

ORIGINAL ARTICLE

Flaws of Formal Relationism

Mahrad Almotahari

University of Illinois at Chicago

Formal relationism in the philosophy of mind is the thesis that folk psychological states should be individuated, at least partially, in terms of the purely formal inference-licensing relations between underlying mental representations. It's supposed to provide a Russellian alternative to a Fregean theory of propositional attitudes. I argue that there's an inconsistency between the motivation for formal relationism and the use to which it's put in defense of Russellian propositions. Furthermore, I argue that formal relationism is committed to epiphenomenalism about singular mental content.

Keywords Frege's puzzle; folk psychology; content; inference; relationism

DOI:10.1002/tht3.104

A family of doctrines in the philosophy of mind has been winning adherents. The doctrines promise to individuate propositional attitudes so as to vindicate everyday folk psychological explanations and evaluations. The guiding idea is simple: individuating propositional attitudes for the purposes of folk psychology requires that we, in the words of Kit Fine, "go relational." In particular, it requires acknowledging that propositional attitudes have (in a sense to be made precise momentarily) an indispensably relational character. But authors have implemented this guiding idea in different ways. Fine (2007), for example, takes the indispensably relational character of propositional attitudes to be an aspect of content: propositional attitudes that stand in different inferential relations (that license different inferences) have different contents and are thus distinct. When the guiding idea is put to work in this way—as a basis for individuating mental contents—the resulting view is *semantic relationism*. In contrast, Jerry Fodor (1990) and, more recently, Richard Heck (2012) claim that propositional attitudes should be individuated, at least partially, in terms of the purely formal inference-licensing relations between underlying mental representations. According to this view, propositional attitudes that stand in different inferential relations consist of mental representations with different logical forms and are thus distinct, but their corresponding contents may well be the same. When the guiding idea is put to work in this way—as a ground for individuating mental representations with the same content—the resulting view is *formal relationism*, or *formalism* in short.

The motivation for formalism about propositional attitudes is illuminating (Section 1). But I believe the view suffers from two serious flaws. The first is inconsistency

Correspondence to: E-mail: almotaha@uic.edu

(Section 2). There's a latent conflict between the role formalism plays in accounting for some normative features of inference and the role it plays in motivating a view of mental content according to which contents are at least as fine grained as Russellian propositions. The second is that formalism threatens to render singular contents epiphenomenal (Section 3).

1. What is it to "go relational"? Perhaps the best way into *relationism* is via consideration of a *Frege case*: a situation in which someone rationally believes that $F(a)$, appears not to believe that $F(b)$, even though $a = b$.

Consider Fred. He's unaware that his neighbor, Sam Clemens, is Mark Twain. While reading the newspaper one morning, he learns that Mark Twain has died. Fred admires Twain's work, but he isn't terribly saddened by the news. That afternoon, Fred overhears someone he takes to be reliable say, 'Sam Clemens has died'. Fred accepts the report and feels a great loss (Heck 2012, pp. 134–5).

Now we face a familiar puzzle.

(BELIEF) Distinct beliefs have distinct contents.

(RUSSELLIANISM) Distinct contents have either distinct objects and/or properties (not senses) as constituents, or the same constituents arranged in distinct ways.

(PREMISE) The content that Sam Clemens has died and the content that Mark Twain has died have the same objects and properties as constituents and they're arranged in the same way.

(FPC) The content that Sam Clemens has died and the content that Mark Twain has died are identical.

(SPC) Fred's belief that Sam Clemens has died and Fred's belief that Mark Twain has died are identical.

RUSSELLIANISM plus our PREMISE jointly entail the first puzzling conclusion, FPC. Assuming belief is a binary relation between subject and content, it seems that BELIEF and FPC jointly entail the second puzzling conclusion, SPC. To appreciate why FPC and SPC are puzzling conclusions, we make two additional assumptions.

(EVALUATION) Whether an inference is correct is entirely determined by the contents involved.

(EXPLANATION) Fred's mood changed in the afternoon *because* he acquired a new belief.

Now consider the inferences below.

- (1) If Sam Clemens has died, then Fred is upset.
- (2) Sam Clemens has died.
- (3) Therefore, Fred is upset.
- (4) If Sam Clemens has died, then Fred is upset.

- (5) Mark Twain has died.
 (6) Therefore, Fred is upset.

Inferring (3) from (1) and (2) would be a correct inference. Given EVALUATION, substituting one step in the reasoning with another that preserves content ought to preserve correctness, as well. With FPC in the background, replacing (2) with (5) amounts to a content-preserving substitution. So inferring (6) from (4) and (5) ought to be a correct inference, too. But it manifestly wouldn't be—not without the additional premise that Sam Clemens is Mark Twain. In contrast, the correctness of (1)–(3) doesn't require a subsidiary premise. It's correct as is. So it appears that FPC and EVALUATION fail to cohere with our considered logical judgments.

What about SPC? It appears to conflict with EXPLANATION, since one wants to say that the mood-altering belief Fred acquired in the afternoon was the belief that Sam Clemens has died. But SPC bars one from saying that. The belief that Sam Clemens has died is, according to SPC, the belief that Mark Twain has died. And Fred had acquired *that* belief in the morning.

Resourceful authors have suggested ways to reconcile SPC and EXPLANATION. One obvious way to achieve reconciliation is based on the suggestion that, in the afternoon, Fred comes to believe that *his neighbor* has died, and his mood change is due to *that* belief. Since Fred didn't acquire the belief that his neighbor has died when he acquired the belief that Mark Twain has died, EXPLANATION appears to be satisfied in a way consistent with SPC. In general, the idea is to explain Frege cases in terms of descriptive beliefs.¹ The suggestion may seem promising, but, as authors have noted, it carries a significant cost: if singular beliefs don't play an explanatory role in Frege cases, then they don't play an explanatory role in non-Frege cases.²

Whether Fred's mood changed because he acquired the belief that Sam Clemens has died shouldn't be hostage to whether Fred had come across 'Mark Twain has died' in the morning paper. One way of motivating the point is by considering how we would react to being told only half of Fred's story. Suppose we're told no more than that, one afternoon, Fred is informed that Sam Clemens has died, and that Fred accepted the news and became very sad. Suppose we're then asked whether his mood change was due to the newly acquired belief that Sam Clemens has died. I take it that we wouldn't be disposed to think of the case as under-described. We wouldn't require being told first whether Fred had antecedently formed the belief that—as one would naturally put it—Mark Twain has died. In other words, our answer isn't hostage to whether Fred's situation is a Frege case or a non-Frege case. The point here is about naive folk psychological practice, not about the metaphysics of belief. Folk psychological explanations are insensitive to whether Fred's situation is a Frege case or a non-Frege case. But our objective, remember, is to vindicate folk psychology. Since the view under consideration says that Fred's actual mood change wasn't due to the belief that Sam Clemens has died, but to a materially equivalent descriptive belief, a hypothetical mood change unaccompanied by acquaintance with the newspaper report, 'Mark Twain has died', shouldn't be due to the belief that Sam Clemens has died, either. So the view

under consideration begins to look like a form of epiphenomenalism about singular belief.

The considerations above impose a constraint on theorizing about Frege cases.

(CONSTRAINT) Changes of behavior ought to have parallel explanations in Frege and corresponding non-Frege cases.³

The two sorts of case receive parallel explanations iff the explanation of each identifies the same feature (the acquisition of the singular belief that Sam Clemens has died, say) as the reason why the agent's behavior changed. Keep CONSTRAINT in mind; it will be relevant a little later.

With the puzzle now in place, we can say what it is to go relational.

(RELATIONISM) Beliefs standing in different inferential relations are distinct, and facts about inferential relations aren't grounded in intrinsic features of content (e.g., Fregean senses).

To unpack RELATIONISM, consider the belief that Sam Clemens has died. It's related to (i) the belief that *if* Sam Clemens has died, then Fred is upset and (ii) the belief that Fred is upset in such a way that licenses inferring (ii) from (i). The belief that Mark Twain has died isn't inferentially related to those beliefs in the same way. So, by RELATIONISM, the belief that Sam Clemens has died is distinct from the belief that Mark Twain has died, but not because they have different hyperintensional constituents, like Fregean senses. Thus RELATIONISM is incompatible with at least one of the SPC-entailing principles. Which principle one identifies as the source of the problem depends on what one takes to be the basis for facts about inferential relations. According to formalism, "[...] the correctness of an inference is not determined by any relation just between contents, but by a relation between *representations* of those contents" (Heck 2012, p. 172, emphasis in original). In other words,

(FORMALISM) Whether an inference is correct is partly determined by whether the mental representations involved are *formally* related to each other in logically relevant ways.

This presupposes that mental representations have logical form.

How is logical form realized in a mental representation? If there's a language of thought, then logical form is realized in the syntax of Mentalese.⁴ But I don't believe there's very much to recommend the Language of Thought Hypothesis, and I worry that accounting for the normative features of inference by presupposing Mentalese is objectionably psychologistic. In any case, it would take us too far afield to develop these worries here.⁵ What matters presently is that formalists reject EVALUATION and BELIEF. Here is Fodor: "It seems to me plausible that you can have two beliefs with the same object [...], where the difference between the beliefs comes from differences in the inferential/causal roles that the vehicles [of cognition] play" (1990, p. 168).

The very same contents may be encoded by mental representations with different logical forms, in which case the beliefs would stand in different inferential relations and thus be distinct.

FORMALISM not only explains what correct inference involves, it allows for finer distinctions to be made between incorrect inferences.

Consider the arguments below.

(7) If Clemens has died, then Martha is upset.

(8) Martha is upset.

(9) Therefore, Clemens has died.

(10) If Twain has died, then Martha is upset.

(11) Martha is upset.

(12) Therefore, Clemens has died.

While it's certainly incorrect to infer (9) from (7) and (8), "one can nonetheless understand the mistake" (Heck 2012, p. 155). The inference is incorrect but *intelligible*. In contrast, inferring (12) from (10) and (11) is both incorrect *and* unintelligible.⁶ As Heck puts it, "only seriously confused people" would treat (10) and (11) as a reason to believe (12). "Yet the only difference between these arguments is that (7) is formally related to (9) in a way that (10) is not formally related to (12)" (ibid.). Because this observation will be relevant a little later, let's give it a name so that we can easily refer back to it.

(INTELLIGIBILITY) Intelligible inference requires that the mental representations involved be formally related.

Formalists say it's because (10)–(12) fails to satisfy INTELLIGIBILITY that it's worse than (7)–(9) as a piece of reasoning.

2. The formalist isn't committed to RUSSELLIANISM; she can coherently ally her view with a far more coarse grained theory of content. One might wonder, then, whether there's any positive reason to combine FORMALISM with RUSSELLIANISM. Heck (2012, p. 169) sketches an argument to the effect that "the contents of beliefs should be individuated at least as finely as Russellian propositions." Let me reproduce the argument in Heck's words.

First, the fact that belief-states stand in certain formal relations to one another (and not in others) is essential to our ability to explain what needs explaining in Frege cases. I therefore take us to be licensed to regard these states as being logically articulated in a sense that should be uncontroversial. [. . .] For example, saying that Fred's belief that Twain has died has the logical form $\phi(\alpha)$ is a way of characterizing the formal relations in which it stands to other beliefs. [. . .]

Second, this logical articulation is not merely syntactic but also has a semantic aspect: *Other beliefs that are formally related to Fred's belief that Twain has died via its*

α -component (if I may put it that way) share an intentional feature with it, namely, that they too are about Twain. And this shared intentional feature is implicated in at least some of the explanations in which these beliefs are implicated. [. . .]

If so, then psychological explanations that mention Fred's belief that Twain has died appeal to intentional features of this belief that are determined neither by its truth-value nor by the set of worlds in which it is true. [. . .] The intentional features of the belief simply outstrip both its truth-value and the set of worlds in which it is true. Russellian propositions are designed precisely to remedy this problem (ibid., emphasis added).

I'm especially interested in the italicized statement at the center of this passage. It expresses a principle that links sameness of formal components with sameness in features of intentional content. For ease of exposition, let's give the principle a canonical formulation.

(LINK) Formally related mental representations have the same intentional feature (are about the same thing).

LINK plays a key role in Heck's argument for contents at least as fine grained as Russellian propositions. But I believe that LINK and INTELLIGIBILITY are incompatible. My reason, in outline, is that they jointly entail the thesis below.

(CONSEQUENCE) Intelligible inference requires that the mental representations involved have the same intentional feature (are about the same thing).

And CONSEQUENCE is false.

Suppose Kurt is the greatest logician who ever lived. Additionally, suppose Kurt is tracking an object in his visual field. At t_1 he judges that that object is F . Then he blinks and the object in his line of sight is quickly replaced with a duplicate. When Kurt opens his eyes, he judges that that object is G . Kurt then infers that something is both F and G . Assume that Kurt's inference doesn't rely on a suppressed judgment that that object (at t_1) is identical to that object (at t_2). As a matter of fact, Kurt lacks a piece of information, namely, that that object (at t_1) is not identical to that object (at t_2). And, in light of this information, the conclusion of the argument below may well be false even if both premises are true.

- (13) That object is F .
- (14) That object is G .
- (15) Therefore, something is both F and G .

But what about Kurt's inference—the mental transition from two antecedent judgments to a third judgment. Is *it* rational? Well, opinions vary. Some say yes, some say no, and others say it's unclear. I vacillate between thinking yes and thinking it's unclear. But

surely we can all agree that Kurt's inference is *intelligible*. It would be unfair for Kurt to classify his inference with (10)–(12) rather than (7)–(9). We can understand Kurt's inference; it isn't a display of *total* confusion.

So I assume we're agreed: Kurt's inference is intelligible even if irrational. By INTELLIGIBILITY, the two occurrences of 'that object' in the mental analogs of (13) and (14) are formally related. It follows by LINK that those mental representations have the same intentional feature (are about the same thing). But that's false, since the mental analog of 'that object' in (13) is about one object in Kurt's visual field at t_1 while the mental analog of 'that object' in (14) is about a distinct object in Kurt's visual field at t_2 . So although Kurt's inference is intelligible, the mental representations involved don't have the same intentional feature. Thus CONSEQUENCE is false.^{7,8}

One of either LINK or INTELLIGIBILITY must be abandoned, but which? To give up LINK is to give up on the idea—central to the argument above—that the “logical articulation” of belief states has a semantic aspect. *A fortiori*, one would give up the idea that the logical articulation of belief states has an “explanatorily relevant semantic aspect” (Heck 2012, p. 170). If one opts for this response, then worries about epiphenomenalism begin to set in. (We will explore this worry in greater detail momentarily.) On the other hand, to give up INTELLIGIBILITY is to give up a central motivation for FORMALISM itself—specifically, that a certain normative feature of inference (namely, intelligibility) is to be explained in terms of formal relations. I leave it to the formalist to pick her poison.

3. The formalist account of Frege cases renders singular mental content epiphenomenal. To see why, think what would happen if we put FORMALISM to work in the service of a super coarse grained, “psycho-Fregean” view about content.

According to the psycho-Fregean, there are just two contents, Truth and Falsity. All true representations encode the former; all false representations encode the latter. For the psycho-Fregean, Fred's situation is to be explained not in terms of differences between contents, but in terms of differences between inferential relations between logical forms. As Heck acknowledges, “It *would* be a problem if the psycho-Fregean view could be defended along [such] lines: That would seriously call into doubt whether content plays any role in psychological explanation” (p. 166). In response, however, he writes:

But the psycho-Fregean cannot mimic the account I have offered. Consider, for example, this explanation:

[A]

Fred had a belief b_1 with the content <Clemens, having died>; Fred also had a belief b_2 with the content <Clemens, =, his neighbor>; these beliefs were formally related via their respective first terms. He was therefore able to infer the belief with

the content <his neighbor, having died>, where this belief is formally related to b_1 via their second terms and to b_2 via their first and last terms, respectively.

If we try to adapt this pattern on behalf of the psycho-Fregean, we get something like the following:

[B]

Fred had a belief b_1 with the content Falsity⁹; Fred also had a belief b_2 with the content Truth; these beliefs were formally related via . . . what? He was therefore able to infer a belief with the content Falsity, where this belief is formally related to b_1 via . . . what? And to b_2 via . . . what?

The problem is that there is not enough structure in the ‘contents’ of these beliefs for us to state how they are formally related (ibid.).

But why should a formalist regard this as a problem?

The formalist thinks inferences are licensed not by a relation between contents, “but by a relation between *representations* of those contents” (Heck 2012, p. 172). So talk of contents being “formally related via their respective first terms” must be short for *representations* of those contents being formally related via *their* respective first terms.

A*

Fred had a belief b_1 consisting of a mental representation of the form $D(c)$ encoding the content <Clemens, having died>; Fred also had a belief b_2 consisting of a mental representation of the form $c = \text{the } N$ encoding the content <Clemens, =, his neighbor>; these beliefs were formally related via the first term of each mental representation, ‘ c ’. Fred was therefore able to infer the belief consisting of a mental representation of the form $D(\text{the } N)$ encoding the content <his neighbor, having died>, where this representation is formally related to b_1 via the second term of its underlying representation, ‘ D ’, and to b_2 via . . .

Thus A* reflects FORMALISM more faithfully. But now the absence of structure in psycho-Fregean ‘contents’ is no impediment to stating how the beliefs are formally related, since the formal relations are specified completely in terms of the logical form of underlying mental representations. For the psycho-Fregean to mimic the formalist account, all she has to do is delete reference to Russellian propositions and insert ‘Truth’ and ‘Falsity’ in the appropriate places.

My point isn’t that the psycho-Fregean can claim all of the advantages of the Russellian’s account. FORMALISM isn’t a cure for psychosis. But it offers a local treatment, illustrating a more modest point: the psycho-Fregean *is* able to mimic the Russellian formalist’s account of Fred’s situation. Now I take it that, in the psycho-Fregean’s explanation of Fred’s situation, contents make no explanatory contribution at all. What then entitles the Russellian to think that, *in her account of Fred’s situation*, they *are* performing an explanatory service?

Formalists say the difference between the belief that Sam Clemens has died and the belief that Mark Twain has died in virtue of which the former, but not the latter, is well suited to explain Fred's mood change isn't a matter of content; it's a matter of form. After all, the two beliefs have the same content. So the feature of Fred's psychology that the formalist—whether Russellian or psycho-Fregean—identifies as the reason why his mood changes is the form of a certain mental representation that Fred acquires upon being told that Sam Clemens has died. Now this observation by itself doesn't raise the specter of epiphenomenalism, for it may be a peculiarity of the Frege case. But, in light of CONSTRAINT, there is a great deal of pressure to say that even in the corresponding non-Frege case—in which Fred hasn't antecedently come to believe that Mark Twain has died—the content of the belief that Sam Clemens has died makes no explanatory contribution. For, otherwise, the two cases wouldn't receive parallel explanations—not in the intended sense of “parallel,” anyway. But if singular contents are idle in Frege *and* non-Frege cases, then epiphenomenalism sets in.

The discussion has been about propositional attitudes, but the consequences potentially expand. If one is attracted to a unified model of linguistic and mental intentionality—and why should the medium of representation alone make for disunity?—then the flaws of FORMALISM suggest that the most promising way to go relational is to embrace semantic RELATIONISM about mental *and* linguistic representation.

Acknowledgments

The author benefitted from discussions with Walter Edelberg, Aidan Gray, Dave Hilbert, Crispin Wright, and an audience at the Northern Institute of Philosophy in Aberdeen, Scotland, where this material was presented under the misleading title ‘Flaws of Formalism’. Aidan deserves special recognition for many months of stimulating conversations about naming, referring, and inferring.

Notes

- 1 See David Braun (2001) and Michael Thau (2002).
- 2 See Heck (2012) and Gabriel Segal (1996, 2000).
- 3 ‘Behavior’ should be construed quite broadly so as to include things like changes in mood and inferences.
- 4 Fodor (1990) defends a view of just this sort. And, in several places, Heck expresses sympathy for the Language of Thought Hypothesis (2012, pp. 152 and 159). But, for the most part, he seems to want to kick the question upstairs, or to postpone it for another occasion.
- 5 See Robert Matthews (2007).
- 6 Or perhaps not *even* incorrect because unintelligible.
- 7 My story about Kurt is based on an example in John Campbell (1988). But Campbell uses it to make a different point. Situations very like Kurt's have received some attention in the literature, and are labeled cases of “confusion”. It's not clear to me whether all of the cases that have been so labeled should be grouped together, but I omit a fuller discussion for the sake of brevity. See Joseph Camp (2002) and Krista Lawlor (2007).

- 8 Two points are worth noticing. First, the argument here doesn't essentially rely on demonstratives. The same point can be made using names. Suppose an impersonator had taken over Napoleon's role early in Napoleon's life (in, say, 1793). Before the switch, suppose Kurt forms a judgment that he would express as follows: 'Napoleon is *F*'. Decades pass and the impersonator becomes for Kurt "the dominant source of his associated body of information" (Gareth Evans 1973, p. 303). In 1815, Kurt forms a judgment that he would express as follows: 'Napoleon is *G*'. He then combines the two judgments and infers that someone is both *F* and *G*. Kurt's inference is intelligible, I claim, but the judgments aren't about the same person. The judgment in 1793 is about the original Napoleon; the judgment in 1815 is about the impersonator, since he's the dominant source of information at that time. Thanks to Aidan Gray for drawing my attention to this example in this context. The second point worth noticing is that, as far as I can tell, the argument here applies with equal force to Fregean senses, since they're supposed to determine the normative features of inference *and* the intentional features of mental representations.
- 9 At this stage Heck is assuming that the reports about Twain's demise were inaccurate.

References

- Braun, David. "Russellianism and Explanation," in *Philosophical Perspectives 15: Metaphysics*, edited by James Tomberlin. Malden, MA: Blackwell, 2001, 253–89.
- Camp, Joseph. *Confusion: A Study in the Theory of Knowledge*. Cambridge, MA: Harvard University Press, 2002.
- Campbell, John. 1988. Is Sense Transparent? *Proceedings of the Aristotelian Society*, LXXXVIII: 273–92.
- Evans, Gareth. "The Causal Theory of Names." *Proceedings of the Aristotelian Society*, XLVII (1973): 187–208 Reprinted in Martinich AP, ed. *The Philosophy of Language: Fourth Edition*. New York: Oxford University Press, 296–307. Page references to this volume.
- Fine, Kit. *Semantic Relationism*. Oxford: Wiley-Blackwell, 2007.
- Fodor, Jerry. "Substitution Arguments and the Individuation of Belief," in *A Theory of Content and Other Essays*. Cambridge, MA: MIT Press, 1990.
- Heck, Richard. "Solving Frege's Puzzle." *Journal of Philosophy* CIX (2012): 74–132.
- Lawlor, Krista. "A Notional Worlds Approach to Confusion." *Mind and Language* 22 (2007): 150–72.
- Matthews, Robert J. *The Measure of Mind: Propositional Attitudes and their Attribution*. Oxford: Oxford University Press, 2007.
- Segal, Gabriel. "Frege's Puzzle as Some Problems in Science." *Revista de Linguística* VIII (1996): 375–88.
- . *A Slim Book About Narrow Content*. Cambridge, MA: MIT Press, 2000.
- Thau, Michael. *Consciousness and Cognition*. Oxford: Oxford University Press, 2002.