle is. But an argument similar to Whittle's leads to the conclusion that this is not so: Let C be a circle,

6 A bijection type here is a class X for which the following condition holds: For some set a, X is the class of all possible sets x such that there is a bijection between x and a.

doi: 10.48106/dial.v77.i3.02

and consider an impossible situation S' that is exactly like the actual world except that there is no point at equal distances from all the elements (points) of C. (Reasoning about S' seems to make as much sense as reasoning about S does.) Then:

(1) In S', no point is at equal distances from all the elements of C.
(2) In S', C is a circle. (This seems as plausible as Whittle's premise that A and B have the same size in S.)
(3) From (1): If CIRCLE is correct, then, in S', C is *not* a circle.
(4) From (2) and (3): CIRCLE is incorrect.

(Note that I am not claiming that this is a very convincing argument—just that it is similar to, and not less appealing than, Whittle's argument.)

This "bad company" indicates that something is wrong with Whittle's argument. But what? There are, I believe, two major problems with the argument. First, it is not at all clear that we can make sufficient sense of mathematically impossible situations like S to determine the truth value of statements like Whittle's premise (2) ("In S, A and B have the same size"). Whittle claims that we *are* capable of judging what would be true in S, but it is far from clear that he is correct about that. Whittle also claims that an argument similar to his can be given not in terms of what is true under an impossible hypothesis but, instead, in terms of what is an immediate consequence of the hypothesis. If this is indeed possible, then the resulting (modified) argument would not be any more problematic than a standard reductio. It is unclear, however, how the imagined modification of the argument is supposed to proceed, and Whittle gives no indication of that. At least on the face of it, his argument is very much *un*like a reductio: It is more similar, it seems, to a demonstration of non-entailment using a counterexample; for, rather than deriving an absurdity from an impossible hypothesis, he seems to be relying on judgments made under such a hypothesis in order to reject a universal statement (to the effect that certain properties always coincide).

Second, assuming (for the sake of argument) that we can make sufficient sense of mathematically impossible situations like S, it is unclear why we should think that A and B have the same size in S. Perhaps Whittle holds that this is so since (allegedly) A and B are still of size 2 in S. But it is unclear why we should think that this is so. Especially if, in S, *there is no way of counting the elements of A or those of B*, and there is indeed no way of counting those elements in S, at least if the following jointly hold: (a) a way of counting

the elements of, e.g., A is a way of correlating them 1-1 with the elements of {"one", "two"}; (b) a way of thus correlating the elements of these two sets is a *bijection* between them; (c) in S, there are no bijections between A and any other set.

Perhaps Whittle holds that, e.g.,

> SIZE2
> The size of A is 2 in *every* situation, mathematically possible or not, in which A exists.[7]

But this, it seems, cannot serve as a ground for Whittle's premise (2), since an argument similar to his leads to the conclusion that SIZE2 is *false*: Consider a situation S'' that is exactly like the actual world, except that there are no numbers other than 0 and 1. (Reasoning about S'' seems to make as much sense as reasoning about S does.) Then:

(1) It is not the case that, in S'', something equals 2.
(2) If SIZE2 is true, then, in S'', the size of A equals 2.
(3) From (2): If SIZE2 is true, then, in S'', something equals 2.
(4) From (1) and (3): SIZE2 is false.

I am not claiming that (1)–(4) is a convincing argument or that reasoning in this way about S'' makes any sense; I only claim that this is so *by Whittle's standards*. If this claim is correct, then Whittle cannot rely on SIZE2; consequently, premise (2) of his argument remains unfounded.

I conclude that Whittle's argument from S fails as a refutation of functional accounts of size and that it is therefore ineffective against THESIS 1.

3 Benacerraf's Problem

Whittle (2018, 862–863) argues that the specific functional account that featured in his argument from an impossible situation faces a version of "Benacerraf's problem."[8] That functional account, recall, identified sizes with properties of the form

7 He may think that this is so, since the following (presumably) holds in any situation in which A exists: $\exists u \exists v (u \neq v \wedge u \in A \wedge v \in A \wedge \forall w [w \in A \rightarrow (w = u \vee w = v)])$.

8 The original problem, presented in Benacerraf (1965), arises for theories that take the natural numbers to be a particular collection of sets.

doi: 10.48106/dial.v77.i3.02

⟨There is a bijection from __ to κ⟩,

where κ is a von Neumann cardinal number. The problem is this: By replacing the cardinal numbers with other, equipollent sets, we can obtain different, incompatible functional accounts of size. Since, moreover, there is no reason to favor any one of those competing accounts over the others, each of them seems arbitrary and therefore, according to Whittle, incorrect.

This problem does not seem to afflict *all* functional accounts of size. Consider, for instance, the account I mentioned in the previous section, on which a size property is the property of belonging to a given bijection type. This account does not commit to a particular metaphysical theory of what properties are, and the Benacerraf problem, it seems, does not arise for it. Whittle would seem to agree, but he does not consider this to be a problem for his attack on THESIS 1, since he takes the functional accounts immune to Benacerraf's problem to be refuted by the argument discussed in the previous section (see his 2018, fn.15). If my arguments in the previous section are correct, however, then there *is* a genuine problem here for Whittle; for the argument he is relying on fails.

4 Are Cardinalities Size-Like?

THESIS 2 was this:

> The notion of cardinality defined in set theory is at least "size-*like*": it is at least similar to our ordinary notion of size (of sets) and perhaps a natural generalization of the notion of finite size.

Whittle takes cardinalities to be properties of the form

⟨There is a bijection from __ to κ⟩,

where κ is a cardinal number. He makes two related points against THESIS 2 (Whittle 2018, 864):

1. The notion of cardinality is not a generalization of the notion of finite size, since the collection of all cardinalities does not contain the finite sizes. This is allegedly established by the argument from *S* (discussed in section 3 above), as that argument is supposed to show that finite sizes, unlike finite cardinalities, cannot be understood in terms of bijections.

2. Cardinalities are "just a completely different sort of property from sizes" (Whittle 2018, 864). For, sizes have nothing in particular to do with functions; they are, rather, *intrinsic*: they can be accounted for only in terms of the things that constitute the set, such as its elements, their parts, or their elements. This is supposed to be established by an argument similar to the argument from *S*.

There are, however, two serious problems with this line of argument. First, if my criticism of Whittle's argument from an impossible situation (see section 3 above) is correct, then that argument fails to establish that finite sizes are distinct from finite cardinalities, and it is unclear how a similar argument could establish the supposed intrinsic nature of size properties.

Second, even if the argument from *S* did establish that sizes are intrinsic and that finite sizes are thus distinct from finite cardinalities, it would still *not* follow that sizes are not size-*like*—i.e., not similar to sizes—in any important or interesting ways, or that it is not obvious that they are. For, whether or not sizes are intrinsic, there are several well-known points of similarity between cardinalities and finite sizes that are, arguably, both interesting and important. These include the following:

(i) Finite cardinalities are at least *co-extensive* with finite sizes (Whittle concedes this).
(ii) If a subset *B* of *A* is smaller* than *A* (in the sense of cardinalities), then it is a *proper* subset of *A*; more generally, the pigeonhole principle holds: If a set *B* is smaller* than *A*, then there is no injection from *A* to *B*.
(iii) The following version of Hume's principle (which Whittle accepts for finite sizes) holds: Sets *A* and *B* have the same size* iff there is a bijection between them.
(iv) Assuming the axiom of choice, sizes* (i.e., cardinalities) are well-ordered by the relevant smaller-than* relation.

Given these problems, I submit, Whittle's points against Thesis 2 do not suffice as a refutation of it.

5 Conclusion

Whittle puts forward four arguments against theses 1 and 2; these, as I explained in the introduction, constitute an essential component of his objection

doi: 10.48106/dial.v77.i3.02

to Cantor's conception of infinite size. As I hope to have established, however, none of Whittle's arguments against those two theses succeed.*

Ran Lanzet
0000-0003-2475-562X
University of Haifa
lanzetr@gmail.com

References

BENACERRAF, Paul. 1965. "What Numbers Could Not Be." *The Philosophical Review* 74(1): 47–73, doi:10.2307/2183530.

ENDERTON, Herbert B. 1977. *Elements of Set Theory*. Cambridge, Massachusetts: Academic Press.

HRBACEK, Karel and JECH, Thomas J. 1978. *Introduction to Set Theory*. 1st ed. Pure and Applied Mathematics n. 45. New York: Marcel Dekker.

—. 1999. *Introduction to Set Theory*. 3rd ed. Pure and Applied Mathematics n. 220. New York: Marcel Dekker. First edition: Hrbacek and Jech (1978).

MCGEE, Vann. 2015. "Whittle's assault on Cantor's Paradise [on Whittle (2015a)]." in *Oxford Studies in Metaphysics*, volume IX, edited by Karen BENNETT and Dean W. ZIMMERMAN, pp. 20–32. New York: Oxford University Press, doi:10.1093/acprof:oso/9780198729242.001.0001.

SMULLYAN, Raymond M. and FITTING, Melvin Chris. 1996. *Set Theory and the Continuum Problem*. Oxford Logic Guides n. 34. Oxford: Oxford University Press.

WHITTLE, Bruno. 2015a. "On Infinite Size." in *Oxford Studies in Metaphysics*, volume IX, edited by Karen BENNETT and Dean W. ZIMMERMAN, pp. 3–19. New York: Oxford University Press, doi:10.1093/acprof:oso/9780198729242.003.0001.

—. 2015b. "Reply to McGee (2015)." in *Oxford Studies in Metaphysics*, volume IX, edited by Karen BENNETT and Dean W. ZIMMERMAN, pp. 33–41. New York: Oxford University Press, doi:10.1093/acprof:oso/9780198729242.001.0001.

—. 2018. "Size and Function." *Erkenntnis* 83(4): 853–873, doi:10.1007/s10670-017-9917-z.

* For helpful comments and discussions I thank Balthasar Grabmayr, David Kashtan, Aviv Keren, Eli Pitcovski, Gil Sagi, and four anonymous referees. The research leading to this paper was supported by post-doc fellowships at the University of Haifa and at the Hebrew University of Jerusalem.

Functionalism, Pluralities, and Groups

Emilie Pagano

It's widely accepted that *pluralism* about groups—the view that groups are pluralities—is incompatible with the following: one group can have different individuals as members at both different times and in different worlds (Difference), and more than one group can have the same individuals as members at both the same times and in the same worlds (Sameness). As a result, it's widely accepted that pluralism is false. In this paper, I argue that these "arguments from Difference and Sameness" are unsound. First, I articulate a functionalist account of what it is to be a group that's neutral with respect to pluralism and its primary opponent, *monism*. According to the version of functionalist pluralism I propose, groups are pluralities of functional roles. Second, I argue that because different individuals can play a role at both different times and in different worlds, and because the same individuals can play different roles at both the same times and in the same worlds, functionalist pluralism is invulnerable to the arguments from Difference and Sameness. Lastly, I raise a challenge for functionalist monism: whereas functionalism seems to favor "external" individuation conditions, monism seems to favor "internal" individuation conditions, and it's up to the functionalist monist to square them. In the process, I hope to have shown that functionalism—whether pluralistic or monistic—is worthy of our attention.

Groups are everywhere. We rely on them when we marry, matriculate, and mortgage, when we pray, purchase, and patronize, and when we lend, loot, and lecture. They systematically guide our interactions. They have members, they do things, and they come and go. They matter. As a result, the question arises: What are they?

As with "What is X?" questions generally, this one's ambiguous. On the one hand, to ask what a group is is to ask what it is *to be* a group. That's a question about what defines the kind GROUP. It's asking: What is it for a

particular collection of individuals to be a group rather than a mere collection of individuals? Call it *the definitional question*.[1]

On the other hand, to ask what a group is is to ask what kinds of *things* groups are. That isn't a question about what defines the things we call groups qua groups. It's asking: What, say, instantiates GROUP? Groups are the kinds of things that fit into churches, corridors, and courtrooms. And it might be—indeed, it's quite plausible—that GROUP doesn't fit into churches, corridors, or courtrooms. As a result, we'll want to know what does. Call this *the ontological question*.[2,3]

Interestingly, philosophers have nearly universally privileged the ontological question. Predictably, there's significant disagreement among them. Nonetheless, the vast majority of philosophers accept that groups are "one," that they're fundamentally singular things (e.g., Effingham 2010; Ritchie 2013, 2015, 2020; Hawley 2017; and Fine 2020).[4] I call their view *monism*. Nonetheless, a minority of philosophers accept that groups are "many," that they're fundamentally plural "things" (i.e., pluralities) (Uzquiano 2018; Faller 2019;

1 I have "real" rather than linguistic definition in mind. See Dorr (2016), Correia (2017), and Rosen et al. (2018) for discussion.

2 Should one be disinclined to accept this distinction—in particular, because one is inclined to accept that GROUP is defined by the kinds of things groups are, or conversely—here's an argument:

(i) Groups exist.
(ii) It's possible that nothing defines GROUP.
(iii) If it's possible that nothing defines group, GROUP isn't defined by the kinds of things groups are; that is, the kinds of things that exist.
Therefore, GROUP isn't defined by the kinds of things groups are (i.e., the kinds of things that exist).

Of course, one might deny (ii). But the point is: whether it's possible that GROUP is indefinable is an open question, and, so, one can't deny (ii) without argument.

3 Though the distinction between the definitional and ontological questions is familiar, my conception resembles a distinction metaethicists draw between value and its "bearers." Again, we might ask what it is *to be* good, what it is that defines GOOD. That's a definitional question. However, we might ask what kinds of things are good, what, in other words, "bears goodness." That's an ontological question. (See **?** for an excellent discussion.) Clearly, it's possible that GOOD is indefinable; a whole slew of philosophers accept it. Nonetheless, things are good, things bear value. As a result, GOOD isn't defined by its bearers; that is, by the kinds of things that are good.

4 It's important to note that Epstein (2015, 2018) and Thomasson (2019) accept "hybridism" about groups, according to which some groups are one and some groups are many. In what follows, I set hybridism aside, in particular, because problems for monism *are* problems for hybridism. If groups aren't one, they're not one *and* many. Rather, they're many.

Horden and López de Sa 2021; and Wilhelm 2022). They accept that a group *is* in some sense its members, that they *are* "it."[5] I call their view *pluralism*.

In general, monists accept the following argument against pluralism:

(1) A group can have different members at both different times and in different worlds. (Difference)
(2) If groups are pluralities, groups can't have different members at either different times or in different worlds.
Therefore, groups aren't pluralities.
(3) According to pluralism, groups are pluralities.
Therefore, pluralism is false.

Moreover, they accept:

(4) Different groups can have the same members at both the same times and in the same worlds. (Sameness)
(5) If groups are pluralities, different groups can't have the same members at either the same times or in the same worlds.
Therefore, groups aren't pluralities.
(6) According to pluralism, groups are pluralities.
Therefore, pluralism is false.

I call these "the arguments from Difference and Sameness." In this paper, I argue that they're unsound. Both (2) and (5) are false.

Admittedly, my argument takes some—ultimately necessary—twists and turns. Here's how it'll go. In § 1, I articulate a functionalist account of groups as an answer to the definitional question. In § 2, I argue that functionalism is neutral with respect to both monism and pluralism. In § 3, I argue that by obscuring the definitional question, the arguments from Difference and Sameness assume a particularly naïve version of pluralism, and show that an attractive version of functionalist pluralism is invulnerable to them. Lastly, in § 4, I raise a challenge for functionalist monism: whereas functionalism seems

5 Of course, that's an odd way of speaking. However, I have roughly what Baxter (1988) has in mind when he suggests that "a" group—a *loosely* singular entity—is identical to its members—a *strictly* plural entity (i.e., a plurality). In other words, though we might *speak* of groups as if they're strictly one—we tend to say that *it* nominated me rather than that *they've* nominated me, for instance—that we do follows from, say, our willingness to *count* the department as something over and above its members; that is, from strictly *practical* considerations. And that's quite compatible with its being strictly many.

doi: 10.48106/dial.v77.i3.03

to favor "external" individuation conditions, monism seems to favor "internal" individuation conditions, and it's up to the functionalist monist to square them. I don't claim that my challenge is dispositive, however. Rather, I claim that it exposes an important source of disagreement that's worth pursuing.

Although the implications for monism and pluralism are clear, one of my aims is to generate interest in functionalism, whether monistic or pluralistic. Although functionalism about social goings-on—specifically, about artifacts—has pedigree, its application to groups hasn't been explored.[6] There are details to sort out, of course. And though I'll make suggestions as I go, ultimately, I hope to have provided a framework for sorting them out that's worthy of our attention.

1 The Definitional Question: Functionalism

As an answer to the definitional question, I propose a functionalist account of groups. Functionalism about social goings-on is an established view; in particular, functionalism about artifacts (Searle 1995, 2010; Baker 2007; Thomasson 2019; Evnine 2016; Guala 2016; Koslicki 2018). However, it hasn't been pursued as an account of what groups are. But as Haslanger (1995) suggests, we might think of groups as special kinds of artifacts, as products of some of the things we do, whether intentionally or unintentionally. We might think of them as things we in some sense "use" to do them. And this suggests we might expand functionalism to account for them. In this section, I give it a try.

First, a note. Generally, philosophers think of artifactual functions teleologically, as things that serve purposes artificers impose on their products. However, I take my cue from functionalism about mental states, according to which functions aren't teleological but, e.g., causal. Ultimately, that's a choice point. One might translate the account I propose in terms of teleological functions (see Thomasson 2019 for rumblings). As a result, it's easily assimilated into the wider literature.

6 Sometimes Ritchie (2013)'s account seems like a version of functionalism about groups, as when she claims that "[a group] structure is realized when each of its functionally defined nodes or places are occupied" (Ritchie 2013, 257). Nonetheless, she doesn't pursue functionalism about groups per se.

1.1 *Functionalism*

In general, functionalism about *x* is the view that *x* is a functional kind. A kind, K, is a functional kind when something is an K because of its extrinsic rather than its intrinsic properties; specifically, what it does—the way it functions—within a particular system in which it's embedded. Ordinarily, functionalism is associated with a particular account of mental states.[7] Functionalists about mental states accept that kinds of mental states (e.g., PAIN) are defined by functions (e.g., to avoid physical harm) that are realized[8] by whatever plays the relevant roles (e.g., the "pain role") within a particular cognitive system. Generally, they accept that PAIN's realizations are pains.

Functionalism's claim to fame is the ease with which it accommodates multiple realizability. Multiply realizability is a feature of mental states whereby a single mental state might be grounded in multiple non-mental (e.g., physical) states. In creatures with cognitive systems like ours, PAIN is realized by c-fibers that cause the relevant kinds of responses; flinches, winces, and wails, for instance. In extraterrestrials with cognitive systems unlike ours, however, PAIN might be realized by gunky she-fibers that cause relevantly similar responses; shlinches, shinces, and shails, for instance. But both are pains because pains *are* what function that way.

Although there are many details about which functionalists disagree, I'll assume that PAIN is a higher-order property of the form *an-input-linking-an-output*, where the relevant kind of linking is causal.[9] A particular pain might be a-pinch-causing-a-flinch within a particular cognitive system or a shinch-causing-a-slinch in theirs. But that's merely for convenience. For the important thing is this: to be a pain is to do what pains do. As a result—and with relevant disagreements aside—I propose

> GROUP FUNCTIONALISM (FUNCTIONALISM). For *xx* to be a group, is for

7 Of course, there are many varieties of functionalism. However, the variety of functionalism I've described is "commonsense," represented in, e.g., Block (1978), Fodor (1968), Lewis (1980), and Putnam (1975).

8 Realization is a kind of non-causal determination that's figured in influential arguments for functionalism. See Bickle (2020) for an excellent survey.

9 Regrettably, I can't discuss the many—and radically complex—differences in functionalist accounts of mental properties here. But, ultimately, the relevant differences won't jeopardize the view I propose. In fact, they're bound to enrich it.

doi: 10.48106/dial.v77.i3.03

i. *xx* to be an instance of a group kind, K, and for
ii. *xx* to serve a function that defines K,
iii. within a particular social system

where *xx* is either a singular or a non-singular plurality. Here, I argue that FUNCTIONALISM provides an attractive answer to the definitional question because it does what an account of groups should do; it satisfies several desiderata.

First, a clarification. One might worry that in appealing to group kinds, FUNCTIONALISM is circular. Group kinds are *group* kinds, and one might reasonably insist that GROUP can't be defined by them.[10] (See **?** for the corresponding objection about mental states.) Here, then, is a conception of group kinds I'll assume throughout:

GROUP KIND FUNCTIONALISM. K is a group kind iff K is

iv. a type of collection of individuals, C, such that
v. the individuals comprising C realize roles, *rr*, that
vi. give them reasons to act as members of C

where, again, *rr* is either a singular or non-singular plurality. Clearly, FUNCTIONALISM plus GROUP KIND FUNCTIONALISM isn't circular. GROUP KIND FUNCTIONALISM says nothing about GROUP. Rather, it says that a particular collection of individuals—Jonathan, Jennifer, and Julia, say—is, e.g., a department of philosophy because they realize roles that push them around in particular ways—in ways characteristic of departments of philosophy. In other words, their being a group isn't what makes them a department. Rather, their being a department is what makes them a group. Of course, GROUP KIND FUNCTIONALISM assumes we understand what roles are. But because everyone owes us an account of them, that's okay. As a result, I'll carry on as planned.[11]

10 For instance, one might insist that particular group kinds are defined by GROUP as particular pains might be defined by PAIN. But that's in tension with multiple realizability. For if this or that mental state is essentially a pain, anything that's in that mental state has to be in pain. But that's precisely what the functionalist denies. Because mental states are multiply realizable, something might be in that mental state and not be in pain because of the nature of its cognitive system more generally. That's why pains are defined by their extrinsic rather than their intrinsic features.

11 Alternatively, one might use "group kind" ostensively as a placeholder for whatever things are relevantly like departments, courts, and baseball teams. Again, I take my cue from functionalism about mental states, and it's plausible that we can say that MENTAL STATE is defined by "mental

1.2 *Desiderata*

FUNCTIONALISM satisfies several desiderata. In particular, it accounts for

(a) the distinction between groups and mere collections of individuals;
(b) the fact that groups and the individuals that are their members might malfunction—alternatively, might err[12]—as the kinds of things they are; and
(c) the fact that groups are embedded in wider social systems.

That it does is reason to take it seriously.

As for (a): again, FUNCTIONALISM is the view that groups are what they are because they do what they do.[13] A paradigmatic group like the Department of Philosophy is the group it is because the individuals that are its members in some way "do philosophy" within a particular college, within a particular university, and as part of a particular department of education.[14] Whatever their doing what they do amounts to, the Department of Philosophy is the particular group it is because the individuals that are its members do what they do. It's a department of philosophy, an instance of DEPARTMENT OF PHILOSOPHY, and, so, it's a group.

---

state kinds" without circularity. We know which responses are pain responses, and so we can appeal to them, whatever they are, in defining PAIN. (See Shoemaker 2003; McCullagh 2000; Tooley 2001 for this kind of response.) Similarly, we know which collections of individuals are, e.g., baseball teams, and so we can appeal to them, whatever they are, in defining BASEBALL TEAM, too.

12 The difference between malfunctioning and erring corresponds to the difference between teleological and nonteleological (e.g., causal) functions. Again, I'll set that aside and mark it as a choice point.

13 We might worry about understanding groups in terms of their "doing things." For there's a question about how to understand the relation between groups and what they do, such that groups don't disappear when their members *don't* do it. For instance, it's an open question whether departments of philosophy exist when their members don't do philosophy because, e.g., they're on strike, funding is suspended, it's temporarily out of members. But this isn't a special problem for FUNCTIONALISM. Whether groups are defined in terms of what they do, everyone is on the hook for providing an account of how they persist through changes like these, if they do.

14 There's a question about how to define the relevant group kinds when some of the department's members—its secretaries, say—*don't* do philosophy in any straightforward sense. There's a relatively straightforward solution to this problem, however. There's a slew of ways to play the roles that define particular group kinds. For instance, although secretaries might not give seminars, what they do *enables* professors to give seminars. As a result, defining group kinds in terms of relatively course-grained functions like this one isn't in itself problematic.

doi: 10.48106/dial.v77.i3.03

And this feature of FUNCTIONALISM underwrites a plausible explanation of the distinction between groups and mere collections of individuals. It's uncontroversial that groups, in some sense, consist of collections of individuals. The Department of Philosophy, in some sense, consists of the individuals that are its members, of Jonathan, Jennifer, and Julia. Similarly, the Supreme Court, in some sense, consists of the individuals that are its members, of Roberts and his colleagues, and the Boston Red Sox of the individuals that are its members, of Martinez and his teammates.

But the Department of Philosophy, the Supreme Court, and the Red Sox are relevantly unlike collections of individuals like those wearing white t-shirts, those driving rental cars, and those that prefer chocolate to vanilla ice cream. Intuitively, whereas groups like the Department of Philosophy, the Supreme Court, and the Red Sox are such that the individuals that are their members do what they do *because they're members*, there's nothing the collection of individuals wearing white t-shirts do because they're "members." Of course, there's something they do: they wear white shirts. Unlike the groups in question, however, they're its members *because they're wearing white shirts*. In other words, whereas the individuals that are a group's members play particular kinds of roles—and so can act in their capacity as members—mere collections of individuals don't and can't. And, again, that's what FUNCTIONALISM implies: because groups are individuated by what they do, the Department of Philosophy is a group because the individuals that are its members play the roles that define it, and not conversely.

As for (b): FUNCTIONALISM explains how and why groups and the individuals that are their members might malfunction as the kinds of things they are. Groups are like thermostats. Thermostats are what measure temperature. There's something they do and with respect to which they might fail. If they do, they're bad thermostats. Similarly, departments of philosophy are what do philosophy in the relevant ways. Like thermostats, there's something they do with respect to which they might fail. If they do, they're bad departments of philosophy.[15]

As for (c): FUNCTIONALISM explains how and why particular groups are embedded in wider social systems. Again, particular pains are what protect the

15 When we say the Department of Philosophy does philosophy badly, then we mean that it's failed *as* a department of philosophy. Though it's failing to be, e.g., collegial is grounds for criticism, this isn't a failure qua department of philosophy. On the contrary, its failing to do philosophy collegially is a kind of normative failure. But its normative failures don't entail that it fails to be a department of philosophy altogether.

pained, are causing's to avoid hot stoves, stubbed toes, and unfriendly blows. This entails that they function within wider cognitive systems that enable the relevant kinds of behaviors; in other words, that there are pained things—organisms or otherwise—to protect. They're parts of cognitive systems, are what function to avoid the physical harms that might befall the things they cause to behave in the relevant ways. In a word: no things, no cognitive systems, no pains.

Analogously, particular departments of philosophy are what do philosophy in the relevant ways, what in some sense account for the fact that the individuals that are their members give particular kinds of seminars, invite particular kinds of speakers, and host particular kinds of events. Again, this entails that particular departments of philosophy function within wider social systems that enable these kinds of behaviors. Again, the Department of Philosophy is the group it is because its members do philosophy as part of a particular college that's part of a particular university that's part of a particular department of education, and these ultimately underwrite, say, its seminar offerings. These groups require that the Department follow a particular curriculum and, so, ultimately constrain how the Department of Philosophy does philosophy.

Relatedly, FUNCTIONALISM explains how and why the social systems in which particular groups are embedded are *structured*. What social structures are is an important and underexplored issue in contemporary philosophy. But a few things are clear.

Minimally, structures are arrangements. They're complexes of relations. They consist of "positions" at the intersections of these relations, and things "occupy" them. Naturally, then, group structures are social arrangements; complexes of social relations that consist of intersecting positions things occupy. Baseball teams are structured, for instance. Every baseball team has a pitcher and a catcher. In other words, a baseball team's structure partly consists of particular asymmetric, irreflexive, and non-transitive relations—*pitches to* and *returns to*, say—whose relata intersect in a particular way: pitchers pitch to catchers and catchers return to pitchers. And that the relevant elements—*this* and *that* individual—occupy the relevant positions—that they're arranged in the relevant way—partly grounds the fact that they're part of a baseball team rather than a mere heap of individuals.

More than this, though, that a particular group functions in the particular ways it does is explained by the general arrangements of its elements. For instance, the Red Sox and the individuals that are their members play roles

doi: 10.48106/dial.v77.i3.03

that are determined by the wider social system in which they're embedded.[16] Like the Department of Philosophy and the Supreme Court, in other words, the Red Sox are part of other groups; for instance, the American League East (ALE) and Major League Baseball (MLB).[17] And this wider social system partly determines in which relations the individuals that are members of the Red Sox stand both to one another and to other groups. For instance, we can't explain the fact that the Red Sox won the World Series in 2018 by appealing to how well they played. Rather, we must appeal to the relational fact that they played *better* than the Los Angeles Dodgers—themselves part of the National League West (NLW) and, so, the MLB—and to the rules that legitimated their win. In other words, we must appeal to the social system in which both the Red Sox and the Los Angeles Dodgers are embedded—the MLB—to explain important facts about them. As a result, it's not merely that particular groups are embedded in wider social systems, the social systems in which they're embedded structure them. And, again, FUNCTIONALISM bears this out. (See Haslanger 2000 for a similar thought.)

As a result, FUNCTIONALISM satisfies desiderata (a)–(c) and thereby provides a powerful framework for thinking both about what it is to be a group per se and what it is to be a group of a particular kind. But because it's a sketch of an account, there are details about which we might meaningfully disagree. I'll consider a particularly important detail about which we might disagree in § 3. But there's more.

1.3 *Open Questions*

Here's a brief survey. We might disagree about what a social system is and about what it is for a complex of relations to constitute a structure. Both Ritchie (2013, 2015, 2020) and Haslanger (2016) understand both in terms of Shapiro (1997)'s influential conception of mathematical structure according to which

16 Importantly, FUNCTIONALISM implies that, at any time and in any world, there's a social system that is structured by social relations that *aren't* embedded in any larger social structure. Maximal social systems—like "maximal" cognitive systems—can't be realized more than once at any time or in any world aside from trivial permutations of the same roles within a single group. As a result, every non-maximal social system there is part of a larger social system whose elements are ultimately structured by the most basic social relations there are.

17 I'm assuming that the MLB is itself a group with other groups as "parts." Following Ruben (1985), I take it that being a member of a group and being part of a group are different relations. In what follows, however, I overlook this distinction.

> a system [is] a collection of [entities] with certain relations. […] A structure is the abstract form of a system, highlighting the interrelationships among the [entities], and ignoring any features of them that do not affect how they relate to other [entities] in the system. (Shapiro 1997, 73–74)

However, we needn't accept Shapiro's conception of structure to accommodate this feature of FUNCTIONALISM. What's important is that we accept that the social systems in which groups are embedded are holistic. And, again, FUNCTIONALISM bears this out.

Moreover, we might disagree about whether particular individuals are unified in coming to be elements in a functionalist structure or whether it's sufficient that the relevant network of social relations is interdependent. (Indeed, I discuss its importance in § 4.) Relatedly, we might disagree about what the relevant functions are and what it is to realize them. For instance, there might be "basic" kinds of groups that correspond to basic kinds of functions—for instance, to competition—and non-basic kinds of groups that correspond to ways individuals might compete—for instance, to playing baseball.

Lastly, we might disagree about which types of collections of individuals are groups—in particular, whether genders and races are.[18] Functionalist accounts of genders and races are well-represented. For instance, MacKinnon (1996)'s remarkably influential account of gender—according to which for one to be a woman is for one to be sexually subordinated to men and to be a man to sexually subordinate women—is recognizably functionalist. (See Jenkins 2017 for a compelling case for this functionalist interpretation of MacKinnon; see also Young 1990; Witt 2011; and Haslanger 2012 for additional evidence.) Similarly, Charles Mills (1997)'s account of race—according to which to be, e.g., black is to be positioned within a social system (i.e., a "vertical race system") such that one is treated as a "sub-person"—is recognizably functionalist, too. (See Griffith 2020 for discussion.) But the important point is this: FUNCTIONALISM paves the way for a unification of otherwise disparate literatures about the metaphysics of groups generally.

18 Another thing we might disagree about is whether genders and races are groups at all. Ritchie (2020) claims that they're "group-like kinds," though one wonders whether they might rather be "kind-like groups." Nonetheless, I won't defend the view that they are. Ultimately, it doesn't matter for the question whether monism or pluralism is correct, though I think it's worthy of our attention.

doi: 10.48106/dial.v77.i3.03

Although FUNCTIONALISM is ultimately neutral with respect to these disagreements, it both clarifies what's at issue and provides a framework for thinking about how they might be resolved. (My many footnotes attest to it!) To the extent that they're meaningful disagreements, then we've reason to take it seriously.[19]

2 The Ontological Question: Monism & Pluralism

Though FUNCTIONALISM allows for significant disagreements among functionalists, there's an important detail that's central to what I've called the arguments from Difference and Sameness, namely, whether groups are one or many. In this section, I argue that FUNCTIONALISM is compatible both with what I've called monistic and pluralistic answers to the ontological question. Ultimately, in § 3, I show that the arguments from Difference and Sameness are unsound because of it.

Again, monism is the view that groups are singular things. Some monists accept that groups are sets (Sider 2001; Effingham 2010), some that they're "realizations of structure" (Ritchie 2013, 2015, 2020), some that they're fusions (Hawley 2017), and some that they're "embodiments" of structure (Fine 2020). Sets, realizations of structure, fusions, and embodiments of structure are "something over and above" the collections of individuals out of which they're made up, and, in each case, that something is one.[20]

And though pluralism is widely rejected, some have come to its defense. In particular, some accept that groups are pluralities of individuals that embody "plural conditions" (Uzquiano 2018), and others that they're "instantaneous pluralities" (Wilhelm 2020). In each case, however, groups are many in the sense that they're fundamentally plural "things" (i.e., pluralities), "the very kind of ['object'] of which many is to be asserted," as Russell (1903) suggests.

But the important point is this: one can be a functionalist, whether one is a monist or a pluralist. In this section, I focus on Kit Fine and Gabriel Uzquiano's monistic and pluralistic accounts of groups, respectively. Because each appeals to Fine (1999)'s "theory of embodiment," focusing on theirs tidies things up. Though other monists and pluralists will answer the ontological

19 Griffith (2020) takes FUNCTIONALISM, as I understand it, seriously.

20 It might be that a plurality is something over and above *another* plurality, as Uzquiano (2018) accepts with respect to groups and as Dasgupta (2014) suggests more generally with respect to the relation (i.e., grounding) between groups and their members. As a result, this is an unargued for assumption on the part of monists, though it's perfectly commonplace.

question in meaningfully different ways, what I say in restricting myself to embodiments is ultimately compatible with them.

First, the framework. According to Fine (2020), groups are embodiments. Embodiments are quite like Aristotelean compounds.[21] Generally, Aristotelean compounds consist not only of "matter"—say, of a plurality of flowers—but of "form"—say, of a "being bunched" arrangement—where form is understood to structure matter, to turn a plurality of flowers into a bouquet. Similarly, embodiments consist not only of their parts but of "principles of embodiment" (henceforth: principle/s) that specify the relevant arrangements that structure their parts. The bunch is a plurality of flowers that embody the arrangement the relevant principle specifies; the bunch is the flowers-qua-bunch: a bouquet.[22]

Fine distinguishes two kinds of embodiments, what he calls "rigid" and "variable" embodiments. On the one hand, rigid embodiments have their parts necessarily. For instance, the bouquet is a rigid embodiment because exactly the relevant flowers embody exactly the relevant bunching arrangement at all times and in all worlds. The bouquet is the bouquet it is because it has exactly those flowers arranged in exactly that way; replace one and you'll have another bouquet.

On the other hand, variable embodiments have their parts contingently. For instance, my bike is a variable embodiment because it has different "manifestations" that embody a particular arrangement at particular times and in particular worlds. Clearly, *my* bike has different bits of rubber, plastic, and metal as parts, and these are arranged 'bicycley' at different times and in different worlds. In other words, my bike isn't identical to its manifestations, to the objects consisting of bits of rubber, plastic, and metal arranged 'bicycley' at particular times and in particular words. Rather, it's constituted by them then and there. As a result, it persists when I replace a tire. (We might disagree about examples, of course, but the distinction is clear enough.)

Ultimately, groups can be either rigid or variable embodiments, though Fine (2020) focuses on those that are variable embodiments. The Department

21 See Koons (2014) and Marmodoro (2013) for misgivings.

22 Importantly, embodiment isn't mere instantiation. Should the flowers merely instantiate the property of being bunched, it follows that the flowers are merely contingently bunched. When one tosses the bunch and the flowers scatter, the bunch—the *flowers*—survive. If the bunch of flowers embody being bunched, however—if they're the flowers-qua-bunch, the bouquet—they're necessarily bunched, and, so, the bouquet doesn't survive the scattering. But this feature of Fine's account won't matter for what I go on to say.

doi: 10.48106/dial.v77.i3.03

is the group it is because it has manifestations at particular times and in particular worlds consisting of the individuals that are its members and the arrangements they embody at those times and in those worlds. What this arrangement is is specified by the relevant principle; something like 'being arranged "departmentally".' And in embodying that arrangement, the Department isn't identical to its manifestations but constituted by them at particular times and in particular worlds.[23]

But Fine's theory is remarkably flexible. He places no restrictions on the content of principles of embodiment generally. It's up to say what they are. Uzquiano (2018) exploits this flexibility and argues that groups are structured by principles that encode *plural* rather than singular conditions, as Fine assumes. In particular, he argues that their principles are relevantly like "being scattered," where to be scattered implies that what's scattered *isn't* one but many.

Moreover, he argues that because his account of groups can accommodate what might be groups that others can't—in particular, supposed groups like queues and multitudes that are significantly less structured than, e.g., departments of philosophy—we should prefer it to Fine's. As he puts it, "neither queues nor multitudes appear to demand much of their individual members." They must "[embody] a certain spatial arrangement but they do not seem to require a shared intentionality or agency from their members" (Uzquiano 2018, 423). In other words, though individuals that are "members" of queues embody minimal arrangements such that they're queues, this doesn't entail that they're one.[24]

However, neither Fine nor Uzquiano say what *kinds* of principles define groups rather than other variable embodiments. In other words, neither gives us an account of what it is to be a group—what defines GROUP—such that we can distinguish embodiments that are groups from those that aren't. Ultimately, that's a desirable feature of the theory of embodiment. It was developed as an alternative to traditional accounts of mereological composition. It tells us what kinds of *things* groups are, not what it is to be a group *per se*. As a

23 Fine claims that this account can accommodate the view that groups can have different structures at different times and in different worlds. It's not clear that he's right, however. Whether he is depends on the content of the relevant principles of embodiment.

24 One wonders whether this is compatible with pluralism. It's tempting to say that embodiment itself is unifying and, so, that even queues become one in embodying the relevant spatial arrangement, plural conditions be damned.

result, it doesn't presuppose an answer to the definitional question. And that's good.

But now we can see how both Fine and Uzquiano might answer it. For we can trace the differences between bikes and departments of philosophy to the principles they embody. Whereas my bike's parts might embody one functional arrangement—a principle that specifies what the relevant bits of rubber and metal do (e.g., enable riding)—a department of philosophy will embody another functional arrangement—concerning what *it* does (i.e., philosophy). As a result, Fineans can appeal to FUNCTIONALISM to explain what it is to be a group per se. In particular, it can explain which embodiments are groups and which aren't by appealing to the distinctive kinds of principles the individuals that are their members embody. Because Fine and Uzquiano can agree that groups are functional, then FUNCTIONALISM is compatible with both monism and pluralism.

3 The Arguments from Difference & Sameness

Now we're ready for the biggest bit: the arguments from Difference and Sameness. (Here[25] they are for easy access.) Again, they're supposed to be problems for a pluralistic account of the kinds of things groups are and *not* for a functionalist account of GROUP. They're supposed to imperil the pluralist's answer to the ontological question. In this section, I argue that an attractive version of

25 For reference:

(1) A group can have different members both at different times and in different worlds.
(2) If groups are pluralities, a group can't have different members either at different times or in different worlds.
Therefore, groups aren't pluralities.
(3) According to pluralism, groups are pluralities.
Therefore, pluralism is false.

And:

(4) Different groups can have the same members both at the same times and in the same worlds.
(5) If groups are pluralities, different groups can't have the members either at either the same times or in the same worlds.
Therefore, groups aren't pluralities.
(6) According to pluralism, groups are pluralities.
Therefore, pluralism is false.

doi: 10.48106/dial.v77.i3.03

functionalist pluralism is invulnerable to them. In particular, I argue that the arguments from Difference and Sameness obscure a distinction between relations I call "being grouped" and "being a member." Once we distinguish them, FUNCTIONALISM comes to the rescue, and pluralism is back in business.

3.1 *The Arguments*

Again, that paradigmatic groups can have different members both at different times and in different worlds is widely accepted as a constraint on accounts of what they are. I call it:

> THE DIFFERENCE CONSTRAINT (DIFFERENCE). A group can have different members both at different times and in different worlds.

It's plausible that departments of philosophy can, e.g., hire professors. In particular, it's plausible that the department once had members it no longer does and that it might have had members it never did. Nonetheless, it's precisely the department it either was or might have been. If so, DIFFERENCE is true.

Moreover, that different groups can have the same members both at the same times and in the same worlds is widely accepted as a constraint on accounts of what they are. I call it:

> THE SAMENESS CONSTRAINT (SAMENESS). Different groups can have the same members both at the same times and in the same worlds.

The department and its bowlers, the HaeXeities, might have exactly the same members. Nonetheless, it's plausible that the department isn't identical to the HaeXeities. For instance, whereas members of the department are expected to behave decorously in departmental dealings—and, so, might be sanctioned for misbehaving—the HaeXeities aren't; anything goes on the lanes. If so, SAMENESS is true.

But because it's widely accepted that pluralities have their "members" essentially, it's widely accepted that pluralities can't have different members at either different times or in different worlds. If the plurality of individuals that are members of the department here and now—Jonathan, Jennifer, and Julia—consists of them essentially, it consists of them necessarily and, so, at every time and in every world.

As a result, both DIFFERENCE and SAMENESS presuppose a principle of extensionality for pluralities:

> EXTENSIONALITY. One plurality, xx, is identical to another plurality, yy, if and only if for all z, z is one of the xx's if and only if z is one of the yy's.

Assuming EXTENSIONALITY, pluralism entails that for groups to be different *is* for them to have different members, and for them to be the same *is* for them to have the same members. Because Jonathan, Jennifer, and Julia couldn't be Jonathan, Jennifer, and Julia, *and Jim*, pluralism entails that each is a different plurality and, so, that the department can't have different members at different times and in different worlds. Hence the argument from DIFFERENCE. Similarly, if groups are pluralities, the department and the HaeXeities aren't different groups because they have exactly the same members: Jonathan, Jennifer, and Julia. Hence the argument from SAMENESS.

The arguments from DIFFERENCE and SAMENESS have become *the* arguments against pluralism in the literature, and, so, they're rarely resisted.[26] Nonetheless, the arguments are misleading. In keeping with EXTENSIONALITY, it's important to emphasize that (2) and (5) are true if and only if pluralism is the view that

(i) to be a group, G, is to be identical to a plurality, *ab*; and
(ii) to be a member of G is to be a "member" of *ab* (i.e., to be either *a* or *b*).

Again, EXTENSIONALITY entails that the department *is* its members, that they *are* the department, because the relevant individuals are "members" of the plurality with which it's identical. Similarly, the HaeXeities *is* its members, they *are* the HaeXeities, because the relevant individuals are "members" of the plurality with which it's identical.

However, there's no one thing it is to be either one or many, and, so, there are different versions of pluralism to which the arguments are inattentive.[27]

26 In particular, though Horden and López de Sa (2021) are pluralists, they reject SAMENESS, which neither Uzquiano (2018) nor I are willing to do.

27 The monist will likely insist that there's no conception of plurality on which pluralities don't have their members essentially. And, if they're right, EXTENSIONALITY entails that pluralism is false. I both agree and disagree. On the one hand, I agree that the pluralist ought to accept EXTENSIONALITY. However, I disagree that accepting EXTENSIONALITY entails that there's no conception of pluralities on which they have their members essentially. That's the crux of my argument.

doi: 10.48106/dial.v77.i3.03

As a result, neither (i) nor (ii) is entailed by pluralism per se. Rather, they constitute a—particularly naïve—version of pluralism that both monists and pluralists are right to resist. Because both (i) and (ii) are required to substantiate the arguments from DIFFERENCE and SAMENESS, then they misrepresent pluralism.

The basic idea is this. The arguments from DIFFERENCE and SAMENESS assume that EXTENSIONALITY entails that pluralism per se is false. However, there's an intuitive version of *functionalist* pluralism that's compatible with EXTENSIONALITY. If that's right, (2) and (5) are false, and the arguments from DIFFERENCE and SAMENESS are unsound. That's what I argue in this section.

3.2 *Being Grouped vs. Being a Member*

I begin with a distinction:

> BEING GROUPED. The relation between a group, G, and the plurality of its members.

And

> BEING A MEMBER. The relation between an individual that is a member of a group, G, and G.

Minimally, BEING GROUPED is a multigrade relation. Although it's a relation between G and its members, its members' slot doesn't have a definite degree: at some times and in some worlds, some number of members stand in this relation, and at other times and in other worlds, another number of members do. However, BEING A MEMBER is a unigrade relation; its members' slot does have a definite degree. In particular, it's a binary relation in which a group stands to a particular individual.[28]

That's sufficient to distinguish BEING GROUPED and BEING A MEMBER. They're different relations because they have different properties. But it's especially important to distinguish them because they imply the distinction between (i) and (ii). For the claim that G is identical to *ab* is a claim about the relation between a group and its members. Something makes a particular collection of individuals a group rather than a mere collection of individuals.

28 Whether these relations are symmetric, asymmetric, or, importantly, antisymmetric is something I won't—but that should be—discuss/ed in detail.

Jonathan, Jennifer, and Julia are the department, in other words, because they're related to it in a particular way; they do this and not that. Again, if (i) is true, this relation is identity (i.e., to *ab*).

Moreover, the claim that to be a member of *ab* is to be a "member" of *ab* is a claim about the relation between the individuals that are a group's members and the group of which they're members. Again, something makes a particular individual a member of a particular group. Julia is a member of the department because *she's* related to it in a particular way. And, again, if (ii) is true, this relation is identity (i.e., to either *a* or *b*).

3.3 *Functionalist Pluralism: Roles*

Importantly, to distinguish BEING GROUPED from BEING A MEMBER is to recognize both that they needn't be identity and that neither (i) entails (ii) nor that (ii) entails (i). Here's a version of pluralism that does the work:

ROLES. For G to be a group is for

(A) G to be a plurality of functional roles, *rr*, that are instances[29] of a kind, K, for
(B) K to be defined by *rr* at particular times and in particular worlds,
(C) within a particular social system.

Simply: the collection of individuals we call the Department of Philosophy is a group because the kind of which it's an instance—DEPARTMENT OF PHILOSOPHY—is defined by a set of functional roles the individuals that are its members realize.[30]

29 Note that I'm appealing to a distinction between multiple realizability and multiple instantiability here. Again, I've suggested that functional properties like being a pain or being a baseball team are *second*-order properties that are defined in terms of particular first-order properties that are instantiated by particular individuals at different times and in different worlds. But because the Red Sox can have different instances of the same set of first-order realizers of the kind BASEBALL TEAM, it's multiple instantiability and *not* multiple realizability that explains how groups survive changes in membership. Nonetheless, it's plausible that multiple realizability explains how different baseball teams at the same times and in the same worlds each realize the functional property in different ways.

30 One might insist that it's an instance of finer-grained kind; say, being a department. But that's risky. In particular, there's a risk that in appealing to fine-grained social kinds, we can't accommodate the view that groups can have *different* structures at different times and in different worlds. For it's arbitrary to stop appealing to finer- and finer-grained social kinds in distinguishing these groups. If the Department of Philosophy is structured in the particular ways it is because it's a

doi: 10.48106/dial.v77.i3.03

Importantly, in defining groups in terms of functional roles, ROLES is a structuralist account of groups. As I suggested in § 1, arrangements are essentially relational. There are no roofs without frames, and there are no frames without foundations. Analogously, there are no pitchers without catchers, no catchers without pitchers. The position PITCHER is defined by the pitches to relation and thereby CATCHER, and the position CATCHER is defined by the returns to relation and thereby PITCHER. And that's what ROLES implies. Every group is defined by a plurality of functional roles, each of which is played by particular individuals that are embedded within wider social systems. ROLES is holistic, too.

A quick clarification. It's plausible that a *version* of the ontological question arises for ROLES. Philosophers interested in groups have said remarkably little about the metaphysics of roles and, so, about what it is to play one.[31] Here, then, is another detail about which functionalists might disagree. I'll refer to role-types and role-tokens to simplify matters, but I intend to remain ecumenical with respect to their metaphysics. We can reasonably expect any account of roles to satisfy the corresponding versions of DIFFERENCE and SAMENESS.

3.4 *Responding to the Arguments from Difference and Sameness*

ROLES reveals that (i) doesn't entail (ii). The first of these claims—that G is identical to *ab*—is neutral both with respect to what *a* and *b* are and with respect to what it is to be a member of G. In particular, it tells us that *a* and *b* aren't *individuals* but the roles they play.[32] And, in that case, (ii) doesn't entail (i) either.

department of philosophy, it's as structured in the particular ways it is because it's a department of philosophy at the University of Chicago, because it's *the* Department of Philosophy at the University of Chicago, because it's *the* Department of Philosophy at the University of Chicago *at a particular time and in a particular world*, and on and on. And with each successive refinement, we'll significantly limit the possibility that it'll survive changes in structure. (See footnote 36.)

31 For instance, a functionalist Finean might accept that *roles* are variable embodiments. That is, whereas Fine accepts that groups are variable embodiments, one might accept that groups are pluralities of variable embodiments. In other words, whereas Fine claims that groups are pluralities of individuals that collectively embody a formal principle, it might be that groups are constituted by pluralities of roles that individuals individually embody. And this underwrites an account of roles with which I'm ultimately sympathetic. I won't assume it here, however.

32 The words we use to refer to both groups and their members can—or, as Horden and López de Sa (2021) argue, do—mislead us. For what we might call "member terms," such as "department chair," figure in claims with both *de re* and *de dicto* interpretations.

For if G is identical to a plurality of functional role-types, *rr*, that define the kind in question—and *not* to a plurality of individuals, *ab*, with which they're easily confused—to be a member of G is to be a "member" of *rr*, to be identical to either role-type. To identify a plurality's "members" is to identify that of which it's a plurality; to identify a *group's* members isn't to identify the role-types of which it's a plurality. To claim that role-types are members of groups would significantly strain our—admittedly pre-theoretical—conception of membership. Rather, it's to identify the individuals that *play* the roles of which it's a plurality, that "are" the role-tokens of those role-types in the way that Jonathan "is" the department's chair. As a result, there are grounds for claiming that whereas "membership" is extensional, membership proper is *non*-extensional, and, so, that membership isn't "membership" (contra (ii)).

As a result, the claim that to be a member of G is to be a "member" of *ab* is neutral both with respect to what *a* and *b* are and with respect to what it is for the individuals that are a group's members to be grouped. We can suppose that *a* and *b are* particular individuals rather than the roles they play *and* that to be a member of G is to be a "member" of *ab*. Still, it's compatible with this view that the relation between G and *ab* isn't identity. Again, on FUNCTIONALISM, it's realization. In other words, it's compatible with (ii) that G exists because *rr* does, that G is a plurality of these role-types that's realized by a plurality of individuals that play them at particular times and in particular worlds (contra (i)). And this allows that G itself might be realized by different pluralities at different times and in different worlds. As a result, the functionalist pluralist

On the one hand, "department chair" has a reading on which it refers to the particular individual that happens to play the relevant role at a particular time and in a particular world (e.g., Jonathan). On the other hand, "department chair" has a reading on which it refers to an arbitrary individual; in particular, to the individual—whoever they are—that plays the relevant role at particular times and in particular worlds.

But we've independent reason to interpret claims in which member terms like "department chair" figure *de dicto*. For if "department chair" refers to Jonathan, the claim that Jonathan *is* the department chair is necessarily true; Jonathan is necessarily self-identical. Similarly, if "chief justice" refers to Roberts and "outfielder" to Martinez, the claims that Roberts is the chief justice and Martinez an outfielder are necessarily true. But, of course, that can't be.

This can be avoided by using member terms like "department chair" (e.g., "chief justice" and "outfielder") to refer to role-types and, so, to arbitrary individuals. We can say that Jonathan "is" the department's chair because he plays the role it refers to within a particular social system; in other words, that he "is" a role-token of that role-type. He realizes it at a particular time, in a particular world, and within a particular social system.

doi: 10.48106/dial.v77.i3.03

needn't reject EXTENSIONALITY to reply to the arguments from DIFFERENCE and SAMENESS.[33]

To summarize: ROLES entails that (2) and (5) are false because it distinguishes being grouped (via (A)) from being a member of (via (C)), where BEING A MEMBER is non-extensional (via (B)). Pluralism *is* back in business.[34]

4 Functionalist Monism vs. Functionalist Pluralism: A Challenge

That's interesting enough. Whether the argument that follows is successful, we'll have made some progress: the arguments from DIFFERENCE and SAMENESS presuppose an account of pluralism that we shouldn't accept, and, so, they're unsound. But that does nothing to recommend pluralism. That pluralism is invulnerable to the arguments from DIFFERENCE and SAMENESS is one thing. That we should accept it is another.

In this section, I give it a try. I argue that the fact that groups are partly individuated by the structured social systems in which they're embedded (per FUNCTIONALISM) is an obstacle for functionalist monism. Because monism recommends the view that groups are exclusively individuated by the relations their members realize "internally," its conception of the kinds of things groups

33 This marks a crucial difference between Uzquiano (2018)'s pluralism and ROLES. Whereas he denies EXTENSIONALITY, I don't.

34 Before proceeding, a disclaimer. One might worry that while ROLES isn't vulnerable to the arguments from DIFFERENCE and SAMENESS, it's vulnerable to arguments from nearby metaphysical constraints. In particular, we can distinguish DIFFERENCE and SAMENESS—which are claims about the individuals that are a group's members—from Difference* and Sameness*—which are claims about a group's structure; in this case, about the relevant role-types:

> DIFFERENCE*. Groups can have different structures at different times and in different worlds.
>
> SAMENESS*. Different groups can have the same structures at the same times and in the same worlds.

Provided we accept EXTENSIONALITY, in other words, the arguments from DIFFERENCE* and SAMENESS* immediately arise. Ultimately, I'm inclined to say that functional kinds are hierarchical and, so, that groups can survive important structural changes by realizing "lower-level" kinds at different times and in different worlds. But because consideration of this issue exceeds the scope of this paper, I set it aside. See (**?**), Faller (2019), and Wilhelm (2022) for accounts of persistence that claim to be pluralistic.

are is a liability. To the extent that we favor FUNCTIONALISM, then we've reason to disfavor functionalist monism and to favor functionalist pluralism.

A note. I'm not claiming that the challenge I raise for functionalist monism is dispositive. Functionalist monism is a powerful view of the metaphysics of groups, and it has powerful resources. Rather, I'm claiming that it's a *meaningful* challenge for functionalist monism, both independently and because of its implications for functionalist pluralism. As a result, I offer it to both functionalist monists and functionalist pluralists. It represents a significant point of disagreement among them that's worth exploring.

4.1 *Internal and External Structure*

Let's return to the view that groups are embedded in structured social systems. In a series of influential papers, Ritchie (2013, 2015, 2020) defends an account of groups according to which groups are—*deep breath*—elements-realizing-social-structure. Groups are singular things, and their elements (e.g., their members) are arranged in particular ways.[35] As a result, her account substantiates the arranging/arrangement distinction I introduced in § 1: a group's elements are arranged in *realizing* an arrangement, and arrangements are the social structures they realize. The Red Sox's members realize a particular social structure consisting in part of PITCHER and CATCHER.

Ritchie's is an exceptionally insightful account of groups, innovative and rich with nuance. But one of its innovations is especially important in understanding both groups and the structured social systems in which they're embedded. She distinguishes what she calls "internal" and "external" ways the individuals that are a group's members realize social structure. In particular, she claims that a group is internally structured when precisely its members are arranged in particular ways and externally structured when it and/or its members and other groups and/or *their* members are arranged in particular ways. For instance, Bertrand Russell University is internally structured relative to both the College of Liberal Arts and the Department

35 It's an open question whether things like books, buildings, and buses occupy nodes, too. Certainly, they aren't *members* of groups but, perhaps, parts of them. For instance, Fine (2020) claims that things like buildings, buses, and basins are *parts* of groups. In particular, he proposes that they're "spatial" rather than "temporal" parts of groups rather than that they're members. Of course, one needn't accept that buildings, buses, and basins are parts of groups either. For instance, one might accept that buildings, buses, and basins *enable* the relevant connections among groups and/or their members.

Nonetheless, I'll focus on their members as Ritchie does.

doi: 10.48106/dial.v77.i3.03

of Philosophy, the College of Liberal Arts is externally structured relative to the Bertrand Russell University and internally structured relative to the Department of Philosophy, and the Department of Philosophy is internally structured relative to its members and externally structured relative to both the College of Liberal Arts and Bertrand Russell University.

That's quite plausible. But there are several things to emphasize about Ritchie's account. First, she claims that groups are internally and externally structured *relative* to both their members and other groups and/or their members. The Department of Philosophy is externally structured relative to both the College of Liberal Arts and Bertrand Russell University, yes. But it's externally related relative to the HeXaeities, with which it's co-extensive, too. There are uncountably many social structures particular groups realize and, so, uncountably many ways they're structured relative to one another.

Second, and again, Ritchie develops the internal/external distinction by appealing to Shapiro (1997)'s account of mathematical structure. Shapiro distinguishes "systems" and "structures" and claims that a system is a "collection of [particular elements] with certain relations" (Shapiro 1997, 73), and a structure is "the abstract form of a system, highlighting the interrelationships among the [particular elements]" (Shapiro 1997, 74). Analogously, Ritchie accepts that whereas groups are systems whose elements are arranged in this or that way, a group's structure is the arrangement their elements realize. Groups are internally and externally structured in whatever ways the arrangements they realize specify.

Shapiro represents structures hypergraphically in terms of "nodes"—or positions things occupy—and edges—or relations that link them. Ritchie does, too, claiming that

> the structure of a group can be represented with nodes […] and edges connecting nodes to other nodes. The edges of a structure capture the relations that hold between nodes. Since all members of a group are related to some degree, each node in structure S is connected to every other node in S. (Ritchie 2013, 268)

As a result, she precisifies her internal/external distinction hypergraphically: a group is internally structured when "all the relevant nodes are occupied by its members and every member occupies some node or other" (Ritchie 2020, 409), and externally structured if and only if its elements "occupy only some node/s" of the relevant internal structures and when "other nodes […] are occupied by entities or systems that are not" among them (Ritchie 2020, 410).

Third, the internal/external distinction doesn't tell us how particular groups are individuated. But we can and should ask how they are. In particular, we can and should ask why—that is, in virtue of what[36]—the College of Liberal Arts is externally structured relative to Bertrand Russell University rather than to the ALE and internally structured relative to the Department of Philosophy rather than to the Red Sox in the ways it is. Again, the internal/external distinction doesn't say which of these social structures is privileged relative to the College of Liberal Arts. But, of course, exactly one is: the department of philosophy structure. As a result, we'll want more.

4.2 *A Challenge*

Luckily, FUNCTIONALISM gives us more. Suppose the individuals that are the Red Sox's members are precisely the individuals that are the Department of Philosophy's members. But, of course, they're different groups, and FUNCTIONALISM accounts for the fact that they're different groups by appealing to the functions they serve. It says that whereas the Red Sox function in one way—the baseball team way—the Department of Philosophy functions in another way—the department of philosophy way—whatever these amount to.

As a result, FUNCTIONALISM implies that we can't account for the fact that the Red Sox and the Department of Philosophy are different groups by appealing to the ways the relevant individuals are arranged. We've supposed that the individuals that are members of both groups are arranged in both ways. Nonetheless, the individuals that are the Red Sox's members aren't arranged in being, e.g., teachers, nor are the individuals that are members of the Department of Philosophy's arranged in being pitchers. In other words, although the individuals that are the Red Sox's members are teachers, they don't do philosophy *as* members of a baseball team. Likewise, although the individuals that are the Department of Philosophy's members are pitchers, they don't play baseball *as* members of a department of philosophy. However, we *can* account for the fact that the Red Sox and the Department of Philosophy are different groups by appealing to the different functions they serve.

Here it's especially important to emphasize that, according to FUNCTIONALISM, the Department of Philosophy is the kind of group it is because the collection of individuals that are its members serve a particular function

36 There's a causal "why question" I mean to avoid asking in this context.

doi: 10.48106/dial.v77.i3.03

within a particular social system. The individuals that are members of the Red Sox don't play baseball simpliciter. Rather, they play baseball *in the ALE*. If we can't appeal to the ALE in individuating the Red Sox, we can't distinguish it from either departments of philosophy with the same members or—more importantly—from other baseball teams (e.g., The Dodgers). Similarly, the individuals that are members of the Department of Philosophy don't do philosophy simpliciter. Again, if we can't appeal to the College of Liberal Arts, we can't distinguish it from either baseball teams with the same members or—again, more importantly—from other departments of philosophy (e.g., at David Lewis University). As a result, FUNCTIONALISM implies that a group's external structure plays a distinctive role in individuating them. In particular, it implies that the Red Sox and the Department of Philosophy are different groups because each is embedded in different social systems that thereby structure them.

Now for the important point. The functionalist's emphasis on external structure seems to be in tension with the view that the individuals that are a group's members are singular entities. For if groups are "unified wholes," they have identifiable boundaries that mark them off from one another. In particular, they're marked off by their intrinsic rather than their extrinsic properties. Indeed, that's what intrinsic properties are: properties things have that don't "mention" other things. But because FUNCTIONALISM entails that group kinds are extrinsic to the collections of individuals that realize them, it seems in conflict with the view that they're singular entities that exclude the groups that are external to them.

Although Ritchie isn't my primary target, let's consider an explicitly functionalist version of her account and see whether it has the resources to respond. Surely, her conception of internal/external structure captures the fact that the Department of Philosophy is externally structured relative to the College of Liberal Arts in being bound by its charter. Per Shapirian structuralism, the relevant complex of relations is there, and we abstract it. But as singular entities with identifiable boundaries, it's not clear that she's entitled to the view that particular groups are partly individuated by their external structures. Again, as she suggests, whereas a group is internally structured when "all the relevant nodes are occupied by its members and every member occupies some node or other" (Ritchie 2020, 409), it's externally structured if and only if its elements "occupy *only some node/s*" of the relevant internal structures and when "other nodes […] are occupied by entities or systems *that are not*" *among them* (Ritchie 2020, 410, emphasis added). And it's precisely this that

makes a group's external structure "stand outside" the singular entities—the groups—that are thereby externally related to it. In other words, the view that groups are singular entities seems to imply that they're individuated exclusively by their internal structures. But that's the problem.

There's a good question about how singular entities are individuated, of course.[37] There are certainly accounts of singular entities that don't have this result. (Fine's is one of them, and I'll consider it shortly.) But because Ritchie's account of groups has it that we abstract structures from whatever social systems are there already, it's difficult to see how she might individuate precisely the "right" social systems (i.e., the Department of Philosophy) rather than others (e.g., the Department of Philosophy + the Red Sox) without privileging their internal structures. In other words, it's difficult to see how she isn't committed to the view that a particular group is individuated exclusively by its internal structure when what's there to be extracted is a tangle of relations, both internal and external, only some of which unify the group in question.

Luckily, Ritchie has options, and each is worthy of significant consideration. Again, I don't claim that this challenge to functionalist monism is dispositive, only that it's worth considering. I'll consider one.

She might accept that the Department of Philosophy and the College of Liberal Arts *are* asymmetrically, *internally* related. It's certainly true that they stand in a kind of asymmetric relation. There's a function they serve that entails it, and that's realized when the Department of Philosophy is bound by its charter. But that seems to require that the College of Liberal Arts and Department of Philosophy *aren't* different groups. In particular, it seems to require that the realization of structure to which the Department of Philosophy corresponds *is* the realization of structure to which the College of Liberal Arts since the function they serve unifies them.

And this problem compounds the further up the hierarchy we go. For to retain the view that the relevant functions are served by whatever singular entities they unify, we seem compelled to search out ever larger social systems to accommodate the view that the groups to which they correspond have identifiable boundaries. Again, the Department of Philosophy will become a member of the College of Liberal Arts such that the members of each are

37 This corresponds to a familiar issue, namely, the possibility of Max Black objects. Something that's worth considering is whether there are Max Black *groups*. If so, the monist might have resources to resist this challenge. But the issue is precisely how to square FUNCTIONALISM—that individuates groups by appeal to extrinsic properties—with monism—that individuates groups exclusively by appeal to their intrinsic properties.

doi: 10.48106/dial.v77.i3.03

in fact internally related. Moreover, the College of Liberal Arts will become a member of Bertrand Russell University such that all of their members are internally related, too. But, again, that robs us of the view that these are different groups. Whether this commits Ritchie to the existence of a single group—society itself, say—is beside the point.[38] The point is that in order to accommodate the view that groups are individuated by whatever functions unify their members, she'll commit herself to an implausible view of their interrelations.[39]

Functionalist Fineanism recommends a different response, one that might be available to Ritchie, too. (However, her failure to account for the role group kinds play in individuating groups remains a problem.) But although Fine doesn't have the problem I've raised for Ritchie's account, he has a relevantly similar problem.

Importantly, Fine can reasonably deny that groups are individuated solely by the relations their members realize internally; in particular, because there are no restrictions on the content of the principles of embodiment they manifest. Again, he might accept that though we individuate the department by its relation to the university, the university isn't part of that *thing*, the department. In particular, he might insist that because the relevant principles of embodiment are functional, particular extrinsic—or, in Ritchie's sense,

38 Interestingly, this suggests a kind of priority monism with respect to social goings-on. According to Schaffer (2010), a single thing—the universe—grounds everything else there is. The universe is fundamental. Similarly, the functionalist might accept that a single thing—society—is fundamental relative to the social and, so, with respect to groups. And in the way Schaffer appeals to quantum entanglement to justify his priority monism, the "social priority monist" might appeal to "social entanglement" to justify theirs. And that's an issue worth exploring, monism and pluralism aside.

39 She has another, better option. (I think there's problems with it, too, but it's important to emphasize that it isn't a problem for Ritchie's monism but for her commitment to Shapiro's view of structure.) As a result, I'll mention it, then set it aside. She might accept a plausible distinction between parthood and membership (*pace* Ruben 1985), and she might insist that both a group's members and the groups of which they're *part* are individuative. For the wider social systems of which particular groups are embedded are plausibly groups of which they're part rather than of which they're members. That's a significantly better option and one that's worth pursuing. However, it, too, risks a kind of arbitrariness. Again, it doesn't say which social systems are groups and which are the groups of which they're part. Certainly, it implies that the Department of Philosophy and the College of Liberal Arts are different groups; each serves a different function. But it also implies that they *aren't* different groups. As before, there is a function they *both* serve; there are many. It implies both that they are and aren't different groups. If being a member of a group concerns realizing the relevant functional kind, then even the membership/parthood distinction is inadequate to capture the view that they're different groups.

external—relations are needed to pick out the particular collections of individuals they unify. He's entitled to use Ritchie's distinction in that way.

For instance, he might say that to the extent that the Red Sox are a baseball team, the individuals that are its members are unified by the relations that define BASEBALL TEAM. Nonetheless, they're individuated by their relations to, e.g., the ALE—and, so, to the MLB—and to the City of Boston, too, because the relations between them are what make the Red Sox the unique instance of the kind it is; the very group it is. In other words, whereas the internal relations that make them a baseball team unify them, the external relations that make them *the* baseball team in question—the Red Sox—don't.

Nonetheless, this response makes Finean principles of embodiment intolerably arbitrary.[40] In particular, it suggests that if a group is individuated both by the structures it realizes internally and externally, there isn't a principled distinction between a particular group and the groups to which it's externally related. This is a version of the challenge raised for Ritchie. For if, in order to individuate the relevant collections of individuals, principles of embodiment appeal to relations that aren't definitional of the kinds of groups in question, it will be difficult to say which groups are which and why. That the department is unified by a principle that appeals to relations that don't unify the individuals that are its members—in this case, to the university—is at best stipulative. In other words, if variable embodiment is unifying, it's not clear why that which is externally related to that which is internal to a particular group doesn't have as much a right to be counted as part of the same group. As a result, it's not clear that he's justified in claiming that genuine unification occurs.

Relatedly, this response makes it impossible to tell whether to prefer Fine's monism to Uzquiano's pluralism. Again, Uzquiano claims that the relevant principles of embodiment don't stamp out singular but plural "entities." But since each assumes that principles of embodiment are either singular or plural, it's difficult to know how to decide between them. For both Fine and Uzquiano accept that for variable embodiments to be identical is for "them" to embody the same principle of embodiment. But if principles of embodiment are individuated by their modal profiles—as Fine (1999, 70) and Uzquiano

40 There are other responses available to Fine, too. For instance, he might accept that both internal and external relations *are* definitional of groups such that it's necessary that particular groups are embedded in the social systems in which they in fact are. But this gives rise to a host of other problems, in particular concerning the possibility that groups can have different *structures* at different times and in different worlds. Ultimately, solving that problem requires an account of group structure itself. Again, see (**?**) for an attempt.

doi: 10.48106/dial.v77.i3.03

(2018, 442)'s remarks suggest—it's not clear why we should think that a given group embodies a plural *rather* than a singular condition.[41] (This is as much a problem for Uzquiano as it is for Fine, of course, but I'll set that aside.)

4.3 *Functionalist Pluralism: Redux*

However, this isn't a problem for the functionalist pluralist; in particular, for ROLES. For given the distinction between internal and external structure, we can accept that groups are structured by the internal relations among the roles that ground their existence. We can accept that some of the roles departments of philosophy realize depend on others. For instance, we can accept that the role of being an associate professor is tied to the role of being an assistant professor in the way the role of being a pain is tied to that of being a wince. And though the proponent of ROLES accepts that the roles in question are interrelated, they get to deny that groups are *unified* by the relations among them. In particular, they have principled reasons to deny that groups are individuated solely by these relations and to accept that they're at least partly individuated by the social systems in which they're embedded. Their pluralism is precisely what vindicates their FUNCTIONALISM.

But, again, there are details to sort out. And, again, how a functionalist pluralist ought to conceive of structured social systems is important. Here's what I'm inclined to say. As I suggested in § 1, FUNCTIONALISM ranges over social systems. Given what I've said here, then, we might think of social systems as consisting of clusters of role-tokens, each of which corresponds to a group.[42] How tightly pluralities of roles cluster will correspond to the specificity of the functions they realize. For instance, the Red Sox play roles

41 A related problem arises because of Fine's assumption that groups are *either* rigid or variable embodiments. For it's important to emphasize that the difference between what I've called collective and individual embodiment isn't inconsequential; in particular, because there are considerations that count in favor of individual embodiment. For there seem to be groups that are neither rigid nor variable embodiments. For instance, the Jimi Hendrix Experience (JHE) seems to be a group that has some of its members rigidly—namely, Jimi Hendrix—and some of its members variably—namely, everyone else. As a result, though there's no time or world at which Jimi Hendrix isn't a member of JHE, there are both times and worlds at which other individuals are its other members. But that's incompatible with JHE being either a rigid or variable embodiment. However, it's perfectly reasonable to accept that some groups are constituted by *both* rigid and variable roles. JHE might be constituted both by a role that only Jimi Hendrix can play and by roles that many individuals can play.

42 I say either/or because, in addition to disagreeing about what structure is, we can meaningfully disagree about what social systems are.

that are clearly defined by the function they serve—again, to play baseball in a particular way within a particular set of institutions. However, genders—for instance, women—play roles that aren't as clearly defined and that interact with different roles—for instance, with race and class roles—in complicated ways.

But, again, because clusters of roles aren't singular entities, we can individuate them both by the structures they realize internally—that is, by the relations among the roles in question and because of which they can be said to cluster—and/or the structures they realize externally—that is, by their relation to other clusters. For instance, we can individuate the Department of Philosophy by identifying the roles the relevant individuals realize. And we can identify these by identifying the function they realize within the relevant set of institutions—and, ultimately, the maximal social system—in question. The Department of Philosophy is the department it is because it does philosophy in a particular way within a broader social system within which the other groups to which it's related are embedded, too.

5 Conclusion

In this paper, I've argued that the arguments from DIFFERENCE and SAMENESS are unsound. They obscure both the distinction between the definitional and ontological questions and between BEING GROUPED and BEING A MEMBER. I've articulated a version of functionalist pluralism—what I called ROLES—that bears this out.

Moreover, I've argued that once we make these distinctions, we see that the crucial question is whether the grouping relation is monistic or pluralistic, in particular, whether or not grouping is unifying. I've argued that if groups are one, the grouping relation *is* unifying and that this raises an important difficulty for the functionalist monist, namely, the problem of how to individuate groups. I've argued that if groups are many, this problem doesn't arise.

Although the implications for pluralism are clear, one of my aims is to generate interest in FUNCTIONALISM about groups, whether monistic or pluralistic. As I've suggested, there are important details about which we might reasonably disagree. Nonetheless, I hope to have shown we have reason to attend to them and, so, to treat FUNCTIONALISM as a viable metaphysical framework for theorizing about groups.*

* THANKS

doi: 10.48106/dial.v77.i3.03

Emilie Pagano
University of Texas at Austin
emilie.pagano@gmail.com

References

Baker, Lynne Rudder. 2007. *The Metaphysics of Everyday Life: An Essay in Practical Realism*. Cambridge Studies in Philosophy. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511487545.

Baxter, Donald L. M. 1988. "Identity in the Loose and Popular Sense." *Mind* 97(388): 575–582, doi:10.1093/mind/xcvii.388.575.

Bickle, John. 2020. "Multiple Realizability." in *The Stanford Encyclopedia of Philosophy*. Stanford, California: The Metaphysics Research Lab, Center for the Study of Language; Information. Revision, May 18, 2020, of the version of November 23, 1998, https://plato.stanford.edu/entries/multiple-realizability/.

Block, Ned. 1978. "Troubles with Functionalism." in *Minnesota Studies in the Philosophy of Science, Volume IX: Perception and Cognition: Issues in the Foundations of Psychology*, edited by C. Wade Savage, pp. 261–325. Minneapolis, Minnesota: University of Minnesota Press. Reprinted in Block (2007, 63–101).

—. 2007. *Consciousness, Function, and Representation. Collected Papers, Volume 1*. Cambridge, Massachusetts: The MIT Press.

Correia, Fabrice. 2017. "Real Definitions." in *Philosophical Issues 27: Metaphysics*, edited by Jonathan Schaffer, pp. 52–73. Oxford: Wiley-Blackwell, doi:10.1111/phis.12091.

Dasgupta, Shamik. 2014. "On the Plurality of Grounds." *Philosophers' Imprint* 14(20).

Dorr, Cian. 2016. "To Be *F* Is To Be *G*." in *Philosophical Perspectives 30: Metaphysics*, edited by John Hawthorne, pp. 39–134. Hoboken, New Jersey: John Wiley; Sons, Inc., doi:10.1111/phpe.12079.

Effingham, Nikk. 2010. "The Metaphysics of Groups." *Philosophical Studies* 149(2): 251–267, doi:10.1007/s11098-009-9335-4.

Epstein, Brian. 2015. *The Ant Trap. Rebuilding the Foundations of the Social Sciences*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199381104.001.0001.

—. 2018. "What are Social Groups? Their Metaphysics and How to Classify Them." *Synthese* 196(12): 4899–4932, doi:10.1007/s11229-017-1387-y.

Evnine, Simon J. 2016. *Making Objects & Events. A Hylomorphic Theory of Artifacts, Actions, & Organisms*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780198779674.001.0001.

Faller, August. 2019. "How Groups Persist." *Synthese* 198(8): 7149–7163, doi:10.1007/s11229-019-02514-0.

Fine, Kit. 1999. "Things and Their Parts." in *Midwest Studies in Philosophy 23: New Directions in Philosophy*, edited by Peter A. French and Howard K. Wettstein, pp. 61–74. Boston, Massachusetts: Basil Blackwell Publishers, doi:10.1111/1475-4975.00004.

—. 2020. "The Identity of Social Groups." *Metaphysics* 3(1): 81–91, doi:10.5334/met.29.

Fodor, Jerry A. 1968. *Psychological Explanation*. New York: Random House.

Griffith, Aaron M. 2020. "Realizing Race." *Philosophical Studies* 177(7): 1919–1934, doi:10.1007/s11098-019-01291-3.

Guala, Francesco. 2016. *Understanding Institutions: The Science and Philosophy of Living Together*. Princeton, New Jersey: Princeton University Press, doi:10.1515/9781400880911.

Haslanger, Sally. 1995. "Ontology and Social Construction." *Philosophical Topics* 23(2): 95–126, doi:10.5840/philtopics19952324.

—. 2000. "Gender and Race: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34(1): 31–55, doi:10.1111/0029-4624.00201.

—. 2012. *Resisting Reality. Social Construction and Social Critique*. New York: Oxford University Press, doi:10.1093/acprof:oso/9780199892631.001.0001.

—. 2016. "What is a (Social) Structural Explanation?" *Philosophical Studies* 173(1): 113–130, doi:10.1007/s11098-014-0434-5.

Hawley, Katherine. 2017. "Social Mereology." *The Journal of the American Philosophical Association* 3(4): 395–411, doi:10.1017/apa.2017.33.

Horden, John and López de Sa, Dan. 2021. "Groups as Pluralities." *Synthese* 198(11): 10237–10271, doi:10.1007/s11229-020-02715-y.

Jenkins, Katharine. 2017. "What Women are For: Pornography and Social Ontology." in *Beyond Speech. Pornography and Analytic Feminist Philosophy*, edited by Mari Mikkola, pp. 91–112. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780190257910.003.0005.

Koons, Robert C. 2014. "Staunch vs. Faint-Hearted Hylomorphism: Toward an Aristotelian Account of Composition." *Res Philosophica* 91(2): 151–177, doi:10.11612/resphil.2014.91.2.1.

Koslicki, Kathrin. 2018. *Form, Matter, Substance*. Oxford: Oxford University Press, doi:10.1093/oso/9780198823803.001.0001.

Lewis, David. 1980. "Mad Pain and Martian Pain." in *Readings in the Philosophy of Psychology. Volume I*, edited by Ned Block, pp. 216–222. Language and Thought. Cambridge, Massachusetts: The MIT Press. Reprinted, with a postscript (Lewis 1983b), in Lewis (1983a, 122–130), doi:10.1093/0195032047.003.0009.

—. 1983a. *Philosophical Papers, Volume 1*. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

—. 1983b. "Postscript to Lewis (1980)." in *Philosophical Papers, Volume 1*, pp. 130–132. Oxford: Oxford University Press, doi:10.1093/0195032047.001.0001.

MACKINNON, Catharine A. 1996. *Only Words*. Cambridge, Massachusetts: Harvard University Press, doi:10.2307/j.ctvjk2xs7.

MARMODORO, Anna. 2013. "Aristotle's Hylomorphism Without Reconditioning ." *Philosophical Inquiry* 37(1–2): 5–22, doi:10.5840/philinquiry2013371/28.

MCCULLAGH, Mark. 2000. "Functionalism and Self-Consciousness." *Mind and Language* 15(5): 481–499, doi:10.1111/1468-0017.00146.

MILLS, Charles W. 1997. *The Racial Contract*. Ithaca, New York: Cornell University Press, doi:10.7591/9780801471353.

PUTNAM, Hilary. 1975. *Mathematics, Matter and Method. Philosophical Papers, Volume 1*. Cambridge: Cambridge University Press. Second edition: Putnam (1979), doi:10.1017/cbo9780511625268.

—. 1979. *Mathematics, Matter and Method. Philosophical Papers, Volume 1*. 2nd ed. Cambridge: Cambridge University Press. First edition: Putnam (1975), doi:10.1017/cbo9780511625268.

RITCHIE, Katherine. 2013. "What are Groups?" *Philosophical Studies* 166(2): 257–272, doi:10.1007/s11098-012-0030-5.

—. 2015. "The Metaphysics of Social Groups." *Philosophy Compass* 10(5): 310–321, doi:10.1111/phc3.12213.

—. 2020. "Social Structures and the Ontology of Social Groups." *Philosophy and Phenomenological Research* 100(2): 402–424, doi:10.1111/phpr.12555.

ROSEN, Gideon, BYRNE, Alex, COHEN, Joshua and SHIFFRIN, Seana Valentine, eds. 2018. *The Norton Introduction to Philosophy*. 2nd ed. New York: W.W. Norton & Company.

RUBEN, David-Hillel. 1985. *The Metaphysics of the Social World*. London: Routledge & Kegan Paul.

RUSSELL, Bertrand Arthur William. 1903. *The Principles of Mathematics*. Cambridge: Cambridge University Press. Second edition: Russell (1937), doi:10.4324/9780203822586.

—. 1937. *The Principles of Mathematics*. 2nd ed. London: George Allen & Unwin. Second edition of Russell (1903), with a new introduction.

SCHAFFER, Jonathan. 2010. "Monism: The Priority of the Whole." *The Philosophical Review* 119(1): 31–76, doi:10.1215/00318108-2009-025.

SEARLE, John R. 1995. *The Construction of Social Reality*. London: Penguin Books.

—. 2010. *Making the Social World. The Structure of Human Civilization*. Oxford: Oxford University Press, doi:10.1093/acprof:osobl/9780195396171.001.0001.

SHAPIRO, Stewart. 1997. *Philosophy of Mathematics: Structure and Ontology*. Oxford: Oxford University Press, doi:10.1093/0195139305.001.0001.

SHOEMAKER, Sydney S. 1984. *Identity, Cause and Mind: Philosophical Essays*. 1st ed. Cambridge: Cambridge University Press. Second, expanded edition: Shoemaker (2003).

—. 2003. *Identity, Cause and Mind: Philosophical Essays*. 2nd ed. Oxford: Oxford University Press. Second, expanded edition of Shoemaker (1984).

Sider, Theodore. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford: Oxford University Press, doi:10.1093/019924443X.001.0001.

Thomasson, Amie L. 2019. "The Ontology of Social Groups." *Synthese* 196(12): 4829–4845, doi:10.1007/s11229-016-1185-y.

Tooley, Michael. 2001. "Functional Concepts, Referentially Opaque Contexts, Causal Relations, and the Definition of Theoretical Terms." *Philosophical Studies* 105(3): 251–279, doi:10.1023/a:1010381127525.

Uzquiano, Gabriel. 2018. "Groups: Toward a Theory of Plural Embodiment." *The Journal of Philosophy* 115(8): 423–452, doi:10.5840/jphil2018115825.

Wilhelm, Isaac. 2020. "The Stage Theory of Groups." *Australasian Journal of Philosophy* 98(4): 661–674, doi:10.1080/00048402.2019.1699587.

—. 2022. "Pluralities, Counterparts, and Groups." *Philosophical Studies* 179(7): 2133–2153, doi:10.1007/s11098-021-01755-5.

Witt, Charlotte. 2011. *The Metaphysics of Gender*. New York: Oxford University Press, doi:10.1093/acprof:oso/9780199740413.001.0001.

Young, Iris Marion. 1990. *Justice and the Politics of Difference*. Princeton, New Jersey: Princeton University Press. Second edition Young (2012), doi:10.4324/9781003060963-43.

—. 2012. *Justice and the Politics of Difference*. 2nd ed. Princeton, New Jersey: Princeton University Press. With a foreword by Danielle S. Allen, doi:10.1515/9781400839902.

Causal Inference from Big Data?
A Reply to Pietsch

Serena Galli

In his book *Big Data* (2021), Wolfgang Pietsch defends the view that variational induction, which stands in the tradition of Mill's methods, allows us to derive conclusions about causal relationships from observational data and that the algorithms that are most successfully applied for big data analysis implement variational induction. In his opinion, the fact that the analysis of big data by machine learning algorithms enables reliable predictions and effective interventions in the world supports the assumption that these algorithms correctly identify causal relationships. In the first part of the paper, I argue that attempts to infer causal relationships from observational data by variational induction face fundamental difficulties. Furthermore, I contend that these difficulties are not due to the specific way in which the method is spelled out but are manifestations of a general underdetermination problem. In the second part, I consider Pietsch's claim that the practical benefit of big data approaches indicates that variational induction implemented by machine learning algorithms generates causal knowledge. I provide a critical assessment of his notion of causal knowledge, and I argue that his conclusion relies on an inaccurate depiction of scientific practice.

In developing his views on variational induction, Pietsch relies on a difference-making account of causation. More specifically, he defines causal relevance and causal irrelevance as three-place relations between two variables and a given context or background:

> In a context B, in which a condition A and a phenomenon C occur, A is causally relevant to C, in short A $\mathcal{R}$ C | B, iff the following counterfactual holds: if A had not occurred, C would also not have occurred.
>
> In a context B, in which a condition A and a phenomenon C occur, A is causally irrelevant to C, in short A $\mathcal{I}$ C | B, iff the following

> counterfactual holds: if A had not occurred, C would still have occurred. (Pietsch 2016b, 5)

The truth value of the defining counterfactual statement is assessed in terms of difference-making, taking into account instances that are or were realized in our world (Pietsch 2016a, 11).[1] Methodologically, this assessment rests on the framework of variational induction, which stands in the tradition of Mill's (1889, 253ff) methods and comprises two key methods, namely, the method of difference and the strict method of agreement. To determine if a circumstance A is causally relevant for a phenomenon C with respect to background B, the *method of difference* must be employed:

> If two instances with the same background B are observed, one instance, in which circumstance A is present and phenomenon C is present, and another instance, in which circumstance A is absent and phenomenon C is absent, then A is causally relevant to C with respect to background B, iff B guarantees homogeneity. (Pietsch 2021, 33)[2]

Simply put, the homogeneity of the background ensures that all the circumstances that are potentially causally relevant for phenomenon C are held fixed, except for circumstance A, whose influence on phenomenon C is explicitly studied.[3]

In contrast, the *strict method of agreement* allows us to identify relations of causal irrelevance:

> If two instances with the same background B are observed, one instance, in which circumstance A is present and phenomenon C is present, and another instance, in which circumstance A is absent and phenomenon C is still present, then A is causally irrelevant to C with respect to background B, iff B guarantees homogeneity. (Pietsch 2021, 33)

1 This conception of counterfactual statements differs fundamentally from traditional counterfactual approaches to causation, such as those advanced by Lewis, who analyzes the truth conditions of counterfactual statements by referring to possible worlds (1973, 560–561).

2 In principle, causal relationships between continuous variables can be established likewise by extending the framework of variational induction by the method of concomitant variation (Pietsch 2021, 34). In the following, I will be concerned with binary variables exclusively.

3 I examine the homogeneity condition more closely in the context of epiphenomena. For a detailed discussion, cf. Pietsch (2021, 33–34; 2016b, 11–13).

For variational induction to yield reliable results, several conditions must be fulfilled. Most importantly, (i) all variables that are potentially causally relevant for the phenomenon of interest must be known, and (ii) the dataset must contain a sufficiently large number of observations covering all relevant constellations of the variable values. Pietsch acknowledges that, due to the fact that he makes these presuppositions, his account is (what he calls) *externally theory-laden*. However, he contends that his account avoids *internal theory-ladenness*, i.e., assumptions about causal connections between the variables considered. In other words, he claims to avoid the kind of theory-ladenness that is distinctive of hypothesis-driven approaches.[4] With the framework of variational induction, by contrast, the causal structure of the phenomenon of interest is supposed to be elaborated from the data alone.[5]

1 Causal Inference by Variational Induction and the Underdetermination Problem[6]

To point out the fundamental difficulty of inferring causal relationships by variational induction, I first focus on more complex causal structures, namely, (i) *symmetric overdetermination and preemption* and (ii) *epiphenomena*. Then, I evaluate whether the (iii) *directionality* of the relation of causal relevance can be established or not. For this assessment, I take for granted that the above-mentioned conditions for variational induction are met. In particular, I assume that every possible constellation of variable values that could have been generated by the causal structure in question is indeed observed and, moreover, that the set of observations involves neither measurement errors nor accidentally correlating variables.

(i) Let us consider the following dataset consisting of observations 1–4, which all share the same background B:

4 A prominent advocate of such an approach is Pearl, who maintains that "causal questions can never be answered from data alone" and that answering those questions "require[s] us to formulate a model of the process that generates the data, or at least some aspects of that process," also in the context of big data (Pearl and Mackenzie 2018, 351).

5 If the requirements for variational induction are met, "then there are enough data to avoid spurious correlations and to map the causal structure of the phenomenon without further internal theoretical assumptions about the phenomenon" (Pietsch 2015, 910–911). See also Pietsch (2021, 65–66).

6 Woodward uses the term *underdetermination problem* to refer to the circumstance that, given a set of variables, different causal structures encompassing these same variables can generate an identical pattern of correlations and conditional correlations (2003, 106–107).

doi: 10.48106/dial.v77.i3.04

Observation 1: A_1 & A_2 & C
Observation 2: A_1 & $\neg A_2$ & C
Observation 3: $\neg A_1$ & A_2 & C
Observation 4: $\neg A_1$ & $\neg A_2$ & $\neg$C

Further, let us suppose that the causal relationships that do, in fact, underlie these observations are those depicted by the model in figure 1a. In this model, we have two potential causes of C, A_1 and A_2, with A_1 preempting A_2 when both obtain. How does Pietsch propose to deal with this dataset? *Preempted* or *alternative causes* such as A_2 require other circumstances, in this case A_1, to be absent in order to have an impact on the phenomenon of interest. He claims that such alternative causes can be singled out based on the method of difference and the method of strict agreement too, but the background B must be specified by an additional condition X, the *preempting* cause:

> A is an 'alternative cause' to C with respect to background B, iff there exists an X such that A is causally relevant to C with respect to a background B & ¬X, but causally irrelevant to C with respect to a background B & X (i.e., C is always present in B & X). (Pietsch 2021, 34)

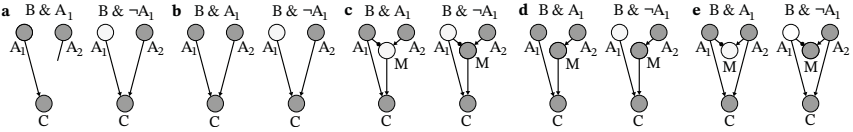

Figure 1: Symmetric overdetermination and preemption. In scenario (a), the exact mechanism of preemption is not specified and, therefore, symbolized by the prematurely terminated line originating from A_2.

Since A_2 is a preempted cause of C, it is causally relevant to C with respect to B only in the absence of A_1, which can be deduced by contrasting observations 3 and 4: A_2 $\mathcal{R}$ C | B & $\neg A_1$. Comparing observations 1 and 2, A_2 turns out to be causally irrelevant for C with respect to B & A_1, relying on the strict method of agreement (A_2 $\mathcal{I}$ C | B & A_1). Yet, two problems arise from Pietsch's approach for the identification of alternative causes: First, A_1, which is the *preempting cause* of C, is determined to be causally irrelevant for C when applying the strict method of agreement to observations 1 and 3. As seen before, it is only

in the absence of A_2 that the causal relevance between A_1 and C with regard to background B can be established (A_1 $\mathcal{R}$ C | B & $\neg A_2$, observations 2 and 4). However, by definition, the impact of A_1 on C does not depend on the value of the variable A_2. Thus, variational induction fails to depict the asymmetry between the preempting and the preempted cause. Second, the above-listed observations 1–4 are compatible with another underlying causal structure, namely a model of symmetric overdetermination, as displayed in figure 1b. Hence, without any prior knowledge about the causal connections between the variables, it is impossible to decide which causal structure really underlies the observed constellations of variable values.

Then, a common feature of big data is its high dimensionality, meaning that each observation includes numerous different variables. So, it could be put forward that this problem only arises because the dataset is not sufficiently complex and not enough variables were regarded. For example, introducing the variable M, which mediates the causal effect of A_2 on C and whose instantiation is prevented in the presence of A_1, would definitively allow to distinguish between the case of preemption and the case of symmetric overdetermination, as depicted in figures 1c and 1d.[7]

Observation 5: A_1 & A_2 & $\neg$M & C
Observation 6: A_1 & $\neg A_2$ & $\neg$M & C
Observation 7: $\neg A_1$ & A_2 & M & C
Observation 8: $\neg A_1$ & $\neg A_2$ & $\neg$M & $\neg$C

Observation 5 is indeed not compatible with the model of symmetric overdetermination in figure 1d because, according to that model, the variable A_1 has no impact on the other cause A_2 or its mediating variable M. Yet, these four observations are consistent with another model of symmetric overdetermination, where the instantiation of M depends both on the presence of A_2 and the absence of A_1, but M does not mediate the causal effect of A_2 on C, as illustrated in figure 1e. Hence, including more variables does not solve, but, at best, deflects the above-mentioned difficulties.

(ii) In connection with *epiphenomena*, similar problems arise. Epiphenomena, such as A_4 in figure 2c, have a common cause with the phenomenon of interest but have no causal impact on it. Let us consider another dataset con-

7 According to Lewis' terminology, the causal model displayed in figure 1c is an example of early preemption (1986b, 200).

doi: 10.48106/dial.v77.i3.04

sisting of observations 9 and 10, which, in turn, share the same background B:

Observation 9: A_3 & A_4 & C
Observation 10: ¬A_3 & ¬A_4 & ¬C

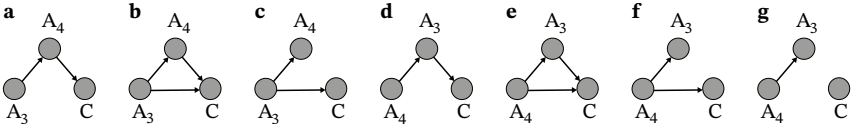

Figure 2: Epiphenomena and directionality of causal connections.

Then, let us suppose that these observations, which are compatible with all the models depicted in figure 2a–f, were generated by the causal model in 2c. Since A_3 and A_4 covary, neither the method of difference nor the strict method of agreement can be applied to determine whether A_4 stands in a relation of causal relevance or irrelevance to C with respect to background B.[8] By contrast, A_3 proves to be causally relevant to C with respect to B, which, in this case, does guarantee homogeneity.[9] On behalf of Pietsch, it could be put forward that this problem can be circumvented by considering the combination of the variables A_3 and A_4 instead of examining them separately. Following this approach, the method of difference establishes either the conjunction or the disjunction of A_3 and A_4 to be causally relevant for C with respect to background B (A_3 & A_4 $\boldsymbol{\mathcal{R}}$ C | B or $A_3 \vee A_4$ $\boldsymbol{\mathcal{R}}$ C | B).[10] Still, in all cases, variational induction fails to establish that A_4 is causally irrelevant to C with respect to B.

Pietsch acknowledges the issue that algorithms employing variational induction may mistakenly single out epiphenomena as causally relevant for a

8 A_3, which is potentially causally relevant for C as well, cannot be held fixed, as it strictly covaries with A_4. Therefore, B does not guarantee homogeneity.

9 B guarantees homogeneity with respect to the relationship between A_3 and C if "only circumstances that are causally irrelevant to C can change" or that "lie on a causal chain through $A_{[3]}$ to C or that are effects of circumstances that lie on this causal chain" (Pietsch 2021, 33–34). Since A_4 is an effect of A_3, B does guarantee homogeneity, although A_4 cannot be held fixed. However, presuming that A_4 is connected to A_3 in this way contradicts Pietsch's claim that his account avoids assumptions about causal connections between the variables considered.

10 Such Boolean expressions are, as Pietsch maintains, a possible result of variational induction (2021, 50).

phenomenon of interest, although they lack the difference-making character of a cause (Pietsch 2021, 55; 2016a, 153–154).[11] However, he attributes it to the fact that either the dataset is incomplete or the algorithm does not fully implement variational induction. By contrast, this example demonstrates that this erroneous conclusion is not due to missing observations because it is drawn despite considering all observations compatible with a given causal model. On the other hand, it cannot be ascribed to the algorithmic implementation either, as the manual, non-algorithmic application of variational induction does not satisfyingly deal with epiphenomena either.[12]

(iii) Finally, the fact that observations 9–10 could have been generated by a causal structure with A_3 being a cause of A_4 (figures 2a–c) or a causal structure with A_4 being a cause of A_3 (figures 2d–f) demonstrates that the *direction of the relation of causal relevance* cannot be established by variational induction. Besides, the same holds true for phenomenon C, which could as well be a cause and not an effect of variable A_3 or A_4 if not predefined as the phenomenon of interest. To solve this problem, Pietsch has suggested introducing a time index for the phenomenon of interest and the variables examined (2014, 424). Yet, from a conceptual point of view, this seems like an ad hoc solution since the truth condition he specifies for the counterfactual defining causal relevance fails to capture the asymmetry in the relation of causal relevance. Additionally, this solution is not practicable for cross-sectional data, where all the variables are recorded at the same time and, accordingly, the timepoint of their occurrence remains unknown.

In light of these difficulties, Pietsch's assertion that it is possible "to determine the true causal relationships by means of variational induction" seems to be unwarranted (2021, 61). In the causal discovery literature, it is well established that, given the causal Markov condition and the causal faithfulness condition, certain features of the underlying causal structure can be deduced from the probability distribution in the data. But, aside from special cases, it

11 Strictly speaking, he refers to *proxies*, which I take to be the equivalent of epiphenomena.

12 Needless to say, some of such wrong conclusions can be traced back to issues regarding data acquisition. For instance, a sampling error can result in an accidental correlation between variables that is not present in the population from which the sample was drawn. Let us suppose that observations 9–10 were generated by the causal structure displayed in figure 2g. In this case, the observed correlation between A_3 and A_4, on one side, and C, on the other side, must have occurred by chance. Yet, by employing variational induction, the conjunction or disjunction of A_3 and A_4 is mistakenly identified as causally relevant for C with respect to B, and this misattribution can be recognized as such and corrected only when analyzing another, possibly larger dataset devoid of this accidental correlation.

doi: 10.48106/dial.v77.i3.04

is not possible to uniquely determine the true causal structure.[13] In my view, variational induction similarly faces the problem of underdetermination as it rests ultimately on the analysis of patterns of (conditional) dependencies and independencies in the data. That is to say, variational induction aims at identifying the constellation of variables V that has difference-making character with respect to background B: If this exact configuration of variables is present, C is always present as well (i.e., $P(C \mid V) = 1$); in its absence, C is always absent as well (i.e., $P(C \mid \neg V) = 0$). From this dependence between C and V, a relation of causal relevance is inferred ($V \mathrel{\mathcal{R}} C \mid B$). Thus, the procedure of variational induction can be viewed as the comparison of the conditional probabilities of C rather than the comparison of individual observations. From the pattern of conditional probabilities based on observations 9 and 10, for example, it can be deduced that it is either the conjunction or the disjunction of A_3 and A_4 that makes a difference for the value of the variable C with respect to background B (i.e., $V = (A_3 \,\&\, A_4)$ or $V = (A_3 \vee A_4)$).[14] While these two Boolean expressions are highly useful for predicting the value of C, the pattern of dependencies is, as already stated for the direct comparison of individual observations, compatible with all the models depicted in figures 2a–f. Accordingly, Pietsch's claim that "the difference-making circumstances identified by variational induction are exactly the circumstances that need to be manipulated or intervened upon in order to change a phenomenon" does not seem justified. Although a given configuration of circumstances might unequivocally determine the value of the phenomenon of interest C in an observational setting, the exact connection between these circumstances and C remains elusive. Therefore, successful intervention strategies cannot be deduced from the Boolean expression of these circumstances. While in 2a, 2d, 2e, and 2f a single intervention on A_4 is an effective way of manipulating the value of C, in 2c this is clearly not the case. The Boolean expression may

13 These two conditions are so-called *bridge principles*, which are required to connect the observations of a given set of variables to the underlying causal model that generated these observations. More specifically, the causal Markov condition allows the inference from a probabilistic dependence between two variables to a causal connection, whereas the causal faithfulness condition allows the inference from a probabilistic independence to causal separation. Cf. Eberhardt (2017, 82–85). For a discussion of underdetermination in causal inference in relation to different success criteria and background assumptions, see Zhang (2009).

14 Since the configurations $(A_3 \,\&\, \neg A_4)$ and $(\neg A_3 \,\&\, A_4)$ do not occur in a purely observational setting, these two possibilities cannot be distinguished.

encompass the necessary and together sufficient circumstances for *observing* phenomenon C but not for *producing* it.[15]

2 Objectives of Big Data Analysis and Causal Knowledge[16]

Pietsch distinguishes two central functions of big data approaches, namely, *prediction* and *intervention*, and claims that the exertion of both requires some access to causal knowledge.[17] Arguably, the view that causal knowledge is indispensable for effectively manipulating a phenomenon of interest is hardly contested. However, causal knowledge is usually not considered a prerequisite for predictive success.[18] In that regard, it is useful to touch upon Pietsch's notion of causal knowledge, which bears on his distinction of direct and indirect causal connections: If a certain variable is causally relevant for another variable in a given context, as it is the case for A_3 and C in figure 2c, the relationship between those two variables constitutes a *direct causal connection*, as Pietsch suggests. If, by contrast, two variables are not causally relevant for one another but are related via a common cause, then there exists an *indirect causal connection* between these two variables, as it is the case for A_4 and C in

15 As an anonymous reviewer pointed out, a promising way of dealing with this problem of underdetermination is the appeal to theoretical virtues such as parsimony. For example, Forster et al. introduce the principle of frugality that favors those causal structures with the fewest causal connections (2018). I fully agree that, technically, the procedure of variational induction could be combined with an algorithm that ranks the possible causal structures in terms of simplicity. Yet, this constraint regarding the total number of causal connections involves an assumption about the causal connections between the variables since causal models with more numerous connections, such as 2b or 2e, are dismissed in favor of models with fewer connections, although perfectly compatible with the data. Therefore, this strategy is *internally theory-laden* and not reconcilable with the concept of variational induction as a purely data-driven approach. An alternative strategy to determine the true causal structure is experimentation. For a detailed discussion of experimentation as a means for resolving underdetermination, cf. Eberhardt (2013).

16 In this section, which is concerned with variational induction as a means of causal inference from big data specifically, I acknowledge Pietsch's claim that the most successful algorithms are based on variational induction without further examination.

17 Rather than distinguishing between different functions, I would propose to differentiate between two questions that are to be answered by big data analysis. To specify *intervention* as a function of big data approaches presupposes what is under consideration. Besides, it remains unclear how to discern which function, prediction or intervention, is exerted in a given case.

18 For example, Woodward maintains that accurate predictions can be made based on correlations solely; furthermore, he points out that "inferences from effect to cause are often more reliable than inferences from cause to effect" (2003, 31–32).

doi: 10.48106/dial.v77.i3.04

figure 2c, which are both effects of A_3 (Pietsch 2021, 55).[19] Certainly, only for successful interventions upon a phenomenon of interest must it be known whether there is a direct causal connection between that same phenomenon and the variable that is to be manipulated or not. But, as he argues, for accurate predictions, either such a direct causal connection or an indirect causal connection between the phenomenon of interest and a potential predictor variable is required. Thus, when a machine learning algorithm singles out a variable as a promising predictor variable for a given phenomenon of interest, the algorithm thereby generates causal knowledge to a certain degree. In my opinion, this broad notion of causal knowledge allowing for different degrees is particularly problematic in three respects:

(i) First of all, as a cause (usually) correlates with the phenomenon of interest, so does an epiphenomenon of this cause. The first correlation is indicative of a (direct) causal connection, whereas the second is indicative of a common cause. As discussed for epiphenomena, variational induction does not allow us to distinguish between a correlation ascribable to a direct causal connection and a correlation ascribable to a common cause. It follows that not only variational induction but also the analysis of (conditional) correlations yields causal knowledge in this wide sense. Accordingly, it does not seem consistent to specify correlation as a contrasting notion for causation. Furthermore, since the procedure of variational induction makes use of the pattern of dependencies in the data, it does not even allow for a distinction between correlations that are indicative of some sort of causal connection and purely accidental correlations. Therefore, it remains unclear in what sense big data algorithms are capable of delimiting causation from correlation, as Pietsch maintains.[20]

(ii) This broad notion of causal knowledge stands in tension with Pietsch's claim that the primary function of causal knowledge is to guide us on how to effectively intervene in the world (2021, 54). If the knowledge of an indirect causal connection between two variables is regarded as causal knowledge as well, having access to causal knowledge in this wide sense does not help to discriminate between effective and ineffective strategies to manipulate a phenomenon of interest.

19 Pietsch's distinction of direct and indirect causal connections differs from the conventional view, according to which the difference between direct and indirect causal connection results from the absence or presence of a mediating variable. See, for example, Woodward (2003, 55).

20 "By relying on variational induction, big data approaches are to some extent able to distinguish causation from correlation" (Pietsch 2021, 57).

(iii) And, finally, it risks obscuring the distinction between questions of prediction and questions of intervention, which are addressed in scientific practice: Causal knowledge in the strict sense, that is to say, knowledge about relations of causal relevance and irrelevance in a given set of variables, is no precondition for predictions. Thus, the fact that an algorithm implementing variational induction yields accurate predictions cannot be cited in support of the view that variational induction is capable of establishing causal relationships. Conversely, interventions indeed depend on causal knowledge in the strict sense. If such an algorithm truly did enable us to efficiently intervene in the world, this could speak in favor of Pietsch's view that variational induction is capable of inferring direct causal connections. As an example, he refers to "algorithms [that] are designed to determine the best medicine to cure a certain cancer" (Pietsch 2021, 54).[21] In fact, there are a number of studies that relied on machine learning in order to predict the response to a given drug. In a recently published work, the tumor tissue of patients with breast cancer was analyzed with different methods at diagnosis (Sammut et al. 2022). Patients were subsequently treated with chemotherapy, and treatment response was evaluated. Using a machine learning algorithm, the authors built a model to predict the response to chemotherapy, which was based on the molecular profile of the tumor as well as clinicopathological features, and model performance was successfully validated on a different dataset. Amongst other things, they drew the conclusion that patients predicted to show a poor response to standard-of-care chemotherapy should be enrolled in clinical trials investigating novel therapies. Therefore, the results of this big data approach may allow for better stratification of patients that will or will not benefit from conventional chemotherapy and are inasmuch action-guiding. However, Pietsch maintains that these algorithms are, moreover, designed to *determine the best treatment* for a given cancer, in this way allowing us to effectively intervene upon the phenomenon of interest, namely, tumor growth. For the sake of argument, let us suppose the algorithm revealed that three signaling pathways are hyperactive in tumors poorly responding to chemotherapy compared to tumors displaying a good treatment response. But, as outlined above, it is impossible to determine if (or which of) these three pathways are indeed driving tumor growth and which are rather an epiphenomenon of the cause of excessive tumor growth or even a consequence thereof. Accordingly, the question whether one of these hyperactive signaling

21 He does not cite any specific publication to underpin his assertion.

doi: 10.48106/dial.v77.i3.04

pathways truly constitutes a promising therapeutic target or not cannot be answered based on the observational data alone but requires experimentation. Besides, in order to successfully intervene in the world, it is essential not only to identify the causes of the phenomenon but also to understand how these causes can be manipulated, specifically which drug effectively targets a given pathway (compare figure 3). This kind of causal knowledge may be generated in randomized controlled trials or *in vitro* studies but, again, cannot be derived from observational data. To trace back the practical benefit of big data approaches to the generation of causal knowledge by the algorithms used does not accurately reflect the scientific practice, which builds upon different sources of knowledge to determine effective interventions.

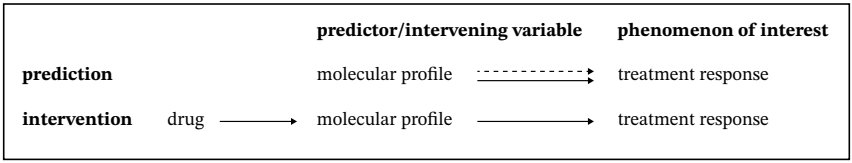

Figure 3: Prediction of and intervention upon a phenomenon of interest. Indirect causal connections are represented by dashed arrow lines, direct causal connections by solid arrow lines.

3 Conclusions

In my view, variational induction fails to elucidate causal structures involving preemption, symmetric overdetermination, or epiphenomena, establishing causal relationships between variables that actually are conjoined in a relation of causal irrelevance and vice versa. Furthermore, the direction of the relation of causal relevance cannot be specified by variational induction, which poses a problem for even the simplest causal models possible. These shortcomings are neither specific to the method of variational induction nor ascribable to an imperfect dataset with missing observations, an insufficient number of observed variables, or any measurement errors. Rather, the attempt to infer causal relationships from observational data (including big data) itself faces important limitations: Since for a given set of observations multiple underlying causal structures are usually conceivable, generally it is impossible to uniquely determine the true causal model from this set of observations alone without

endorsing any background assumptions about or having any prior knowledge of the causal relationships between the variables involved.

Pietsch's notion of causal knowledge explains, at least partially, why he reaches another assessment of variational induction as a method for generating causal knowledge. Supposing a broader notion of causal knowledge, he seems to have in mind a less strict success criterion: It suffices for variational induction to approximate causal relationships, namely, to determine if there is any causal connection, direct or indirect, between two variables. This could create the appearance that the conflicting assessment of variational induction as a means to infer causal relationships is merely due to two divergent, equally plausible notions of causal knowledge. However, in my opinion, Pietsch's broad notion of causal knowledge is problematic because it blurs the distinction between causation and correlation and between the prerequisites for prediction and for intervention.

If the practical benefit of big data approaches cannot be attributed to the elucidation of causal relationships, an alternative explanation is needed. The identification of predictive markers may indeed improve patient care by sparing patients who are unlikely to respond to the adverse reactions of an ineffective treatment. Randomized controlled trials can yield negative results only because patients are not selected appropriately. This could be obviated by a more adequate patient stratification based on reliable predictor variables. The analysis of the molecular profile of a tumor can generate promising hypotheses about chemoresistance in a relatively unbiased way, which may prove to be true in experimental assays. Undoubtedly, the results of machine learning algorithms contain very valuable information, which, in conjunction with knowledge derived from other sources, provide reasons to act in a certain way. In this sense, Pietsch is right in stating that precisely data-rich sciences such as medicine are fundamentally concerned with difference-making relationships and that the correlations unveiled by machine learning algorithms certainly do not replace causation. But, although such results of big data analysis can be action-guiding and aid in singling out potentially effective interventions, this should not be taken as a confirmation of the claim that machine learning algorithms indeed elucidate the causal structure underlying the observational dataset.*

Serena Galli

* I thank Peter Schulte and two anonymous reviewers for their helpful comments on an earlier version of this paper.

doi: 10.48106/dial.v77.i3.04

0000-0002-9630-8645
University of Zurich
serenacarolina.galli@uzh.ch

References

EBERHARDT, Frederick. 2013. "Experimental Indistinguishability of Causal Structures ." *Philosophy of Science* 80(5): 684–696, doi:10.1086/673865.

—. 2017. "Introduction to the Foundations of Causal Discovery." *International Journal of Data Science and Analytics* 3: 81–91, doi:10.1007/s41060-016-0038-6.

FORSTER, Malcolm, RASKUTTI, Garvesh, STERN, Reuben and WEINBERGER, Naftali. 2018. "The Frugal Inference of Causal Relation." *The British Journal for the Philosophy of Science* 69(3): 821–848, doi:10.1093/bjps/axw033.

LEWIS, David. 1973. "Causation." *The Journal of Philosophy* 70(17): 556–567. Reprinted, with a postscript (Lewis 1986b), in Lewis (1986a, 159–213), doi:10.2307/2025310.

—. 1986a. *Philosophical Papers, Volume 2*. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

—. 1986b. "Postscript to Lewis (1973)." in *Philosophical Papers, Volume 2*, pp. 172–213. Oxford: Oxford University Press, doi:10.1093/0195036468.001.0001.

MILL, John Stuart. 1843. *A System of Logic, ratiocinative and inductive*. London: Longmans, Green & Co.

—. 1889. *A System of Logic, Ratiocinative and Inductive, Being a Connected View of the Principles of Evidence and the Methods of Scientific Investigation*. London: Longmans, Green & Co. First edition: Mill (1843).

PEARL, Judea and MACKENZIE, Dana. 2018. *The Book of Why: the New Science of Cause and Effect*. New York: Basic Books, http://bayes.cs.ucla.edu/WHY/.

PIETSCH, Wolfgang. 2014. "The Structure of Causal Evidence Based on Eliminative Induction." *Topoi* 33(2): 421–435, doi:10.1007/s11245-013-9190-y.

—. 2015. "Aspects of Theory-Ladenness in Data-Intensive Science." *Philosophy of Science* 82(5): 905–916, doi:10.1086/683328.

—. 2016a. "The Causal Nature of Modeling with Big Data." *Philosophy & Technology* 29(2): 137–171, doi:10.1007/s13347-015-0202-2.

—. 2016b. "A Difference-Making Account of Causation." Unpublished manuscript, https://philsci-archive.pitt.edu/id/eprint/11913.

—. 2021. *Big Data*. Cambridge: Cambridge University Press, doi:10.1017/9781108588676.

SAMMUT, Stephen-John, CRISPIN-ORTUZAR, Mireia, CHIN, Suet-Feung, PROVENZANO, Elena, BARDWELL, Helen A., MA, Wenxin, COPE, Wei, et al. 2022.

"Multi-Omic Machine Learning Predictor of Breast Cancer Therapy Response ." *Nature* 601(January): 623–629, doi:10.1038/s41586-021-04278-5.

WOODWARD, James F. 2003. *Making Things Happen. A Theory of Causal Explanation*. Oxford Studies in the Philosophy of Science. Oxford: Oxford University Press, doi:10.1093/0195155270.001.0001.

ZHANG, Jiji. 2009. "Underdetermination in Causal Inference." *Studies in Logic* 2(4): 16–47.

doi: 10.48106/dial.v77.i3.04

Retro-Closure Principle and Omniscience

Ciro De Florio & Aldo Frigerio

Todd and Rabern (2021) have argued that if we assume that future contingents are untrue and if we accept the Retro-closure principle ($p \rightarrow \mathbf{PF}(p)$), then the existence of a temporal omniscient entity becomes metaphysically impossible. Since the truth of a metaphysical and theological theory should not be dependent on questions of temporal semantics, Todd and Rabern conclude that, if one wishes to maintain that future contingents are untrue, one must abandon the Retro-closure principle. The aim of this paper is to propose a temporal semantics system in which future contingents are untrue, the Retro-closure principle is valid, and the possibility of the existence of an omniscient and temporal being is guaranteed.

The future is uncertain. We do not know if it will rain tomorrow or if a drought will persist. Over centuries, philosophers have been wondering whether this uncertainty concerns only the epistemic dimension—and, therefore, it is only the result of our ignorance—or whether it is rooted in the ontological dimension, such that the world itself is at least partly indeterminate. Those who believe that the future is—at least to a certain extent—indeterminate should account for this intuition through an adequate semantics of future tense sentences. Let us assume that the actual state of the world and its natural laws do not determine the weather of tomorrow. According to many future tense semantics, "It will rain" is not true in this situation. However, suppose that time passes, and the following day, it rains. What would be the truth value of the proposition expressed yesterday by the sentence, "It will rain tomorrow," if considered from today's perspective, given that it rains today? Many may have the intuition that this proposition is true today. Following Todd and Rabern (2021), we call the content of this intuition the Retro-closure principle (RCP). In fact, several future tense semantics—though not all—account for this intuition.

Todd and Rabern (2021) have advanced an interesting argument against what they call Open-closurism (OC), which is the conjunction of two theses: 1) Open futurism (OF), following their terminology, which refers to the view that future contingents are untrue (i.e., false or neither true nor false); and 2) RCP. Theirs is an indirect argument: they do not argue against OC directly but instead show that OC is incompatible with the possibility of an omniscient temporal being. Even though we do not want to be committed to the existence of an omniscient temporal being, it is odd that semantic principles would determine a complex metaphysical question, such as the possibility of an omniscient entity. Todd and Rabern's conclusion is that we have to drop OC and, consequently, those OF semantics that validate RCP.

In this paper, we present a future tense semantics that accepts both OF and RCP without implying the impossibility of an omniscient temporal being. As a consequence, Todd and Rabern's objection fails within this semantics or semantics similar in kind. As we will see below, one of the main features of our semantics is the presence of two indices, the first of which indicates the time at which a formula is evaluated, while the second indicates the perspective from which it is evaluated.

The paper is structured as follows: In section 1, we put forward some considerations about the plausibility of RCP. In section 2, Todd and Rabern's objection is presented in detail. In section 3, the branching time semantics is introduced, and the difficulties of Thomason's supervaluationism in responding to Todd and Rabern's objection are considered. In section 4, a particular OC semantics is presented. It is shown that this semantics allows for the possibility of an omniscient temporal being, thus refuting Todd and Rabern's argument. Finally, section 5 contains some closing remarks.

1 Intuitions in Favour of RCP

As indicated above, OC is—at least prima facie—a plausible view. In this section, we would like to further elaborate on our intuitions about OC and show that OC appears to be quite natural when it is "referred," so to speak, to the past and present; the extension to the future case is quite natural, and, therefore, the sceptic about OC has much work to do.[1]

1 In this paper, we will defend a particular version of OC for which future contingents are neither true nor false. There is another version of OC for which future contingents are false. The combination of this view with RCP seems to us much less plausible, so we will ignore this version of OC here.

Suppose today is December 9th, 2022, and the world is such that it is not determined whether it will rain tomorrow. However, it is determined that tomorrow, Paul will believe the proposition expressed by "It's raining today." In other words, the current state of the world and the laws of nature determine a particular configuration of the neurons of Paul's brain, such that on December 10th, 2022, Paul believes the proposition expressed by the indexical sentence "It's raining today." Is Paul's belief of tomorrow correct today? In other words, when Paul believes tomorrow that it is raining that day, will he have a correct belief? Intuitively, we can respond neither positively nor negatively to this question. The answer depends on how things will go: if it rains tomorrow, then Paul's belief will be correct; if it does not rain, it will be incorrect. However, since it is not determined today whether it will rain tomorrow, it is also not determined whether Paul's belief will be correct. Tomorrow, however, it will rain, or it will not rain, and Paul's belief will be either correct or not. Therefore, the correctness of Paul's belief in the proposition expressed by the sentence "It is raining today," uttered on December 10th, is determined by what happens on December 10th and cannot be determined before that date.

Let us now make a change to our initial scenario. Suppose that today is December 9th, but in this case, Paul is not determined to believe something tomorrow; rather, he believes something the day after tomorrow, on December 11th. In particular, he is determined to believe the proposition expressed by the sentence "It rained yesterday." Let us assume again that on December 9th, it is undetermined whether tomorrow, December 10th, it will rain or not. Once again, we can ask whether Paul's belief of December 11th that it had rained the day before would be correct. Once again, however, today, December 9th, we can neither say that it is correct nor that it is incorrect because it is not determined whether it will rain tomorrow on December 10th. Only tomorrow will the correctness of that belief be determined. Tomorrow, we will be able to say whether Paul's belief of the day after tomorrow will be correct or not. In fact, the correctness of Paul's belief on the day after tomorrow depends on what happens tomorrow. However, today, it is indeterminate what will happen tomorrow, and, therefore, it is also indeterminate whether Paul's belief will be correct. Nevertheless, time passes, and the world determines itself. Tomorrow, it will rain, or it will not rain. Consequently, from tomorrow's perspective, Paul's belief of December 11th, that it rained the day before, will be correct or incorrect. Therefore, the correctness of Paul's belief in the proposition expressed by the sentence "It rained yesterday," uttered on December 11th, is

doi: 10.48106/dial.v77.i3.05

determined by what happens on December 10th and cannot be determined before that date.

Now, suppose that Paul believes today, December 9th, that it will rain tomorrow. Again, it is not determined today whether it will rain tomorrow or not. Is Paul's belief correct or incorrect? By symmetry with the previous cases, we would say that from today's perspective, his belief is neither correct nor incorrect. The correctness of Paul's belief depends on what will happen tomorrow, but today, what will happen tomorrow is indeterminate; therefore, the correctness of Paul's belief is not determined either. However, the world moves forward and determines itself. Tomorrow, it will rain, or it will not rain. On the basis of what will happen tomorrow, it will be possible to say whether Paul's belief of December 9th was correct or not. If it rains, it was correct; otherwise, it was incorrect. Therefore, from tomorrow's perspective, Paul's belief will have been correct or incorrect. The correctness of Paul's belief in the proposition expressed by the sentence "It will rain tomorrow," uttered on December 9th, is determined by what happens on December 10th and cannot be determined before that date.

Those who deny OC must deny that it is indeterminate today whether Paul's belief that it will rain tomorrow is correct or incorrect and that instead, from tomorrow's perspective, it can be said that it was correct or not. To deny this, they must either 1) deny that the correctness of Paul's beliefs is determined in the course of time or 2) deny the symmetry between the future case and the present and past cases. Let us consider these two possibilities in detail:

(1) One could reject that the correctness of Paul's belief in the proposition expressed by the sentence "It is raining today," uttered on December 10th, is determined only when the present is December 10th and cannot be determined before that date. This can be done in at least two ways: (i) Either it is claimed that it is already true today, December 9th, that such a belief is correct or incorrect; or (ii) it is denied that tomorrow, on December 10th, Paul's belief becomes correct or not. Since it is indeterminate today, on December 9th, whether it will rain tomorrow or not, (i) is only possible assuming that one of the possible futures is the true one. In other words, (i) is only possible when assuming a Thin Red Line semantics of the future. This means that OC must be abandoned. We will ignore this solution because our goal here is to defend the consistency of OC and, thus, the compatibility of OF with RCP, and not to argue for OF. Then, (ii) is completely implausible: suppose it rains

tomorrow, and that Paul believes it rains. How can we deny that Paul's belief is correct? Suppose that someone says to Paul, "It is true that you believe that it is raining today, and it is true that it is raining today, but your belief is not correct because it was uncertain yesterday whether it would rain today, and therefore it was also uncertain yesterday whether your belief would be correct or not." Paul, and we with him, would find this reasoning absurd. Paul would probably say, "Regardless of how things were yesterday, it is a fact that it is raining today and that I believe it is raining today. Therefore, my belief about today is correct *today*."

(2) More plausibly, the symmetry between past and present cases, on the one hand, and the future case, on the other hand, might be denied. It is true that the correctness of Paul's belief of December 10th, that it is raining that day, is determined on December 10th, and it is true that the correctness of Paul's belief of December 11th, that it rained the day before, is determined on December 10th. However, the correctness of Paul's belief of December 9th that it will rain the following day is not determined on December 10th. This amounts to saying that while the correctness of Paul's belief that it is raining today is determined by what happens today and that the correctness of Paul's belief that it rained yesterday is determined by what happened yesterday, the correctness of Paul's belief that it will rain tomorrow is not determined by what will happen tomorrow. However, it is hard to see why the case of the future should not be similar to the present and past cases.

The only plausible argument to deny the symmetry between these cases could be this: in the first two cases, the passage of time determines the correctness of a present or future belief; in the third case, the passage of time determines the correctness of a past belief. However, the past is fixed and unchangeable; therefore, the flow of time cannot determine something in the past. Nevertheless, as Todd and Rabern (2021, 106) point out, a change in the correctness of a belief is an extrinsic change (or a so-called "Cambridge change"), not an intrinsic one. When the past is said to be fixed, it is usually assumed that it is fixed with respect to intrinsic changes. However, extrinsic changes do not seem to be barred by the fixity of the past. For instance, World War I acquired the property of having ended 21 years before the outbreak of World War II in 1939. However, acquiring this property poses no problem for the fixity of the

doi: 10.48106/dial.v77.i3.05

past because it is an extrinsic property. Similarly, it is not a problem for the fixity of the past that a past belief becomes correct after the fact.[2]

A further argument can be advanced in favour of OC; something similar to this principle seems to be valid for other verb tenses as well, particularly for the progressive. Bonomi (1997) gives the following example. Suppose that Leo has just begun a journey in France. In the first stage, he drives from Milan to Dijon, where he arrives on July 14th at a quarter to three p.m. He does not stop there because he plans to spend his first night in France in one of the following cities: Besançon, Metz, or Paris. Actually, he must make a decision since three different routes correspond to these alternatives. However, at a quarter to three p.m., while driving around the Dijon ring road, he has not yet decided where to go; he is thinking it over because, for several reasons, all these cities attract him in exactly the same way. In this situation, the following propositions are arguably untrue if evaluated at a quarter to three p.m.:

(1) Leo is going to Besançon.
(2) Leo is going to Metz.
(3) Leo is going to Paris.

Since Leo has not yet decided where to go, none of these propositions is privileged with respect to the others. Suppose, however, that Leo ultimately decides to go to Besançon, where he arrives two hours later. Moreover, suppose that traffic police have photographed Leo on the Dijon ring road at a quarter to three p.m. The day after, someone asks, "What was Leo doing yesterday, at a quarter to three p.m., when the picture was taken?" Intuitively, the following answer seems to be correct:

(4) Leo was going to Besançon.

The analogy with OC is obvious: while from the perspective of July 14th at a quarter to three p.m., it is not true that Leo was going to Besançon at that time, from the perspective of July 15th, it was true that Leo was going to Besançon on July 14th at a quarter to three p.m. The progressive, moreover, has an obvious connection with the future. According to the standard semantics of the progressive (Landman 1992; Portner 1998), Prog(e) is true if a first

2 The correctness or incorrectness of past beliefs about the future are soft facts—that is, past facts that depend on future facts. For the distinction between soft and hard facts, cf. Hoffman and Rosenkrantz (1984), Todd (2013), and de Florio and Frigerio (2018).

stage of event *e* has occurred so far, and if all will go on normally without any unexpected interruptions, *e* will be completed in the future. Thus, our intuitions about the truth values of (1) to (4) provide further support for OC.

2 OC and Omniscience

Our intuitions in favour of OC are very strong; this means that our formal semantics should validate it. Otherwise, we would be in the presence of a conceptual tension; although the principle is intuitively valid, our logic fails to characterise it. Indeed, many of the major semantic frameworks for branching time account for OC. Specifically, it is valid in Thomason's supervaluationism (cf. Thomason 1970) because $p \rightarrow \mathbf{PF}p$ holds in all histories, and supervaluationism universally quantifies on all histories. It is also valid in MacFarlane's relativism (cf. MacFarlane 2003, 2014), which indeed seems to be designed to account for the intuition that $\mathbf{F}p$ is untrue with respect to yesterday's assessment context but true with respect to today's assessment context. However, OC is not valid in Peircean or Thin Red Line (TRL) semantics—in Peircean semantics, OF is valid, but RCP is not, whereas in TRL semantics, neither OF nor RCP are valid[3]—and this might be a problem for these semantics.

Todd and Rabern (2021) are not convinced by OC. For this reason, they construct an ingenious argument against it. Their strategy is indirect: if OC is supposed to be valid, then counterintuitive consequences regarding the logic of divine omniscience follow. In other words, the validity of OC is incompatible with the existence of an omniscient and temporal entity. Todd and Rabern are not committed to the actual existence of an omniscient entity, but—and this is their point of argument—it is very odd that a question concerning the semantics of future statements has metaphysical consequences for the existence of omniscient entities. Our intuition suggests that the conditions for the possibility of omniscient supernatural entities should not be related to the structure of the future and RCP. However, if Todd and Rabern are right, and if OC is accepted, there can be no omniscient temporal entity. Since this is a metaphysical thesis, it follows that OC is a suspicious doctrine. Incidentally, this is a point in favour of semantics that do not validate OC. In particular, if

3 For the invalidity of RCP in TRL semantics, cf. Belnap and Green (1994). It is possible to amend the TRL semantics in order to validate RCP. For instance, the TRL can be relativized to instants of time (cf. McKim and Davis 1976; Øhrstrøm 2009), or the TRL can be initialized at the post-semantic level rather than at the semantic level (cf. Iacona 2014; Wawer 2014; Wawer and Malpass 2020). We will ignore these issues here.

doi: 10.48106/dial.v77.i3.05

OF is supposed, this constitutes a point in favour of Peirceanism, which does not validate RCP.

Therefore, let us examine Todd and Rabern's argument in detail. We may, for ease of presentation, sometimes refer to the omniscient entity as "God"; we also assume a doxastic operator of belief (B), which we intend to always refer to the omniscient subject, whereby $\mathrm{B}p$ indicates that God believes p. Todd and Rabern establish the principle of omni-accuracy as follows:

OMNI-ACCURACY. $p \leftrightarrow \mathrm{B}p$[4]

The OMNI-ACCURACY principle thus establishes a double implication between p and God's belief that p: every divine belief is true, and there is no truth that is not believed by God. Using OMNI-ACCURACY and RCP, we obtain the following:

(5) $p \rightarrow \mathbf{P}_1\mathrm{B}(\mathbf{F}_1 p)$[5]

As a result, if it is true that it is raining today, then yesterday, God believed it would rain today. However, suppose yesterday it was not determined whether it would rain today. Then yesterday, God could not believe that today it would rain. In fact, yesterday $\mathbf{F}_1 p$ was untrue because today's weather was indeterminate, and, therefore, God could not believe it.

Advocates of OC have two possibilities at this point:

(i) They can claim that yesterday it was not true that God believed $\mathbf{F}_1 p$, but that today it is true that God believed it. This is equivalent to stating that the past changes with the passage of time. This does not seem

4 An aspect of OMNI-ACCURACY seems to be problematic; if p is true, then it is true that God believes p; analogously, in the case in which it is false (and therefore $\neg p$ is true), we have that God does not believe p. If p is indeterminate, how does the principle behave? One might answer that it is indeterminate whether God believes p, but this seems strange; if p is an indeterminate proposition, then God should not believe it. However, the principle remains silent on this point. It could be argued that it is for this reason that Todd and Rabern introduce the principle of OMNI-CORRECTNESS (see below in the main text). However, this move does not seem to solve the problem of the indeterminacy of divine belief. Suppose that the truth value of p is undetermined. It would, therefore, be the case that $\mathrm{T}(p)$ is also indeterminate (because of the Tarskian T-schema). However, through the OMNI-CORRECTNESS principle, we obtain that $\mathrm{B}(p)$ is indeterminate, and, therefore, the point raised against OMNI-ACCURACY is reiterated.

5 $\mathbf{F}_1$ and $\mathbf{P}_1$ are metric temporal operators. Their semantics is straightforward: fixing the day as the unit of time, $\mathbf{P}_1 p$ means that yesterday, it was true that p; analogously, $\mathbf{F}_1 p$ means that tomorrow, it will be true that p.

acceptable. As mentioned earlier, we can assume that the correctness of beliefs changes over time because correctness concerns the relationships between beliefs and states of affairs in the world. Since changing these relationships does not imply any intrinsic change in beliefs, they do not appear to have any impact on the fixity of the past. However, the change that seems required here does not concern the correctness of beliefs. Rather, what is required here is an intrinsic change of the past: while at time t_1 it is true that God did not believe at t_0 that it would rain (where $t_0 < t_1$), at a later time t_2 it is true that God believed at t_0 that it would rain. This intrinsic change in the past is clearly in conflict with the fixity of the past.

(ii) They can deny Omni-Accuracy. This amounts to denying the very *possibility* of an omniscient being. However, it seems strange that a future semantic theory could imply the denial of such a possibility. As the authors observe, "In general, one could argue that a semantic theory—a theory concerned with the logic and compositional structure of the language—ought not to settle certain substantive non-semantic questions" (Todd and Rabern 2021, 116).

Therefore, it seems that OC must be denied. If we accept OF, RCP must be denied. Todd and Rabern discuss another possible defence by advocates of RCP. Instead of assuming Omni-Accuracy, the defender of RCP could assume Omni-correctness:

Omni-Correctness. $\mathsf{T}p \leftrightarrow \mathrm{B}(p)$

where T is the truth predicate; in other words, an omniscient being believes that p iff p is true. They could then insist that yesterday, it was not true that $\mathbf{F}_1 p$. In other words, they could argue that today, $p \rightarrow \mathbf{P}_1\mathbf{F}_1 p$ holds, but $p \rightarrow \mathbf{P}_1\mathsf{T}\mathbf{F}_1 p$ does not hold. Since it was not true that it would rain yesterday, the principle of Omni-Correctness is not violated. This seems to reconcile OC with the possibility of an omniscient being.[6]

Todd and Rabern do not find this solution convincing. In fact, they find it strange that it could be said that yesterday it was the case that it would rain today but that yesterday it was not true that it would rain today. In their view,

6 Notice that this notion of truth is not the one defended by Thomason (1970), who proposed a completely transparent treatment of T: $t \vDash \mathsf{T}p$ iff $t \vDash p$.

doi: 10.48106/dial.v77.i3.05

> Very plausibly, if one is moved by the backward-looking intuition that, given that a sea-battle has occurred, it was always going to occur, it seems that one should likewise be moved by the intuition that given that a sea-battle has occurred, it was always true—which is not to say determined!—that it was going to occur. (Todd and Rabern 2021, 114)[7]

Todd and Rabern also reject MacFarlane's relativism. According to MacFarlane, it is necessary to evaluate propositions on the basis of both the context of utterance and the context of assessment: a statement such as "It will rain tomorrow" uttered on December 9th has December 9th as its context of utterance. However, it can have different contexts for assessment. When evaluated with respect to December 9th, it is neither true nor false, but when evaluated with respect to December 10th, it is true or false. Todd and Rabern state that apart from the technicalities with which this relativistic intuition is implemented, it remains true that "insofar as the Open-closurist view has a notion of truth that vindicates the (updated) Retro-closure principle, they will have to accept the conclusion that *God was genuinely ignorant*. Something was *true* (in the relevant sense) that God didn't believe" (2021, 115). From the point of view of December 10th, it was true on December 9th that it would rain the following day, and since God did not believe on December 9th that it would rain on December 10th, there was something true that God did not believe.

We believe that the two-dimensional semantics we propose in this paper—inspired by MacFarlane's intuitions—has all the resources to demonstrate that from OC, it does not follow that God failed to know something true. It is possible to assume OC and still not deny the possibility of an omniscient being. We will show this in section 4. In the next section, we introduce the branching time semantics and illustrate the difficulties of traditional supervaluationism in responding to Todd and Rabern's objection.

3 Branching Time, Supervaluationism, and RCP

As we have seen above, Todd and Rabern are convinced that there is no way out for the OC adherent who accepts the possibility of the existence of an omniscient entity. Indeed, as we shall see, supervaluationists (*à la* Thomason)

7 Todd and Rabern do not explicitly mention the principle that would be abandoned if one embarked on this strategy. This would amount to denying the Tarskian T-scheme for which φ iff $\mathsf{T}\varphi$. In our opinion, this move has unsustainable theoretical costs.

can formally accept the two conditions without contradiction. However, the theoretical cost that they have to pay is high; for this reason, we will develop an alternative semantics to supervaluationism that is able to account for our intuitions about omniscience and that validates OC. First, we present the key ingredients of a branching time semantics; then, we reconstruct the argument from a supervaluationist perspective and show that, although not inconsistent, the supervaluationist is nevertheless forced to accept very strange conclusions.

3.1 *Branching Time*

A branching time structure[8] is a couple consisting of a non-empty set of time instants and an order relation defined on them: $\mathcal{B} = \langle \mathbb{T}, < \rangle$. Intuitively, the instants are possible instantaneous states of the world, and $<$ is the relation of temporal precedence. This relation is, therefore, asymmetric and transitive and satisfies (at least) the conditions of Backward Linearity (BL) and Historical Connectedness (HC).

(BL) $\forall t \forall t_1 \forall t_2((t_1 < t \wedge t_2 < t) \rightarrow (t_1 = t_2 \vee t_1 < t_2 \vee t_2 < t_1))$

In words, two instants of the past of t are either identical or ordered by $<$; this implies that for every instant t, there is one and only one past history.

(HC) $\forall t_1 \forall t_2 \exists t(t \leq t_1 \wedge t \leq t_2)$

HC asserts that all the instants are connected in the past.

Maximal subsets of instants linearly ordered in t are referred to as histories (h)—the possible courses of events around the world. Ours is a propositional language that includes a possible infinite set of propositional variables (Var) and two temporal operators, **P** and **F**. It is useful, as we will see in a moment, to exploit metric temporal operators, such as $\mathbf{P}_n$ and $\mathbf{F}_n$. As indicated earlier, $\mathbf{P}_n$ means "n units of time before the instant of evaluation," and $\mathbf{F}_n$ means "n units of time after the instant of evaluation."

Now, let us see how to define formula evaluations in our semantics. Here, we will use an Ockhamist framework in which a formula φ is evaluated with respect to a time t and a history h. For the formal feature of the order relation among instants, any instant in the structure has only one past history but

8 For a classical presentation of branching time, see Belnap, Perloff and Xu (2001). Readers familiar with this literature can go directly to section 3.2.

doi: 10.48106/dial.v77.i3.05

one or more future histories, depending on whether there is branching in the future of that instant.

The use of histories becomes crucial in the case of the evaluation of statements in the future. Suppose there are two histories branching off from instant t_0, namely, h_1 and h_2. In h_1, certain things happen, and therefore certain formulas are true, while in h_2, things go differently and, consequently, other formulas are true. How can we interpret the proposition "It will rain" ($\mathbf{F}\varphi$) evaluated at t_0 if in one history it rains the following day and in the other history it does not? A very natural solution might be to relativise truth conditions to histories, as in Ockhamist semantics. Therefore, today, it is true that in the future, it will rain in history, say, h_1, while it is false that in the future, it will rain in history h_2:

$$\begin{array}{lll} \mathcal{M}, t/h_1 \vDash^{\text{ock}} \mathbf{F}\varphi & \text{iff} & \exists t' > t, \mathcal{M}, t'/h_1 \vDash^{\text{ock}} \varphi \\ \mathcal{M}, t/h_1 \nvDash^{\text{ock}} \mathbf{F}\varphi & \text{iff} & \neg\exists t' > t, \mathcal{M}, t'/h_1 \vDash^{\text{ock}} \varphi \end{array}$$

On the Ockhamist semantics, only if a future branch is specified can a truth value be ascribed to a formula. The intuitive problem with this theory consists in the fact that there is no trace of such specification of possible branches in our everyday talk about the future, which we would like to model (Wawer 2014, 366).

In the literature, there are two large families of answers to this problem. The first, the *Open Futurist* semantics, denies that future contingents can be true. Most Open Futurist semantics state that the truth value of a future tense statement depends (in a sense to be specified) on what happens in all the histories that stem from a certain instant. In the second family of views, on the contrary, future contingents can be true. Usually, those who embrace this thesis assume *linearist* semantics—that is, they believe that there is somehow a privileged history and that the truth conditions of a future tense statement concern only what happens in that history.

As is widely known, Open Futurists can be roughly divided into *Peirceans* and *Aristotelians*: for the former, greatly simplifying, future contingents are all false since a future tense statement is true if and only if it is true in all future histories, and by definition, a future contingent is true in some future histories and false in others. Aristotelians, on the other hand, maintain that future contingents are neither true nor false.

It is not the purpose of this paper to precisely characterise the options on the table nor to argue in favour of one solution or another (for this, we refer,

among many others, to Thomason 1970; Todd 2021). Our aim here is only to demonstrate that Open Futurism and RCP are not in conflict with the possibility of an omniscient being.

3.2 *Supervaluationism and* RCP

Now, let us reconstruct Todd and Rabern's argument from a supervaluationist standpoint.

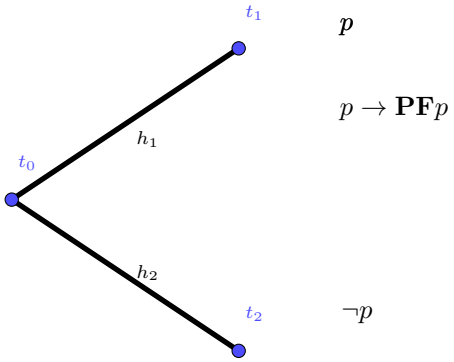

Figure 1: caption

Supervaluationism distinguishes truth from super-truth. Formulas are evaluated with respect to instant/history pairs; they are then supervaluated with respect to instants only. The idea can be described as follows: Formula φ is super-true at t if and only if φ is linearly true in every history that passes through t; formula φ is super-false at t if and only if φ is linearly false in every history that passes through t; finally, φ is indeterminate if and only if it is neither super-true nor super-false.

Figure 1 represents a fork made up of two histories (h_1 and h_2). Within the first, p is true; within the second, it is not true. Since p is true at t_1, in

doi: 10.48106/dial.v77.i3.05

accordance with RCP, it was true at t_0 that p would be true because, for every history that passes through t_1 (that is, h_1), $\mathbf{PF}p$ is true at t_1. Now, $t_0 \not\models^{\text{sup}} \mathbf{F}p$ holds because future contingents are never super-true. By OMNI-ACCURACY and logic, we have that $t_0 \not\models^{\text{sup}} \mathrm{B}(\mathbf{F}p)$—that is, it is not (super)true that God believes at t_0 that it will rain.[9] Therefore, supervaluationists who endorse OMNI-ACCURACY are forced to accept both of the following semantic statements:

(i) $t_0 \not\models^{\text{sup}} \mathrm{B}(\mathbf{F}p)$
(ii) $t_0/h_1 \models^{\text{ock}} \mathrm{B}(\mathbf{F}p)$

Although (i) and (ii) are not contradictory and do not violate OMNI-ACCURACY, this is a weird situation for supervaluationists. They must claim that at a given instant of time, it is not (super)true that God believes that tomorrow it will rain, but it is true with respect to some history that God believes that it will rain tomorrow. It is not straightforward how the advocate of this view could account for these results in a coherent, general picture.

In addition, there is a more general problem concerning the adequacy of supervaluationist semantics in characterising epistemic predicates, such as belief. It is well known that supervaluationism has been developed to rigorously account for the semantics of vague predicates (cf., for instance, Fine 1975). Now, it is not clear how this could apply to belief operator B. If a subject is *agnostic* about p, then it seems natural to claim not only that it is not true that she believes p (and that she believes $\neg p$) but also that it is *false* that she believes p. However, this entails that $\neg\mathrm{B}(p)$ is true. The export of negation from the metatheory to the theory—that is, the passage from the metatheoretical general claim $\mathcal{M} \not\models \varphi$ to $\mathcal{M} \models \neg\varphi$—is invalid within the supervaluationist semantics: from the fact that it is not (super)true that there will be a sea-battle tomorrow, it does not follow that it is (super)true that there will be no sea-battle. However, for the doxastic operator B, things look different. Supposing that Emma is agnostic about the presence of beer in the fridge, the supervaluationist semantics would force the following reading: it is not (super)true that Emma believes that there is any beer in the fridge, but, from that, it does not follow that it is false that Emma believes that there is any beer in the fridge. Given our pre-theoretical stances about the semantics

9 MacFarlane (forthcoming) correctly notices that supervaluationism and OMNI-ACCURACY imply $t_0 \not\models^{\text{sup}} \mathrm{B}(\mathbf{F}p)$ and not $t_0 \models^{\text{sup}} \neg\mathrm{B}(\mathbf{F}p)$, as Todd and Rabern seem to suppose. However, we find supervaluationism still problematic for the reasons given in the main text.

of the verb *believe*, this seems to be rather odd. To solve this problem, perhaps supervaluationists might postulate that if φ is indeterminate, then $\neg \mathrm{B}(\varphi)$ is true.

There is a more serious problem for supervaluationists: the question arises as to the truth value of formula $\mathbf{P}(\mathrm{B}(\mathbf{F}p))$ evaluated at t_1. Supervaluationists have two possibilities: either $t_1 \nvDash^{\mathrm{sup}} \mathbf{P}(\mathrm{B}(\mathbf{F}p))$ or $t_1 \vDash^{\mathrm{sup}} \mathbf{P}(\mathrm{B}(\mathbf{F}p))$. However, both seem to be problematic. If the former holds, then a fact of the past (i.e., a divine belief) is indeterminate. This is against the intuition according to which past facts are metaphysically determined. Therefore, the second possibility must hold. However, in this case, Todd and Rabern's objection seems cogent: yesterday, it was indeterminate (or false) that God believed $\mathbf{F}p$ since $\mathbf{F}p$ is a future contingent, lacking a (super)truth value, but today, it is determinate that yesterday God believed $\mathbf{F}p$. Therefore, endorsing this view is quite implausible because it implies an intrinsic change in past facts.[10] Therefore, we agree with Todd and Rabern's claim that supervaluationism is in trouble with OC and the possibility of the existence of an omniscient entity (Todd and Rabern 2021, 110–111). Now, the following question arises: Is it possible to develop an OC semantics compatible with the existence of an omniscient entity?

4 Retro-Believing and Retro-Truth

4.1 *Double-Indices Semantics*

In the following, we develop an OC double-indices semantics. Our system is, in a sense, inspired by intuitions surrounding MacFarlane's relativist semantics; however, unlike MacFarlane's system, our semantics considers both

10 MacFarlane (forthcoming) claims that Todd and Rabern's argument presupposes a substantive metaphysical claim: past and present beliefs are settled. However, it is difficult to see why past and present beliefs should not be fixed as any other past or present fact. MacFarlane appeals to Jackman (1999) to support the idea that past and present beliefs are not settled. However, Jackman believes that a past belief is not settled when it involves indeterminate meanings that are determined over time. Future uses ultimately determine past uses of a word. By Jackman's own admission, these cases might be quite rare. However, this would not be the case with divine past beliefs about future contingents, which should always be determined by what happens in the future, even when dealing with wholly determinate meanings.

doi: 10.48106/dial.v77.i3.05

MacFarlane's contexts of evaluation and of assessment as semantic indices.[11] A formula is evaluated at a particular time and with respect to a perspective; the perspective indicates the present time within the structure. We call this framework *perspectival semantics*. From a formal point of view, this means that a formula is evaluated with respect to an instant and any history that is included between that instant and the particular instant that is the present, or the "now." Therefore, we evaluate the truth value of a formula at a certain instant *when* another instant has the property of being now. In other words, the second index is the perspective from which we "see" the structure.[12] This parameter is essential, as we will see soon, since it cuts off the histories against which a formula is evaluated. In fact, the advancement of the world determines the future; time flows, and the bundle of possible available histories is reduced. Today (t), it is indeterminate whether it will rain or not tomorrow, but tomorrow, when day $t + 1$ will be now, the weather will no longer be indeterminate.

To clarify, let us take the above example: today, December 10th, it rains. Assuming that today's rain is a contingent feature of the world, is it indeterminate on December 9th whether it will rain the next day? The answer would be that it depends on the *perspective* from which we locate ourselves. If we place the now on December 9th or at an instant preceding December 9th, then the rain of the following day is an indeterminate event. The world has arrived—so to speak—at just a certain point, and the future is open. But if we place the now at an instant following December 9th, such as December 10th, the world has been determined, and some histories are no longer available; in particular, the possibility of a December 10th with no rain has expired. Therefore, *from the perspective of December 10th*, it is determined on December 9th that it will rain the following day.

Why adopt a two-indices semantics? The grounding idea is that the advancement of the present prunes certain histories and leaves others open. This feature is crucial when we evaluate the truth value of sentences that are future-tensed but evaluated at a *past* instant. Since time has passed, some

11 We use this semantics since we believe that it makes our argument clearer. However, we assume that our argument could be formulated in any semantics or post-semantics that involves two evaluation indices.

12 One might wonder what the intended interpretation of perspectival semantics is and whether it favours an A-theoretic construal. Although we have argued elsewhere that perspectival semantics is an adequate framework for advocates of the A-theory, it is nevertheless wholly compatible with a purely indexical reading of the now.

histories are no longer available; they were open *before*, but not now because things went a certain way. As said before, our framework explicitly ffavours no specific metaphysics of time; consequently, we do not take a stance towards the phenomenon of pruning. But let us consider that our linguistic practices often refer both to available and no longer available histories. In other terms, we evaluate the formula ψ at an instant t from a perspective t'. Sometimes, the instant of evaluation is connected to the perspective, that is, it lies in its past or future; sometimes not. In the first case, we have a *factual* situation; in the second, a *counterfactual* one. The two-indices semantics seems to be a promising conceptual tool in order to characterise these scenarios.[13]

From a formal point of view, our semantics accounts for this possibility by using two temporal indices: the first is the evaluation instant, and the second is the present. Therefore, let us consider the following expression:

(a) $\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \varphi$

(a) should be read as follows: φ is true at t_i when the present is t_j. The former index (t_i) refers to the instant at which the formula is evaluated. The latter index (t_j) refers to the position of the present within the structure.

The two instants can coincide. This would be a case in which we would evaluate, say, φ at t when the world has arrived at t. In any case, the evaluation instant (e.g., t) and the present (e.g., t') must be *connected*; in other words, one of the following conditions must hold: $t < t'$ or $t \approx t'$ or $t' < t$.[14] The truth clause for atomic formulas is as follows:

(b) $\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \varphi \quad \Leftrightarrow \quad \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), t_i/h \vDash^{\text{ock}} \varphi$

where $\mathcal{H}_{t_i}$ is the set of histories that pass through the instant t_i ($\mathcal{H}_{t_i} = \{h | t_i \in h\}$); therefore, $(\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j})$ is the intersection between the two sets of histories. Here, we assume that the satisfiability operator is not bivalent in perspective semantics. Therefore, we have the following:

(c) $\mathcal{M}, t_i, t_j \nvDash^{\text{prs}} \varphi \quad \Leftrightarrow \quad \neg\forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), t_i/h \vDash^{\text{ock}} \varphi$

13 For a more extensive defence and articulated exposition of this two-indices framework, cf. de Florio and Frigerio (2020).

14 A perspectival semantics in which the instant of evaluation and the now need not be connected has been exploited to provide a counterfactual semantics in de Florio and Frigerio (2020). We thank you, an anonymous referee, for this point.

doi: 10.48106/dial.v77.i3.05

(d) $\mathcal{M}, t_i, t_j \dashv^{\text{prs}} \varphi \quad \Leftrightarrow \quad \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), t_i/h \nvDash^{\text{ock}} \varphi$

The following truth clauses are straightforward:

$$\begin{array}{lll}
\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \neg\varphi & \Leftrightarrow & \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), t_i/h \nvDash^{\text{ock}} \varphi \\
\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \varphi \wedge \psi & \Leftrightarrow & \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), t_i/h \vDash^{\text{ock}} \varphi \\
 & & \text{and } \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), t_i/h \vDash^{\text{ock}} \psi \\
\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \mathbf{P}\varphi & \Leftrightarrow & \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), \exists t' < t_i, t'/h \vDash^{\text{ock}} \varphi
\end{array}$$

Clauses without temporal operators are a natural extension of the Ockhamist linearist evaluation. As for the past case, the second index is vacuous. In the past of the instant of evaluation, there is just one history, and, therefore, the evaluation is linear. Things become more interesting in the future case because the second index plays a significant role.

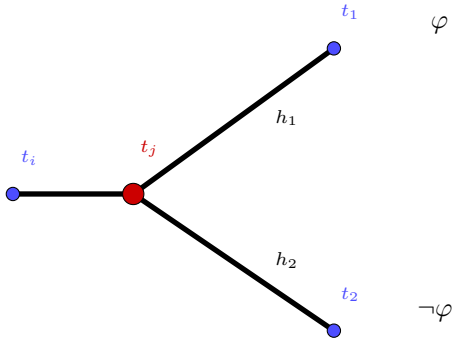

Figure 2: caption

In figure 2, the instant of evaluation is t_i, while the now is at t_j; the truth conditions of $\mathbf{F}\varphi$ at t_i are the following:

$$\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \mathbf{F}\varphi \quad \Leftrightarrow \quad \forall h \in (\mathcal{H}_{t_i} \cap \mathcal{H}_{t_j}), \exists t' > t_i, t'/h \vDash^{\text{ock}} \varphi$$

We now have two cases (see the figures below): the case in which the present coincides (or precedes) the instant of evaluation, and the case in which the present follows that instant:

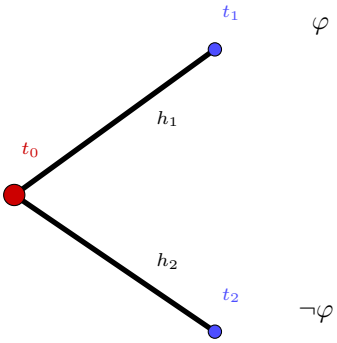

Figure 3: I Case

(I Case). In this case, the now (t_0) coincides with the point of evaluation. Since $\mathcal{H}_{t_0} \cap \mathcal{H}_{t_0} = \mathcal{H}_{t_0}$, both branches (the φ-branch and the $\neg\varphi$-branch) are available. Therefore, $\mathcal{M}, t_0, t_0 \nvDash^{\text{prs}} \mathbf{F}\varphi$.

doi: 10.48106/dial.v77.i3.05

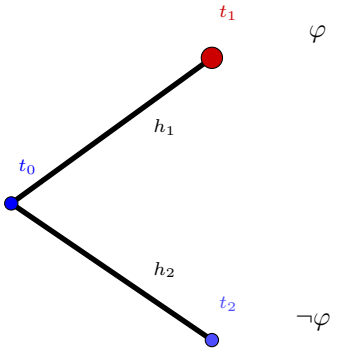

Figure 4: II Case

(II Case). Here, the now is at t_1 and follows the instant of evaluation (t_0). The history in which $\neg\phi$ is true is no longer available because $h_2 \notin \mathcal{H}_{t_0} \cap \mathcal{H}_{t_1}$. Consequently, $\mathcal{M}, t_0, t_1 \vDash^{prs} \mathbf{F}\varphi$.

In the double-indices framework, the truth values of propositions change with the flow of time. This accounts for RCP and the intuitions reviewed in section 1.[15]

4.2 *The Possibility of Omniscience*

Is our perspectival semantics able to answer Todd and Rabern's challenge? We think so, and in order to show this, we must elaborate on a further important point. The belief operator is usually characterised through Kripkean semantics using doxastically possible worlds—namely, possible cognitive representations of the subject. It is debatable whether this approach is ade-

15 Our two-dimensional semantics has some similarity with the two-dimensional possibility framework proposed by Cariani (2024), even though in a rather different context. In his framework, too, the evaluations of formulas depend on two different possibilities, which correspond to moments in branching time semantics.

quate to model divine beliefs; in fact, God has just one representation of the world—the correct one. This fact is correctly grasped by Todd and Rabern's OMNI-ACCURACY principle. However, there is another interesting point here. The concept of belief is clearly representational: a subject believes that p in that she has a representation of the (actual) world as a world in which p is true. This holds true both for human subjects and for omniscient entities. Believing that tomorrow p will be true has to do with the representation of the future actual world.

These reflections may seem to be trivial, but they are of a certain interest if we take into account the double-indices analysis we provided. Since the concept of belief is representational, it is natural to *centre* it on the now, where we represent the present structure of the world. This does not mean that the belief cannot concern the future or the past; we can believe that it will rain tomorrow or that it rained yesterday, but we do so from the actual (present) perspective, whereby if Emma believes (now) that it will rain (tomorrow), this means that it is from the perspective of today that Emma believes that it will rain tomorrow. Analogously, if Emma believed yesterday that it would rain today, it is from the perspective of yesterday that Emma represented the following day.

This peculiar feature could be formally presented through a particular semantic clause about the belief operator B; in a nutshell, when one construes, at instant t and from the perspective t', a formula in which the B operator occurs, then one has to reformulate that interpretation, *moving* the now to the instant of evaluation of the formula. The justification for this manoeuvre follows from the fact that the belief operator must be fixed to the now. In other words, we represent the world based on the information available at the moment of the representation.

This means that we cannot rightly construe a past doxastic state if we locate the epistemic subject at a time different from that at which she represents the world. When we look back and wonder what we believed in the past, we must *backdate* the now, bringing back the perspective to the one that is contemporaneous with the instant of evaluation.

Formally, all of this can be characterised through the following *belief semantic norm*:

(bsn) $\mathcal{M}, t_i, t_j \vDash^{\text{prs}} \mathsf{B}(\varphi) \Rightarrow \mathcal{M}, t_i, t_i \vDash^{\text{prs}} \mathsf{B}(\varphi)$

doi: 10.48106/dial.v77.i3.05

where φ can have any logical complexity. A formula that describes a belief attitude of any complexity must be evaluated from a perspective that coincides with the instant evaluation of the formula. Therefore, (bsn) moves the index of the perspective from t_j to t_i.[16]

This point is important, and it deserves some clarification.[17] (bsn) is a principle that is not derived from other semantical axioms. On the contrary, it is assumed as a formal counterpart of a conceptual reflection about the very notion of belief. Let us consider an ascription of a doxastic state, such as $\mathrm{B}(x, \psi)$, whose intended meaning is: the doxastic agent x believes that ψ is true. Well, ψ could have any complexity; in particular, it can be a future or past tense sentence. So far, so good. But let us also consider x, that is, the doxastic agent. In a temporal framework, x is located (so to speak) somewhere; she believes something at a given time. Then, the perspective from which to evaluate $\mathrm{B}(x, \psi)$ must be centred on the actual temporal position of the doxastic agent.

As a consequence, the OMNI-ACCURACY principle $\varphi \Leftrightarrow \mathrm{B}(\varphi)$ becomes the following:

(Omn-prs) For every $t, \mathcal{M}, t, t \vDash^{\mathrm{prs}} \varphi \leftrightarrow \mathrm{B}(\varphi)$

For every perspective, God believes what is true at the instant of that perspective. For example, if $\mathrm{now}(t_0)$, God believes every proposition that is true at t_0 from the perspective of t_0. Notice that since φ can have any logical complexity, it can contain any number of temporal operators. Therefore, at t_0 from the perspective of t_0, God believes what is true at the other points of the structure from the perspective of t_0. For instance, if φ is true at t_{-1} when $\mathrm{now}(t_0)$, then God believes at t_0 from the perspective of t_0 that $\mathbf{P}_1\varphi$. In symbols:

$$\mathcal{M}, t_0, t_0 \vDash^{\mathrm{prs}} \mathbf{P}_1\varphi \leftrightarrow \mathrm{B}(\mathbf{P}_1\varphi)$$

Therefore, God believes at any instant what is true at any point in the structure from the perspective of that instant.[18]

16 In the following, we also exploit the (bsn) principle for cases of dis-beliefs; this is reasonable, however, since they are representational attitudes toward the world. The idea is that if a proposition is untrue (perhaps because its truth value is indeterminate), then an omniscient entity does not believe it is true.

17 We want to thank an anonymous referee for having prompted these considerations.

18 As previously noted, the OMNI-ACCURACY principle is silent about cases in which the truth value of the formula at play is indeterminate. We think it is plausible and in agreement with our intuitions about an omniscient entity (but, in general, this should be valid for any epistemic

The fundamental ingredients of our semantics can be summarised as follows: propositions are true at an instant of evaluation and with respect to a perspective. This entails that, in some cases, the truth value of a proposition, evaluated at instant t_i, is indeterminate from a certain perspective, while from another perspective, it is true. Divine beliefs always track the truth from the perspective of the evaluation because they are representations of the world at a certain instant from the perspective of that instant.

Let us now see how our perspectival semantics, enhanced with (bsn), is able to answer Todd and Rabern's objection. First, let us resume, in a semi-formal fashion, Todd and Rabern's argument. For convenience, we use metric temporal operators. Consider figure 5.[19] Since $t_1 \vDash \varphi$ and since, by hypothesis, RCP holds, we have that $t_1 \vDash \mathbf{P}_1\mathbf{F}_1\varphi$. However, $\mathbf{F}_1\varphi$ is not true at t_0 because it is future contingent. It follows that $t_0 \vDash \neg \mathrm{B}(\mathbf{F}_1\varphi)$ (cf. footnote 12) because God does not believe what is untrue, but since it is true at t_1 that at t_0 φ would be true the following day, God should have believed that. Instead, $t_1 \vDash \mathbf{P}_1(\mathbf{F}_1\varphi \wedge \neg \mathrm{B}(\mathbf{F}_1\varphi))$. In other words, let us locate ourselves at t_1 (when φ is true). In this case, two things were true yesterday: on the one hand, φ would be true the next day, and on the other hand, God did not believe that φ would be true the next day. However, this means that God is ignorant about the future truth of φ.

It is easy to realise that, within our framework, we get two theoretically interesting results. First, the framework allows us to distinguish the case of the "genuine" future from the case of the retrospective future (i.e., the future in the past). At the same time, Todd and Rabern's argument is no longer reproducible. Let us see why.

---

subject) that the following condition must be satisfied: if a proposition is indeterminate with respect to its truth value, then the omniscient entity does not believe it (as true). We translate this indeterminacy situation via the following meta-theoretical conditional:

(ind-bel) If $\mathcal{M}, t_0, t_0 \nvDash^{\mathrm{prs}} \varphi$, then $\mathcal{M}, t_0, t_0 \vDash^{\mathrm{prs}} \neg \mathrm{B}\varphi$

(Omn-prs) and (ind-bel) describe, therefore, the relationships between propositions and their truth values from one side and God's beliefs on the other side.

19 Todd and Rabern never specify the semantic system in which they carry out their arguments. For this reason, we assume that their satisfaction relation ($\vDash$) involves a quantification on times.

doi: 10.48106/dial.v77.i3.05

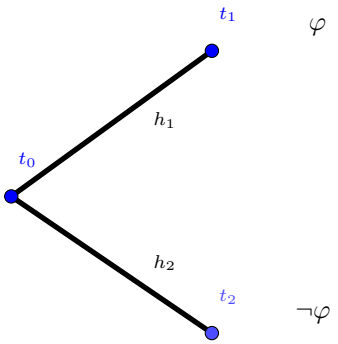

Figure 5: Figure 3

From $\mathcal{M}, t_1, t_1 \vDash^{\text{prs}} \mathbf{P}_1(\mathbf{F}_1\varphi \wedge \neg \mathrm{B}(\mathbf{F}_1\varphi))$, it follows that $\mathcal{M}, t_1, t_1 \vDash^{\text{prs}} \mathbf{P}_1\mathbf{F}_1\varphi \wedge \mathbf{P}_1\neg \mathrm{B}(\mathbf{F}_1\varphi)$ and then that $\mathcal{M}, t_1, t_1 \vDash^{\text{prs}} \mathbf{P}_1\mathbf{F}_1\varphi$. We also have that $\mathcal{M}, t_1, t_1 \vDash^{\text{prs}} \mathbf{P}_1\neg \mathrm{B}(\mathbf{F}_1\varphi)$, whose meaning is: from today's perspective, it was true that yesterday, God did not believe that φ would be true. From this, it follows that $\mathcal{M}, t_0, t_1 \vDash^{\text{prs}} \neg \mathrm{B}(\mathbf{F}_1\varphi)$—that is, again, from today's perspective, yesterday, God did not believe that φ would be true. Is this sufficient to claim that God was ignorant? No. God appears to be ignorant only because we see the world from a metaphysically privileged standpoint, and we look at the world from the perspective of how things actually happened.

If we really want to locate ourselves at t_0, we must also backdate the now; we must, in other terms, rewind the tape of history. This is precisely what (bsn) does: $\mathcal{M}, t_0, t_1 \vDash^{\text{prs}} \neg \mathrm{B}(\mathbf{F}_1\varphi)$ becomes $\mathcal{M}, t_0, t_0 \vDash^{\text{prs}} \neg \mathrm{B}(\mathbf{F}_1\varphi)$. Thus, when the present was yesterday, God did not believe that φ would happen. However, $\mathcal{M}, t_0, t_0 \nvDash^{\text{prs}} \mathbf{F}_1\varphi$ also holds because $\mathbf{F}_1\varphi$ is a future contingent. Therefore, assuming the present version of OF, φ is neither true nor false at t_0. However,

the fact that an omniscient entity does not believe what is not (yet) true is not a problem for that entity's omniscience.[20]

Todd and Rabern's argument hinges on a theoretical passage that retrogrades the truth and charges an omniscient being with the alleged failure to grasp that truth. However, it is precisely a retrograded truth: it is a truth only because the world has moved forward, and what was indeterminate is now determinate. If we place ourselves at the temporal perspective of the omniscient entity at t_0 (i.e., if we move the now to t_0), we naturally get that the omniscient entity does not believe that φ would be the case since, from t_0's perspective, it is not true that φ will be the case.

At the same time, from t_1's perspective, it was true that φ would be true: RCP entails $\mathcal{M}, t_0, t_1 \vDash^{\text{prs}} \mathbf{F}_1\varphi$. However, this is not a problem for the possibility of omniscience because Todd and Rabern suppose that the belief of the omniscient entity occurs when the present is t_0, not when it is t_1.

In addition, we have that $\mathcal{M}, t_1, t_1 \vDash^{\text{prs}} \mathbf{P}_1\mathbf{F}_1\varphi \wedge \mathrm{B}(\mathbf{P}_1\mathbf{F}_1\varphi)$, from which it follows that $\mathcal{M}, t_1, t_1 \vDash^{\text{prs}} \mathrm{B}(\mathbf{P}_1\mathbf{F}_1\varphi)$. In other words, from today's perspective, God believes that yesterday, it was true that it would rain today. Actually, as we have seen, it is true from today's perspective that yesterday, it was true that it would rain today.

To sum up, when $\mathrm{now}(t_0)$, God does not believe that $\mathbf{F}_1\varphi$ because she sees the world from t_0's perspective, from which $\mathbf{F}_1\varphi$ is untrue. When $\mathrm{now}(t_1)$, God believes that $\mathbf{F}_1\varphi$ was true the previous day because she sees the world from t_1's perspective, from which the previous day it was true that φ would be true the next day. Therefore, God's beliefs always track the truth. Time flows, and with its advancing, the truth values of propositions change; an omniscient God always believes at a time t what is true from the perspective of that time. Put differently, God's beliefs are changing in the same way that the truth values of propositions change due to the flowing of time. However, this is what is reasonable to demand of an omniscient entity.

Obviously, our argument can be opposed by objecting (bsn). However, this does not seem to be a plausible objection if we assume that God is temporal, as Todd and Rabern do. Such a God is temporally located, and He, therefore,

20 An anonymous referee suggests that Todd and Rabern's argument shows the incompatibility between the Retro-closure principle and the idea of *permanent omniscience*. The idea is as follows: in our framework, we have cases in which an omniscient being realises—so to speak—that in the past, there were true propositions not known. However, this does not happen in our framework since the omniscient being realises that in the past, He did not believe φ, but, nevertheless, φ was indeterminate from yesterday's perspective. Only from today's perspective, φ was true.

doi: 10.48106/dial.v77.i3.05

knows the world from the perspective of the present. If we embrace the Open theism view, His beliefs evolve when tracking the evolution of the world. If this concept of God is assumed, it is quite natural to endorse (bsn): God is an entity located within the present who, therefore, knows the world from the present's perspective, like human beings.

Of course, we could have different theistic views. We could hypothesise that God is timeless and that He knows the evolution of the world from an eternal perspective. Within these views, (bsn) can be safely discharged. However, if these views are presupposed, Todd and Rabern's argument is no longer valid: being timeless, God neither remembers nor anticipates anything. Rather, He sees the whole unfolding of the world from His eternal standpoint, and thus, He knows all that happens at every instant. If God is timeless, it is meaningless to wonder whether $t_0 \vDash^{\mathrm{prs}} \mathbf{P}(\mathrm{B}(\mathbf{F}p))$ holds since the belief operator B cannot be within the scope of the temporal operators and, in general, cannot be evaluated with respect to an instant.[21]

Therefore, either one assumes that God is temporal, and then Todd and Rabern's argument does not succeed since it is reasonable to claim that (bsn) holds, or one assumes that God is timeless, and then (bsn) does not hold. In the latter case, however, Todd and Rabern's argument cannot even be formulated since their argument presupposes a God located within time.

5 Conclusions

In our semantic framework, there is no instant at which a formula is true, and the omniscient entity does not believe it to be true. The theoretical cost we have to pay is the acceptability of (bsn); specifically, the principle according to which the ascription of belief to a subject at an instant is constrained by the state of the universe at that time and cannot be legitimately forward dated. We think this is a highly affordable cost based on a reasonable theoretical proposal. We conclude that Todd and Rabern's argument fails to show that

21 It is reasonable to require that a timeless God knows the truth values of propositions relatively at every instant of time and every temporal perspective, that is, from any "now." For this omni-perspectival view of God's knowledge, see de Florio and Frigerio (2019, chap. 6). This view seems to be naturally connected with a B-theoretic metaphysics of time, where all the "presents" exist on a par. If one wants to keep together a timeless God and an A-theoretic metaphysics, one needs to appeal to non-standard A-theories, like Fragmentalism. On this, again, see de Florio and Frigerio (2019, chap. 6).

OF is incompatible with the possibility of an omniscient entity. OF, therefore, remains a viable alternative in the tense semantics market.*

Ciro De Florio
0000-0002-5503-7463
Università Cattolica of Milan, Italy
ciro.deflorio@unicatt.it

Aldo Frigerio
0000-0003-4698-2110
Università Cattolica of Milan, Italy
aldo.frigerio@unicatt.it

References

Belnap, Nuel D., Jr. and Green, Mitchell S. 1994. "Indeterminism and the Thin Red Line." in *Philosophical Perspectives 8: Logic and Language*, edited by James E. Tomberlin, pp. 365–388. Oxford: Basil Blackwell Publishers, doi:10.2307/2214178.

Belnap, Nuel D., Jr., Perloff, Michael N. and Xu, Ming. 2001. *Facing the Future: Agents and Choices in Our Indeterminist World*. Oxford: Oxford University Press, doi:10.1093/oso/9780195138788.001.0001.

Bonomi, Andrea. 1997. "The Progressive and the Structure of Events." *Journal of Semantics* 14(2): 173–205, doi:10.1093/jos/14.2.173.

Cariani, Fabrizio. 2024. "Modeling Future Indeterminacy in Possibility Semantics." Unpublished manuscript, draft of March 21, 2024, https://philpapers.org/archive/CARMFI.pdf.

de Florio, Ciro and Frigerio, Aldo. 2018. "Two Kinds of Soft Facts." *Grazer Philosophische Studien* 95(1): 34–53, doi:10.1163/18756735-000031.

—. 2019. *Divine Omniscience and Human Free Will. A Logical and Metaphysical Analysis*. London: Palgrave Macmillan, doi:10.1007/978-3-030-31300-5.

—. 2020. "The Thin Red Line, Molinism, and the Flow of Time." *Journal of Logic, Language, and Information* 29(3): 307–329, doi:10.1007/s10849-019-09304-4.

Fine, Kit. 1975. "Vagueness, Truth, and Logic." *Synthese* 30(3-4): 265–300, doi:10.1007/bf00485047.

Hoffman, Joshua and Rosenkrantz, Gary S. 1984. "Hard and Soft Facts." *The Philosophical Review* 93(3): 419–434, doi:10.2307/2184544.

* The authors gratefully thank two anonymous referees from *Dialectica* for their constructive comments and recommendations, which definitely helped to improve a previous version of this paper.

IACONA, Andrea. 2014. "Ockhamism Without Thin Red Lines." *Synthese* 191(12): 2633–2652, doi:10.1007/s11229-014-0405-6.
JACKMAN, Henry. 1999. "We Live Forwards but Understand Backwards: Linguistic Practices and Future Behavior." *Pacific Philosophical Quarterly* 80(2): 157–177, doi:10.1111/1468-0114.00078.
LANDMAN, Fred. 1992. "The Progressive." *Natural Language Semantics* 1(1): 1–32, doi:10.1007/BF02342615.
MACFARLANE, John. 2003. "Future Contingents and Relative Truth." *The Philosophical Quarterly* 53(212): 321–336, doi:10.1111/1467-9213.00315.
—. 2014. *Assessment Sensitivity: Relative Truth and Its Applications*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199682751.001.0001.
—. forthcoming. "Why Future Contingents Are Not All False." *Analytic Philosophy*, doi:10.1111/phib.12334.
MCKIM, Vaughan R. and DAVIS, Charles C. 1976. "Temporal Modalities and the Future." *Notre Dame Journal of Formal Logic* 17(2): 233–238, doi:10.1305/ndjfl/1093887526.
ØHRSTRØM, Peter. 2009. "In Defence of the Thin Red Line: A Case for Ockhamism." *Humana.Mente* 3(8): 17–32, https://www.humanamente.eu/index.php/HM/issue/view/8.
PORTNER, Paul H. 1998. "The Progressive in Modal Semantics." *Language* 74(4): 760–787, doi:10.2307/417002.
THOMASON, Richmond H. 1970. "Indeterminist Time and Truth-Value Gaps." *Theoria* 36(3): 246–281, doi:10.1111/j.1755-2567.1970.tb00427.x.
TODD, Patrick. 2013. "Soft Facts and Ontological Dependence." *Philosophical Studies* 164(3): 829–844, doi:10.1007/s11098-012-9917-4.
—. 2021. *The Open Future: Why Future Contingents are All False*. Oxford: Oxford University Press, doi:10.1093/oso/9780192897916.001.0001.
TODD, Patrick and RABERN, Brian. 2021. "Future Contingents and the Logic of Temporal Omniscience." *Noûs* 55(1): 102–127, doi:10.1111/nous.12294.
WAWER, Jacek. 2014. "The Truth About the Future." *Erkenntnis* 79(suppl., 3): 365–401, doi:10.1007/s10670-013-9454-3.
WAWER, Jacek and MALPASS, Alex. 2020. "Back to the Actual Future." *Synthese* 197(5): 2193–2213, doi:10.1007/s11229-018-1802-z.

Review of Schofield (2021)

Agnès Sophia Constance Baehni

Schofield, Paul. 2021. *Duty to Self: Moral, Political, and Legal Self-Relation*. Oxford: Oxford University Press.

Paul Schofield's *Duty to Self* (2021) is an excellent contribution to recent moral philosophy. It is a much-needed addition to a literature that has, up until now, largely ignored the possibility of reflexive moral relationships. Thorough and challenging, the book is an indispensable read for students and scholars with an interest in ethics, metaethics, and political philosophy. In this review, I outline what I perceive to be the book's main contributions and discuss some areas of concern about Schofield's innovative framework. It should be noted from the outset that Schofield does not claim to tell us what we owe to ourselves but rather seeks to establish the possibility of moral duties to the self. As he insists: "The project is most concerned with the metaphysics of morals, rather than the content of morality" (p. 18).

In its traditional 'social' understanding, the term 'moral' is applied to our relationships with others and is only rarely applied to our relationships with ourselves. It is often taken for granted that the normative dimension of the self-relationship, how one should act towards oneself, is rather to be understood in terms of prudence. Schofield questions this assumption by focusing on moral duties to oneself. As we shall see, he is the first to offer a compelling way out of the objection raised by Marcus Singer in the early 1960s, according to which the notion of moral duties to oneself is incoherent. In a nutshell, Singer's (1959, 1963) argument is that we cannot be bound by a duty if we are at any time free to release ourselves from it. We typically think of duties as binding because the power to release us from them lies with another individual. Since we are always free to release ourselves from what we owe to ourselves, Singer argues, there can be no genuine duties to the self. This is the 'waivability objection'.

The first chapter, "On the Significance of Duties to the Self," seeks to elucidate whether duties to the self can be moral obligations. Drawing on Stephen Darwall's (2006) influential insights, Schofield argues that having a genuine

moral duty to ourselves entails that we adopt a second-personal stance towards ourselves. So, the main issue is "whether a person can have duties second-personally directed to herself, of the kind had by a person standing on another's foot" (p. 26). Singer's waivability objection is based on the idea that we cannot relate morally to ourselves in this way. While using Singer's objection as his target, Schofield sets aside the possibility of conceiving our moral relationships to ourselves on a monadic model of morality, like virtue ethics (p. 23).[1] This is where his approach is most ambitious: it aims to understand our moral relationships to ourselves and others on a unified model.

A first way to reply to the waivability objection is introduced in chapter 3 ("Defending Duties to the Self Part 1: Duties Across Time"). Schofield's idea, roughly, is that we can understand our relationships to ourselves second-personally by making use of the notion of temporal perspective. To illustrate this idea, Schofield discusses the case of a smoker. Let us call him Alan. Why should we think that Alan owes it to himself to quit smoking? Schofield's answer is that Alan can adopt a second-personal stance towards himself by paying attention to the legitimate demands that might be issued from one of his temporal perspectives. We can think, for instance, of his perspective in 20 years' time. When so doing, Alan may be confronted with a conflict between his present and his future interests, i.e., the interests of his future perspective. Alan does not want to quit smoking now since he finds it enjoyable and is not suffering (yet) from any side effects. By contrast, judging from his future perspective, he should quit smoking, as continuing to smoke will be detrimental to his health in the long run. As Schofield explains, "these cross-temporal conflicts between various of a person's interests enable us to tell a story about *intra*personal generation of duties that parallels in its essentials an *inter*personal story" (p. 67).

The waivability objection does not threaten duties owed to others because the power to release us from our obligations lies in another's hand. Understanding duties to ourselves in terms of obligations towards future perspectives allows reconsidering the idea that we can always release ourselves from such duties: we cannot when the power to do so lies with another temporal perspective. A question still remains: Does a person knowing that she will die soon from an incurable disease have no duty to herself? To analyze duties to ourselves as duties over time may not do justice to all these duties.

1 As Schofield observes, in virtue ethics, each virtue "supplies a standard for good action under which all persons are evaluable, without necessarily putting the subjects into normative contact with others" (p. 23).

This limitation is addressed in chapter 4 ("Defending Duties to the Self Part 2: Duties at a Moment"), which seeks to establish the existence of duties to ourselves *at a time*. When considering duties at a time, Schofield introduces the notion of a *practical perspective*. The idea that we can occupy different practical perspectives stems from the observation that we all wear different hats: we are philosophers, parents, sportsmen and women, chess players, etc. Sometimes, our duties as philosophers conflict with the duties of another of our practical perspectives; for example, we wonder whether we should go to the chess club or grade philosophy papers. As Schofield puts it, "When an individual addresses another, a person might address *herself* from the perspective of one of her practical identities, issuing demands that will be received from a perspective associated with a different practical identity" (p. 107). Again, the conflict between different perspectives' interests helps explain how we can have moral duties to ourselves at a time. We cannot release ourselves from our duty to go to the chess club because the power to do so lies with a different practical perspective, that of a chess player.

Schofield's proposal is promising, but it also raises some concerns having to do with the notion of a person. Quoting Schofield:

> It has been my aim to give an account of duties [...] while maintaining that the person herself is the locus of moral value. But one might doubt whether I've succeeded in this. Talk of personal identities and of second-personal interactions between them will suggest to some readers a picture on which multiple "selves" constitute a person at a moment. (p. 124)

Indeed, the idea that we adopt several practical and temporal perspectives in our practical deliberations intimates a kind of division that may threaten the unity presupposed by the idea of a person: "We're thus left to worry: What if the cost of gaining a second person within is losing the person altogether?" (p. 206).

Schofield goes on to answer this worry by insisting that the adoption of the second-personal standpoint only requires the capacity to occupy different practical or temporal perspectives. His proposal is not premised on the claims that one is composed of different selves interacting with one another at a time or that "person-stages, or time-slices" (p. 169) relate to one another over time. Perspectives are epistemic stances, not metaphysical entities, and it is always the person who relates second-personally to herself through the

doi: 10.48106/dial.v77.i3.06

adoption of different perspectives. Schofield rightly emphasizes here a point of tension in his proposal, having to do with the notion of perspective. We might fear that by focusing on the task of telling us what these perspectives are not, Schofield leaves us wondering about what they exactly are and how we can access them. So, while his proposal is the most articulate response to the waivability objection to date, the key notion of a perspective remains somewhat elusive.

I suggested that some people may not have the kind of distant temporal perspectives required to ground duties across time. One may also fear that the appeal to practical perspectives is insufficient to explain why we have duties to ourselves at a time. Consider, for instance, a genius artist who is fulfilled by committing herself exclusively to her art or a monk who is perfectly content with his life of devotion. Arguably, these individuals do not have other practical perspectives, and there is thus no conflict of interest. Granted that they exist, can Schofield's proposal acknowledge that people who are completely coherent, unified, or focused on the present are in a moral relationship with themselves? At first sight, the answer is "no" since his view presupposes the capacity to look at ourselves from a distance, or to see ourselves as someone else, so to speak. The same seems to be true of our moral relationships with other people: I have a moral duty to be nice to my friends even though I *want* to be nice to them anyway. We can have moral duties to the self and to others without conflicting interests.

A second worry is that Schofield's focus on practical and temporal perspectives might be blurring the nature of the issue at stake. There might be alternative ways to gloss the necessary second-personal stance than by reference to temporal and practical perspectives. For instance, we sometimes regard ourselves second-personally when considering that what we did was morally wrong. The duty of self-respect may not be explained in terms of demands issued by one of our practical or temporal perspectives but only by reference to a moral perspective.

This brings me to my final point. Given its reliance on the notion of perspective, Schofield's proposal may not have the resources to explain two prima facie central and interdependent aspects of the reflexive moral relationships: its relations to the aims of being happy and of becoming the best version of ourselves. Indeed, to which practical or temporal perspective do we owe the pursuit of happiness and of ethical or intellectual development? Given that these goods are desirable from any perspective, reference to perspectives seems here both unnecessary and insufficient in explaining why we have such

duties to ourselves.[2] This is where Schofield's model of moral self-relationship meets its limits. As I observed, Schofield is explicit that he does not seek to establish precisely what we owe to ourselves. Nevertheless, it seems reasonable to expect the metaphysics of the moral self-relationship to be consistent with how we intuitively conceive of the content of this relationship.

Chapters 5 and 6 of the book shift from the moral to the political sphere. Having shown how we can relate morally to ourselves, Schofield goes on to show that we can also relate politically to ourselves. In chapter 5 ("Might There Be Self-Directed Political Duties? Troubles for State Paternalism"), he lays out compelling reasons for thinking that we have no self-regarding political duties before arguing, in chapter 6 ("Defending Political Duties to the Self: The Possibility of Liberal Paternalism"), that we in fact have self-directed duties of right and justice and that paternalism is among the state's functions. His aim, ultimately, is to challenge "a conception on which politics is for others exclusively" (p. 140).

In the seventh and final chapter ("Practical Philosophy After Duties to Self"), Schofield draws some important implications of his proposal, among which the fact that it should not be taken "as a mere addendum to whatever ethical theory is already in place" (p. 195). On the contrary, he insists that two difficult and hitherto ignored questions now arise. First, what to do when what we morally owe to ourselves conflicts with the moral interests of others? Second, what should we do when our own interests conflict with one another? These questions have been largely ignored so far, and practical philosophy will certainly be busy in the coming years trying to answer them. We should be grateful to Schofield for his stimulating proposal, which provides a stable foundation for future debates on these issues.

Agnès Sophia Constance Baehni
0000-0002-7605-9346
Université de Genève
Agnes.Baehni@unige.ch

References

Darwall, Stephen L. 2006. *The Second-Person Standpoint. Morality, Respect and Accountability*. Cambridge, Massachusetts: Harvard University Press.

2 Yuliya Kanygina makes a similar point in Kanygina (2022).

doi: 10.48106/dial.v77.i3.06

KANYGINA, Yuliya. 2022. "Duties to Onself and Their Alleged Incoherence." *Australasian Journal of Philosophy* 100(3): 565–579, doi:10.1080/00048402.2021.1909086.
SCHOFIELD, Paul. 2021. *Duty to Self: Moral, Political, and Legal Self-Relation*. Oxford: Oxford University Press, doi:10.1093/oso/9780190941758.001.0001.
SINGER, Marcus George. 1959. "On Duties to Oneself." *Ethics* 69(3): 202–205, doi:10.1086/291210.
—. 1963. "Duties and Duties to Oneself." *Ethics* 73(2): 133–142, doi:10.1086/291439.

Review of Lepine (2023)

Steve Humbert-Droz

Lepine, Samuel. 2023. *La nature des émotions. Une introduction partisane*. Analyse et philosophie. Paris: Librairie philosophique Jean Vrin.

We are now in the age of affectivism (Dukes et al. 2021): while emotions have long been contrasted with cognition, they are now seen as a central element of our rational life. Lepine (2023) joins this paradigm, arguing that emotions are cognitive states, a source of axiological knowledge, and even an essential component of values.

Lepine's original contribution consists of an extremely cautious and impressive interweaving of psychological and philosophical discussions of emotions as well as of values. We may take from *La nature des émotions* a set of entangled statements: (i) emotions are *cognitive states* distinguished by their *evaluative* nature; (ii) they are sui generis psychological modes that *focus our attention* and prepare our body for action; (iii) they are evaluative since we can ascribe a *correspondence* between the emotion and the value instantiated by the emotion's intentional object (i.e., correctness conditions); (iv) the correctness and justification conditions of emotions *partially depend on the background motivations* on which every emotion is based because (v) values depend on some non-evaluative properties of external objects as well as on the *agents' motivations.* These different points fit together to form the most comprehensive introduction to emotions I've read since Deonna and Teroni's *The Emotions* (2012). Let us examine how.

Chapter (1) outlines the so-called "naive features" of emotions: automaticity, valence, intentionality, direction of fit, cognitive and motivational bases, etc. Chapter (2) focuses on the opposition between emotion and cognition. The notion of cognition is discussed with reference to the debate between Zajonc (1984) and Lebens and Folkman (1984). Lepine (2023) endorses the appraisal theory of emotions in psychology (Lazarus' view), arguing that (i) emotions are *cognitive states* insofar as stimuli processing in emotions makes them available for semantic processing (p. 61), the *evaluative nature* of the processing being the mark of emotions in cognition (pp. 59–60).

Chapters (3) and (4) explore the evaluative nature of emotions. Lepine first wonders whether emotions constitute a natural kind. Chapter (3) sets out Griffiths's (1997) well-known objections to the idea that emotions share essential properties, coupled with considerations coming from constructivist theories (e.g., Russell 2003). Lepine concludes, in line with the appraisal theory, that "emotions would indeed constitute a natural kind insofar as each of them shares the function of detecting a specific core relational theme" (p. 87; I translate all quotations).[1]

Chapter (4) then explores philosophical theories of emotions to understand further the notion of a *core relational theme*. According to Lepine, this notion is analogous to what philosophers call "formal objects" (pp. 108–109). A dog, a steep cliff, losing my money in the stock market, and my Ph.D. supervisor are objects I mention when answering the question, What are you afraid of?—they are the *intentional objects* of my fear. What do they have in common (when my fear is appropriate)? They all instantiate the same evaluative property, being dangerous. Danger is, thus, the *formal object* of fear. How are formal objects connected to emotions? Lepine follows Deonna and Teroni's (2012, 2015, 2024) attitudinal view of emotions, according to which (ii) emotions are sui generis *psychological modes* (i.e., they are reducible neither to judging nor to perceiving, and so on)[2] constituted by unified bodily feelings that prepare the subject for action—e.g., fear prepares me to flee. Most importantly, (iii) the content of emotions need not be evaluative; the evaluative nature of emotions lies in the *fittingness relation* between fear, shame, pride…, and the evaluative properties instantiated by the intentional object of these emotions. In other words, the content of my emotion doesn't need to go beyond a non-evaluative representation of the intentional object—"the dog," "the steep cliff," etc.

Lepine illustrates this relationship by interpreting the attitudinal theory through Cummins's (1996) analysis of psychological attitudes. Attitudes are characterized by their cognitive function;[3] the attitude sets a target and processes its content in a way that is correct when the target is reached:

1 We may regret that Lepine does not raise as vigorous objections to appraisal theory as he does to other ones. For example, it is unclear whether appraisals cause or constitute emotions (Moors 2013; Roseman and Smith 2001).

2 Chapter (4) also discusses perceptual, judgmental, mixed views, etc. These discussions are close to those of Deonna and Teroni (2012, chaps. 5–6). It is unfortunate that some recent approaches (e.g., Mitchell 2021; Müller 2019) are not covered.

3 Here, the term "function" refers to a representational function à la Dretske. In this sense, emotions can be said to represent values. Yet the attitudinal view argues that values do not feature in the *content* of emotions.

> My emotion of fear implies an evaluation of the dog as dangerous: I apply to the content of my representation (the dog) a target (dangerousness). And this emotion is correct if it targets a state of affairs that fits with the mobilized representation, i.e., if dangerousness (the target) is indeed one of the properties of the represented dog (the content). (p. 134)

A refinement of the attitudinal theory is then developed to circumvent the common objection that emotions are not reducible to bodily feelings (Goldie 2000; Mitchell 2021). Lepine insists on the way emotions *direct our attention* on aspects of the intentional object and considers that, at least in the case of mild emotions, we need to focus on this rather than on bodily feelings (p. 139). He argues that attention is a necessary component as well, and sometimes more salient than the bodily component (p. 141, see also Deonna and Teroni 2015).

This elegant (and plausible!) refinement makes sense of the idea that bodily feelings are directed toward the external world—since they accompany and are calibrated by an attentional mechanism. In addition, just as Brady (2013), Lepine can explain how emotions, while not representing values in their content, lead us to focus on natural properties in the world that constitute the basis for an understanding of values (p. 143).

Chapter (5) focuses on the correctness conditions of emotions and, therefore, on what it means for an emotion to fit a value. In a new and stimulating way, Lepine contrasts an *independentist* view with a *motivational* view of correctness. According to the first view—attributed to Tappolet (2016) and D'Arms and Jacobson (2000)—correctness is determined solely by the natural properties instantiated by the intentional object of the emotion. For instance, my sense of humor is irrelevant to determining the correctness of my amusement at Julie's joke; the only thing we need to consider is whether the joke is objectively funny. Lepine rejects this analysis and suggests that (iv) the correctness of emotions also depends on their *congruence with our "background motivations"*—i.e., desires, preferences, feelings, character traits…—as well as the coherence of these motivations, as we shall see later (p. 192).

There are psychological and axiological elements in Lepine's view. On the psychological side, motivations are considered as a necessary base of emotions (p. 163, see also Baier 2004; Roberts 2003). On the axiological side, Lepine adopts (v) a form of (neo-)*sentimentalism* according to which values are understood in terms of appropriate emotions (p. 151)—e.g., injustice is what

doi: 10.48106/dial.v77.i3.07

deserves indignation (Brentano 1902). Thus, if appropriate emotions underlie values, and if emotions are based on motivations, then (by transitivity) motivations underlie values (at least partially). This leads to a *subjectivist view* of values: it is only when *x*'s natural properties are prized by humans (or sentient creatures) that *x* deserves pro-attitude A and thus is good.

Note that Lepine considers that motivations are themselves subject to correctness conditions: motivations are considered correct until proven otherwise, i.e., until they contradict our experience, higher-order beliefs, or (social, prudential, moral…) norms (p. 180). It is thus difficult to possess racist motivations without encountering contradictions (p. 182). Motivationalism is thus a naturalist view on value that occupies a middle ground between "raw" subjectivism and naïve realism.

Finally, in chapter (6), Lepine argues that the justification of emotions is also impacted by motivations. The independentist argues that emotions are justified when we can mention natural properties on which the corresponding value may supervene ("Why am I afraid of that dog? Did you see its teeth, its bloodshot eyes, its lowered tail?!"). The motivationalist replies that subjective motivations also play a justificatory role—as we shall see later.

In the very last section of *La nature des émotions*, Lepine attempts to demonstrate that emotions, despite their subjective nature, are a trustworthy tool for evaluative judgment (p. 224). Contrary to the idea that emotions are prone to many "false positives" (see Goffin 2023), Lepine argues that even when we "confabulate" to justify our (inappropriate) emotions, these confabulations are relatively plausible (p. 218) and should not prevent us from trusting our emotions most of the time (p. 224).

Considering the author's clear view on emotions and values, I may suggest only a few challenges aimed at extending the discussion he proposed.

Among these challenges, we might mention that Lepine does not substantiate his parallel between core relational themes and formal objects. Yet, as Teroni (2023) points out, psychologists have a hard time reconciling the core relational theme or "molar value" (such as injustice, dangerousness, sublime…) and the "molecular values" targeted by each appraisal check (such as relevance, urgency, power…). Since psychologists tend to subjectivize molecular values, this might confer an advantage to the motivational view, as long as the shifts between molecular value, molar value, and, finally, formal object are conceptually possible.

Another challenge concerns both psychological and axiological motivationalism. The two aspects seem inseparable in Lepine's mind, and this, in my

opinion, implies that adopting motivationalism carries a lot of presuppositions. For instance, we find very little argument against the realist approaches to values except the evocation of one aspect of the queerness of values (Mackie 1977): isn't it odd that an "objective" property of the world (value) has motivating or normative power (pp. 169–170)? The motivational view dissolves this issue, but it is far from being the only solution (see Enoch 2011), and it comes with costs and concessions.

Consider the cost at the psychological level. A direct and acknowledged consequence of motivationalism is that there can be no such thing as emotional discoveries (p. 167, p. 173). Just as Sartre (1940) claimed that we can never discover anything with imagination other than what we've put into it, Lepine maintains that we never acquire new motivations by feeling an emotion. This is questionable. Consider Pablo being forced to attend opera—a musical genre he has no motivation to listen to. However, this time, he is touched; from now on, he is willing to come back every month and add opera playlists on Spotify. According to Lepine, Pablo's emotion necessarily arises from a pre-existing motivational basis. At a certain degree of generality, this is indisputable: maybe Pablo has a preference or a desire for music (in general), beauty, or pleasant moments. Yet, this seems to miss the point raised by scholars acknowledging the possibility of emotional discoveries. If we work with a fine-grained notion of motivation, we seem to acquire new specific interests or re-evaluate (quite radically) states of affairs thanks to our emotions. The only replies available to Lepine are either to assume that emotions that are not based on pre-existing motivations are inappropriate (see p. 185, p. 222) or that the relevant specific motivations are unconscious (see p. 188). This is unfortunate considering that the latter is painfully ad hoc, whereas the former clashes with cases of "outlaw emotions" that seem fitting even though they contrast with our personality (Silva 2021). Now, if Lepine rejects the fine-grained approach, one might ask why congruence with background motivations should count as correctness conditions (p. 192): a condition that cannot be incorrect is incongruent with the common understanding of correctness conditions.

At the axiological level, the motivational approach is convincing when we consider personal values (see Rønnow-Rasmussen 2007). For instance, my disappointment at a friend's betrayal may be justified by my twenty-year attachment to that friend (see Bell 2011). It is so because betrayal is the kind of value that depends on a relationship; it cannot be instantiated between two strangers. But when we consider impersonal values, motivationalism loses its panache. Epistemic values, for example, hardly seem to depend on

doi: 10.48106/dial.v77.i3.07

our motivations. Invoking subjective motivation to justify an epistemic emotion comes across as odd—my astonishment at discovering a mathematical demonstration cannot be justified by my passion for Pythagoras. This provides a psychological explanation of the emotion (someone who does not share my passion for Pythagoras would not have been astonished), not a *justification*, which depends, e.g., on whether there are reasons to think that the proof is sound and noteworthy.

Lepine is aware that motivationalism blurs the justification/explanation contrast (p. 200) and might reply that impersonal values are based on interests shared by all mankind (see p. 174). I disagree because I think that impersonal values render the world better for human beings, even when they currently have no motivation to promote them. In the world depicted by the movie *Idiocracy*—where the planet's most foolish inhabitants have reproduced in large numbers to the point of engendering a society with no culture or historical knowledge—nobody is motivated to acquire knowledge. I would not say, however, that knowledge has no value in this world. People are just wrong! You may say that they *should* be motivated by knowledge. Then, if knowledge possesses value not because people have motivations but because knowledge deserves to motivate them, we lose the motivationalist view on the way and go back to pure (neo-)sentimentalism.

Samuel Lepine's monograph is subtitled "une introduction partisane": it is introductory in the noblest and most exciting sense of the word; it offers an overview of affective topics in philosophy and psychology without detracting from the precision and complexity of the debates. Written in crystal-clear French (guarantee without any trace of Sorbonnian style!), *La nature des émotions* results in a conceptually plausible and empirically supported defense of the appraisal theory, the attitudinal view (re-visited), and the motivational view (introduced here).

Steve Humbert-Droz
0000-0002-0680-649X
Université de Genève
Swiss Center for Affective Sciences, Geneva
steve.humbert.droz@gmail.com

References

Baier, Annette C. 2004. "Feelings That Matter." in *Thinking about Feeling. Contemporary Philosophers on Emotion*, edited by Robert C. Solomon, pp. 200–213. Series in Affective Science. Oxford: Oxford University Press, doi:10.1093/oso/9780195153170.003.0014.

Bell, Macalester. 2011. "Globalist Attitudes and the Fittingness Objection." *The Philosophical Quarterly* 61(244): 449–472, doi:10.1111/j.1467-9213.2010.684.x.

Brady, Michael Sean. 2013. *Emotional Insight. The Epistemic Role of Emotional Experience*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199685523.001.0001.

Brentano, Franz. 1889. *Vom Ursprung sittlicher Erkenntnis*. Leipzig: Duncker & Humblot. Second edition: Brentano (1921).

—. 1902. *The Origin of the Knowledge of Right and Wrong*. Westminster: Archibald Constable & Co., Ltd. Translation of Brentano (1889) by Cecil Hague.

—. 1921. *Vom Ursprung sittlicher Erkenntnis*. 2nd ed. Hamburg: Felix Meiner Verlag. Second edition of Brentano (1889), edited by Oskar Kraus.

Cummins, Robert. 1996. *Representations, Targets, and Attitudes*. Cambridge, Massachusetts: The MIT Press.

D'Arms, Justin and Jacobson, Daniel. 2000. "The Moralistic Fallacy: On the 'Appropriateness' of Emotions." *Philosophy and Phenomenological Research* 61(1): 65–90, doi:10.2307/2653403.

Deonna, Julien Amos and Teroni, Fabrice. 2008. *Qu'est-ce qu'une émotion?* Chemins Philosophiques. Paris: Librairie philosophique Jean Vrin.

—. 2012. *The Emotions. A Philosophical Introduction*. London: Routledge. Substantially enlarged edition and translation of Deonna and Teroni (2008).

—. 2015. "Emotions as Attitudes." *Dialectica* 69(3): 293–311. Special issue "Beyond Perceptualism," edited by Sabine A. Döring and Anika Lutz, doi:10.1111/1746-8361.12116.

—. 2024. "Emotions and Their Correctness Conditions: A Defense of Attitudinalism." *Erkenntnis* 89(1): 45–64, doi:10.1007/s10670-022-00522-0.

Dukes, Daniel, Adams, Kathryn, Adolphs, Ralph, Ahmed, Mohammed E. and others, many. 2021. "The Rise of Affectivism." *Nature Human Behaviour* 5: 816–820, doi:10.1038/s41562-021-01130-8.

Enoch, David. 2011. *Taking Morality Seriously. A Defense of Robust Realism*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199579969.001.0001.

Goffin, Kris. 2023. "Better Scared than Sorry: The Pragmatic Account of Emotional Representation." *Erkenntnis* 88(6): 2633–2650, doi:10.1007/s10670-021-00470-1.

Goldie, Peter. 2000. *The Emotions: A Philosophical Exploration*. Oxford: Oxford University Press.

doi: 10.48106/dial.v77.i3.07

GRIFFITHS, Paul E. 1997. *What Emotions Really Are. The Problem of Psychological Categories*. Chicago, Illinois: University of Chicago Press.

LEBENS, Samuel R. and FOLKMAN, Susan. 1984. *Stress, Appraisal, and Coping*. New York: Springer Verlag.

LEPINE, Samuel. 2023. *La nature des émotions. Une introduction partisane*. Analyse et philosophie. Paris: Librairie philosophique Jean Vrin.

MACKIE, John Leslie. 1977. *Ethics: Inventing Right and Wrong*. London: Penguin Books.

MITCHELL, Jonathan. 2021. *Emotion as Feeling towards Value: A Theory of Emotional Experience*. Oxford: Oxford University Press, doi:10.1093/oso/9780192846013.001.0001.

MOORS, Agnes. 2013. "On the Causal Role of Appraisal in Emotion." *Emotion Review* 5(2): 132–140, doi:10.1177/1754073912463.

MÜLLER, Jean Moritz. 2019. *The World-Directedness of Emotional Feeling. On Affect and Intentionality*. London: Palgrave Macmillan, doi:10.1007/978-3-030-23820-9.

ROBERTS, Robert C. 2003. *Emotions. An Essay in Aid of Moral Psychology*. Cambridge: Cambridge University Press, doi:10.1017/CBO9780511610202.

RØNNOW-RASMUSSEN, Toni. 2007. "Analysing Personal Value." *The Journal of Ethics* 11(4): 405–435, doi:10.1007/s10892-006-9012-1.

ROSEMAN, Ira J. and SMITH, Craig A. 2001. "Appraisal Theory: Overview, Assumptions, Varieties, Controversies." in *Appraisal Processes in Emotion. Theory, Methods, Research*, edited by Klaus R. SCHERER, Angela SCHORR, and Tom JOHNSTONE, pp. 3–19. Oxford: Oxford University Press, doi:10.1093/oso/9780195130072.003.0001.

RUSSELL, James A. 2003. "Core Affect and the Psychological Construction of Emotion." *Psychological Review* 110(1): 145–172, doi:10.1037/0033-295X.110.1.145.

SARTRE, Jean Paul. 1940. *L'Imaginaire. Psychologie phénoménologique de l'imagination*. Paris: Gallimard.

SILVA, Laura. 2021. "The Epistemic Role of Outlaw Emotions." *Ergo* 8(23): 664–691, doi:10.3998/ergo.1160.

TAPPOLET, Christine. 2016. *Emotions, Values, and Agency*. Oxford: Oxford University Press, doi:10.1093/acprof:oso/9780199696512.001.0001.

TERONI, Fabrice. 2023. "Evaluative Theories in Psychology and Philosophy of Emotion." *Mind and Language* 38(1): 81–97, doi:10.1111/mila.12374.

ZAJONC, Robert B. 1984. "On the Primacy of Affect." *American Psychologist* 39(2): 117–123, doi:10.1037/0003-066X.39.2.117.

Review of Landgrebe and Smith (2022)

Jonathan Simon

Landgrebe, Jobst and Smith, Barry. 2022. *Why Machines Will Never Rule the World. Artificial Intelligence Without Fear*. London: Routledge.

In this formidable book, Jobst Landgrebe and Barry Smith argue that no AI will ever attain human-level intelligence. The book is a challenging read, but it is full of important insights, its master argument is original, it is informed by an impressive array of sources, and it is timely. It merits philosophical attention. The book is also noteworthy because it is the collaboration of an engineer and a philosopher. Landgrebe runs an AI software company: he has an M.D. and a Ph.D. in biochemistry (although notably, his grandfather Ludwig was a famous phenomenologist). Smith is an expert on the Austrian phenomenological tradition and in formal and applied ontology (where, notably, he has pioneered a way to apply philosophical theory and method to data engineering). This kind of collaboration is vital but hard to achieve.

In very broad strokes, the book's master argument is this: Human-level intelligence requires coping with and getting on in environments that are complex dynamical systems—that is, environments that are open and chaotic and subject to feedback effects, with trends and statistics that change over time (think: the weather or the stock market). Data models of such complex dynamical systems are always mere approximations not good enough to enable long-term prediction in a complex world in constant flux (think: why you don't know whether it will rain next Saturday and why you can't reliably beat the stock market). But AI systems are just data models. So, in principle, they can't enable the sort of coping that humans are capable of. What about full-on emulations of the human neuro-cognitive system? That system, too, is a complex dynamical system, so no ensemble of algorithms based on data models can approximate it well enough for full-on emulation (in Landgrebe and Smith's terminology: well enough for a model that is both "adequate" and "synoptic"—where this means a model that enables predictions that are accurate enough for the task at hand).

Here is how they proceed. Chapters 1–6 present their general picture of intelligence and mindedness, as well as language and sociality, and why they think that having these capabilities entails that we can cope *with* and also that our neuro-cognitive systems *are* complex dynamical systems. Highlights from these chapters include fresh insights on the mind-body problem in chapter two, the authors' breakdown of human intelligence into "primal" and "objectifying" intelligence and their critique of reward-optimization conceptions of intelligence in chapter three, and detailed analyses of language and sociality informed by both phenomenology and empirical work in chapters 4–6. Chapters 7–8 deliver the linchpin of the master argument: the claim that AI systems cannot adequately or synoptically model complex dynamical systems. Chapters 9–12 argue that it follows from what they argue in chapters 1–8 that AGI is impossible, that machines will not master human language or sociality, and also that mind uploading is a waste of time, as are attempts to create digital minds to carry on our civilization. Finally, chapter thirteen makes positive recommendations, discussing what Landgrebe and Smith think AI is good for and how they think it should be used.

Let's look closer at the linchpin of the master argument. In section 7.5.2, Landgrebe and Smith enumerate seven key features of complex systems:

> CHANGE AND EVOLUTIONARY CHARACTER (pp. 126–128). Complex systems evolve in various ways: the system's boundaries can shift, new elements come, and old elements go. In many cases, complex systems can undergo changes in the *types* of elements they contain or interactions they participate in.
>
> ELEMENT-DEPENDENT INTERACTIONS (pp. 128–129). Complex systems typically have different kinds of functionally individuated elements, e.g., the different roles played by proteins, kinases, and ATP in phosphorylation (contrasted with the way that mass and velocity are all you need to chart all of the interactions of a Newtonian system). Elements of a system can also change their functions over time.
>
> FORCE OVERLAY (pp. 130–131). Complex systems typically involve interactions between all four of the basic physical interactions (EM, gravity, strong and weak).

> Non-Ergodic/Complex Phase Spaces (pp. 131–132). We cannot predict the trajectory of a complex system over its phase space by averaging over volumes of that phase space.
>
> Drivenness (pp. 132–136). A driven system is a system that does not generally converge to equilibrium because it has access to a reliable energy source.
>
> Context-Dependence (p. 137). The interface between a complex system and its environment is constantly changing, e.g., which elements are part of the system vs. part of the environment, or what states the system can occupy.
>
> Chaos (p. 137–138). Chaotic systems are unpredictable because small differences in initial conditions may lead to large differences down the road.

Say that systems having all seven of these features are *fully complex*. Landgrebe and Smith's master argument is that AGI would only be possible if fully complex systems could be adequately and synoptically modelled (either the ones in the environment or the ones in the brain), but that fully complex systems cannot be adequately and synoptically modelled.

I am not going to say much here about whether they are correct that fully complex systems cannot be adequately and synoptically modelled. It is intractable to find exact solutions to the dynamical equations for most complex systems (even ones that are not *fully* complex, like three-body gravitational problems). Approximation is thus the name of the game. The more chaos a system exhibits, the more its distribution changes over time, etc., the harder it can be to find approximations that are both tractable and accurate enough for the problem at hand. This much is beyond dispute. However, Landgrebe and Smith are arguing for something extremely ambitious: not just that suitable approximations are sometimes or even typically very costly, but that they are, in principle, unavailable for a wide range of cases and will continue to be, even with the increases in computing power that we can expect the future to bring. This is less clear. It is hard not to look at, for example, NASA's recent successes on missions like DART or OSIRIS-REx and come away with the impression that, when there is a will to find suitably accurate approximations, there is a way.

doi: 10.48106/dial.v77.i3.08

For the remainder, though, I'll focus on the more philosophical questions that arise in Landgrebe and Smith's defence of their claim that AGI would only be possible if fully complex systems could be adequately and synoptically modelled. They pursue two routes to this conclusion. I'll call these the argument from coping and the argument from emulation.

According to the argument from coping, there are fully complex systems in our environments; we cope with them, and AGI is only possible if you can achieve this coping by means of adequate and synoptic modelling.

According to the argument from emulation, our neuro-cognitive systems are fully complex systems, and AGI is only possible if you can emulate them by adequately and synoptically modelling them.

The coping argument is mainly developed in an earlier work, Landgrebe and Smith (2021), but it serves as background for the emulation argument, which is the focus of the present book.

On coping: here, I worry that there is an equivocation. I'll grant unequivocally that there are fully complex systems in our environments, like weather or the stock market, but I'm not sure what it means to allow that we cope with them. Do individuals really cope with hurricanes or stock market crashes? Arguably, imperfect though they are, computational models of hurricanes are the best tools we have for coping with hurricanes. Our coping abilities turn on bounded, often flawed approximations of the chaotic world around us. If those models are enough for coping, then clearly calling for adequate and synoptic modelling sets the bar too high (as a necessary condition for an AI system to count as coping). On the other hand, if these models aren't enough for coping, then, presumably, we can't cope. Either way, the argument from coping fails: either we cannot cope, or the computational methods we use to cope (which fall short of adequate and synoptic modelling) suffice for coping.

On emulation: here, I have a few worries. First, it isn't obvious that our neuro-cognitive systems are fully complex. For example, it is debatable how much chaos there is in the healthy brain, as opposed to criticality or near-criticality (see O'Byrne and Jerbi 2022).

Second, there is an equivocation lurking in the notion of 'emulation' at issue. Is the aim of emulation to create a perfect replica of a specific token system, e.g., to build a concrete model of a specific hurricane, accurate enough to predict where and when that particular hurricane will make landfall? Or is the aim simply to generate a new sample from the same distribution, a new token of the relevant type? If we are after our own digital immortality, then

maybe we must pursue the former project. In contrast, if all we want to do is build an AGI, we only need pursue the latter project.

But I worry that from the claim that the human neuro-cognitive system is fully complex, taken together with the claim that it is impossible to adequately and synoptically model fully complex systems, we only get the impossibility of *token-level* emulation, leaving open the possibility of type-level emulation.

Maybe we cannot build a model of an actual specific hurricane currently out at sea that will allow us to predict to the minute or square mile when and where it makes landfall. But we can build models of hurricanes that embody the profile of hurricanes in general (see Weisberg 2013 for a discussion of distinctions between kinds of predictive models). So too here: the AGI we build might not be a perfect copy of you or me, but this does not preclude that it is adequate as a type-level emulation—especially so since the type in question is *system that has human-level intelligence* and not *system that is as complex as humans*.

Of course, the token-level question is important, too; it seems relevant to questions of uploading and digital immortality. But there is no simple refutation of the possibility of uploading or digital immortality here since numerical identity over time does not require qualitative identity over time.

For the type-level question, of course, it remains to be shown that such an emulation is possible. Maybe adequate and synoptic modelling of fully complex systems would still be required, even if the thing we create is not constrained to perfectly resemble an existing intelligent being.

This is one way that the coping argument fits into the dialectic of the book: if human-level coping involves harnessing our full complexity in order to ride out storms, and we can do this as well as we would if we had adequate and synoptic models of those storms, then emulating our type *in silico* presumably entails adequate and synoptic modelling of fully complex systems somewhere or other. But if, as I suggest above, we don't cope as well as that, then it does not follow that emulating us entails adequate and synoptic modelling of fully complex systems somewhere or other.

Landgrebe and Smith might advocate a slightly weaker fallback claim, which is that the full (or nearly-full) complexity of our neuro-cognitive systems surely must have something to do with all of our intellectual successes, and so nothing will succeed in emulating us (at the type level) if it cannot at the very least *instantiate* the seven key features that make a system fully complex. This isn't obvious: again, the type to be emulated is not *system that is as complex*

doi: 10.48106/dial.v77.i3.08

as humans but rather *system that has human-level intelligence*. Still, it merits consideration.

But there is no clear reason to doubt that we can build digital systems that instantiate these features. Think of Conway's Game of Life or other complex systems generated *in silico* via cellular automaton rules. If we assess criteria like drivenness and force-overlay within the simulation, arguably, these are fully complex. So, too, for deep learning systems, especially if we focus on their dynamics during training (as opposed to inference). During training, parameters evolve (and the sampled distribution changes). Stochastic gradient descent is non-ergodic: systems get stuck in local minima all the time. It can also be chaotic or near-chaotic: there are trajectories that pass along the borders of basins of attraction for local minima. Even during inference, some complex features can be seen. For example, the whole point of attentional mechanisms is to allow models to handle context during inference (see Søgaard 2022), and functional differentiation between different neural network layers (attention vs. feed-forward, pooling vs. convolution, etc.) exemplifies element-dependent interactions. Finally, let's not forget about neural organoids, which are programmable assemblies of biological neural cells: these certainly fit the bill if nothing *in silico* does.

Thus, we have a few reasons to doubt that Landgrebe and Smith fully succeed. Even so, their arguments are important and merit further consideration. If our uploads are guaranteed to differ from us, this problematizes the claim that we can survive into them or that they preserve us, even if the matter is far from settled. And I certainly agree with Landgrebe and Smith about the limits of *current* AI systems and that the question of how and to what degree we adaptively harness our underlying complexity is a key open question, one which we must answer to fully understand the difference between biological minds and AI systems. That said, I am still afraid.*

Jonathan Simon
0000-0001-6081-4127
Université de Montréal
jonathan.simon@umontreal.ca

* Thanks to David Anderson, Jaan Aru, Axel Constant, George Deane, Jordan O'Byrne, Steve Petersen, GPT-4, Claude, and the editors of *Dialectica* for comments on an earlier draft.

References

LANDGREBE, Jobst and SMITH, Barry. 2021. "An Argument for the Impossibility of Machine Intelligence." ArXiv preprint, doi:10.48550/arXiv.2111.07765v1.

—. 2022. *Why Machines Will Never Rule the World. Artificial Intelligence Without Fear*. London: Routledge, doi:10.1002/9780470693476.

O'BYRNE, Jordan and JERBI, Karim. 2022. "How Critical is Brain Criticality?" *Trends in Neurosciences* 45(11): 820–837, doi:10.1016/j.tins.2022.08.007.

SØGAARD, Anders. 2022. "Understanding Models Understanding Language." *Synthese* 200(6), doi:10.1007/s11229-022-03931-4.

WEISBERG, Michael. 2013. *Simulation and Similarity. Using Models to Understand the World*. New York: Oxford University Press, doi:10.1093/acprof:oso/9780199933662.001.0001.

doi: 10.48106/dial.v77.i3.08

Published by *Philosophie.ch*

Verein philosophie.ch
Fabrikgässli 1
2502 Biel/Bienne
Switzerland
dialectica@philosophie.ch

https://dialectica.philosophie.ch/

ISSN 0012-2017

This work is licensed under a Creative Commons Attribution 4.0 International License.

Dialectica is supported by the Swiss Academy of Humanities and Social Sciences.

Abstracting and Indexing Services

The journal is indexed by the Arts and Humanities Citation Index, Current Contents, Current Mathematical Publications, Dietrich's Index Philosophicus, IBZ — Internationale Bibliographie der Geistes- und Sozialwissenschaftlichen Zeitschriftenliteratur, Internationale Bibliographie der Rezensionen Geistes- und Sozialwissenschaftlicher Literatur, Linguistics and Language Behavior Abstracts, Mathematical Reviews, MathSciNet, Periodicals Contents Index, Philosopher's Index, Repertoire Bibliographique de la Philosophie, Russian Academy of Sciences Bibliographies.

Contents