

An Argument against Causal Theories of Mental Content

Some mental states are about themselves. Nothing is a cause of itself. So some mental states are not about their causes; they are about things distinct from their causes. If this argument is sound, it spells trouble for causal theories of mental content—the precise sort of trouble depending on the precise sort of causal theory. This paper shows that the argument is sound (§§1-3), and then spells out the trouble (§4).¹

I. Validity.

The argument has two premises, which may both be framed in terms of reflexivity. The first premise claims that the causal relation is irreflexive; the second that the intentional relation is not. So framed, the validity of the argument is a matter of Leibniz's Law.

Irreflexive relations do not apply to any object and itself. The relation of being a brother is an example, as is the smaller-than relation.

Relation R is *irreflexive* =_{df} $\forall x \forall y (R(x,y) \rightarrow \neg R(x,x))$.

Two sorts of relations are not irreflexive. Some are not irreflexive in a strong sense. These relations are reflexive, properly speaking. A reflexive relation applies to an object and itself, if it applies to an object and anything. The identity relation is an example, and, supposing the principle of one object to a place holds, the relation of being in exactly the same place is another. Other relations are not irreflexive in a weaker sense; they are neither reflexive nor irreflexive, properly speaking. These relations apply to an object and itself in some cases but not in others. Call these partly reflexive. The reflection

¹ Others have mentioned the tension developed in this paper. See Kriegel (2005: 43-44) and Levine (2001: 173).

relation a mirror may, in the right circumstances, bear to itself as well as its surroundings is an example. A mirror may bear this relation to distinct things without bearing the relation to itself; but it may bear the relation to itself as well. These two forms of irreflexivity are expressed in the following definitions:

Relation R is *reflexive* =df $\forall x \forall y (R(x,y) \rightarrow R(x,x))$

Relation R is *partly reflexive* =df $\exists x R(x,x)$

The claim that the intentional relation is not irreflexive should be understood in the weaker sense, i.e., as the claim that there is at least one pairing of an object with itself which satisfies the relation.

Applying Leibniz's Law to properties can prove rather shallow distinctions, since distinct properties may agree in extension (e.g., creatures with a heart and creatures with a kidney). It is therefore worth noting that differences with respect to reflexivity are easily parlayed into differences in the extension of relations. Let R be a relation (a binary relation, for the sake of simplicity), and \mathfrak{R} the extension of this relation.

R is *irreflexive* =df $\forall x \forall y (x = y \rightarrow \langle x,y \rangle \notin \mathfrak{R})$

R is *partly reflexive* =df $\exists x \exists y (x = y \ \& \ \langle x,y \rangle \in \mathfrak{R})$.

The extension of any irreflexive relation excludes all ordered pairs of a thing and itself. Relations which are partly reflexive include at least one ordered pair of a thing and itself. Since extensions are sets and sets are identified by their members, the extension of any irreflexive relations is distinct from the extension of any relation that is not irreflexive.

If our premises are true, then, the intentional relation is neither a causal relation nor extensionally equivalent to a causal relation.

II. The Causal Premise.

Why think ‘x is a cause of y’ is irreflexive? Standard accounts take the relation to be two-placed, with distinct events as *relata*. The requirement of distinct *relata* is sufficient to render the causal relation irreflexive. But what motivates the requirement?

An argument may be based on the *ex nihilo nihil fit* principle. Suppose, for *reductio*, the causal *relata* are identical, i.e., the effect is the cause. A cause is, by definition, something that brings about the effect, or brings the effect to be. By our supposition, then, the effect is something that brings about the effect, the effect brings itself to be. To bring itself to be, the effect must both exert causal influence and not exist. But what does not exist exerts no influence. For nothing comes from nothing. So the supposition is false. Causal *relata* are distinct.

This argument shows that the relation of *being a cause* is irreflexive. This is true even for what we, loosely speaking, call self-causation. Many events are composed of distinct smaller events. In such cases, there is nothing absurd about the later parts being causally dependent upon the earlier. David Lewis dubs this “piecemeal” causation (1986: 172-175, also 212-213). He gives the example of a public address system howling with feedback. Each part of the howling (except the first) is caused by an earlier part, and each part (except the last) causes a later part. In this sense, the howling (as a whole) is partly a cause of itself. But no part of the howling is a cause of itself.²

² In personal correspondence, both Rob Rupert and John Dilworth have independently suggested that causal theories of content may look to such causal feedback loops for an analysis of reflexive intentional relations. Such loops may provide the causal theory a way to show how one stage of a temporally extended mental event can be

Causal theories of mental content appeal to more sophisticated causal relations than simply ‘being a cause.’ No causal theorist thinks that ‘x represents or is about y’ is simply the inverse of ‘y is a cause of x.’ Such a crude analysis is far too generous in its attribution of intentionality. Causal theorists hone their analysis of intentionality, however, by *adding* conditions to the fact that ‘y is a cause of x’—for example, y is a cause of x in ideal conditions (Stampe 1977), or y a cause of x which has the function of indicating x (Dretske 1988), or y is a cause of x and anything else, z, that causes x does so because y does (Fodor 1987). Causal theories entail, minimally, that if x is about y then y is a cause of x—this is what makes them *causal* theories. Thus causal theories analyze intentional relations in terms of irreflexive causal relations.

III. The Intentional Premise.

Why think there are reflexive intentional relations? The simplest way to establish an existential claim of this sort is by example. An artist may paint a portrait of a subject holding a portrait. There does not appear to be anything preventing the portrait held by the subject from representing the very portrait the artist is painting. Linguistic expressions also appear to represent themselves, and famously so: The sentence following the last colon is false. Mental representations exhibit a similar structure. A materialist may think: *the very thought I am now thinking has physically necessary and sufficient conditions*. Gilbert Harman recounts the case of Alice, an insomniac, who

about a later stage. But the type of reflexive intentionality presented in the next section is more robustly reflexive; it involves one stage being about itself. In any case, the task of this paper is not to evaluate possible solutions to a problem for the causal theory, but to argue that there is a problem.

thinks to herself, during a particularly restless night: “I am not going to fall asleep because of my having this very thought” (2006: 334).³ The force of such examples is strengthened by the fact that there is nothing about the very idea of intentionality which precludes self-representation; i.e., there is no *a priori* argument for the irreflexivity of the intentional relation like the one just offered for the causal relation.⁴ Since examples in philosophy can be as controversial as the claims they support, and since the lack of an argument *contra* is not an argument *pro*, it is important to see how strong a case may be made for reflexive intentional relations.

Uriah Kriegel (2003) pioneered the best argument on offer. He assumes there are no reflexive intentional relations and then attempts to account for the representational states implicit in certain episodes of consciousness. Kriegel focuses on two attempts to account for the relevant representational states without reflexive intentional relations, arguing that each contradicts plausible auxiliary premises. The argument presented in this section, expands upon Kriegel’s strategy by addressing three, logically exhaustive, explanatory options for the denier of reflexive intentional relations. The contradiction implied by one of the three options has been known at least since Aristotle (Caston 2002), and is attributed to Brentano by Kriegel. Kriegel himself locates the problem with the

³ Harman credits the case to Derek Parfit. He also discusses several more complicated examples of what he calls “self-reflexive thoughts”—more complicated in that they do not contain any explicit representation of themselves.

⁴ For discussion of some unsuccessful arguments against reflexive intentional relations, see Harman (2006: 334-335) and Williford (2006). Harman discusses arguments based on liar-like paradoxes; Williford takes up an alleged infinite regress.

second (2003). The recent work of Brook and Raymond (2006), and Horgan, Tienson and Graham (2006) suffices to block the third.⁵

As indicated by Kriegel's strategy, the supposition that there are no reflexive intentional relations is not self-contradictory. The consequences of the supposition do not contradict one another, but rather a stock of background premises—six to be precise. Each of these premises is a point at which the conclusion that there are reflexive intentional relations may be resisted, and a complete defense of these premises is more than a single paper can accomplish. In five of the six cases, a few supporting comments suffice to establish the plausibility of the premises. One premise demands more attention, however.

The Simple Representational Thesis. Kriegel's strategy is driven by the need for a coherent explanation of the representational states implicit in certain episodes of consciousness. That representational states are implicit in certain episodes of consciousness is therefore a crucial background premise.

Following David Rosenthal (1997), we distinguish state from creature consciousness, and intransitive from transitive consciousness. Transitive consciousness is awareness, i.e., consciousness of something. Creatures are transitively conscious of objects by having mental states which represent (are about) those objects. The states themselves are not, like the creatures who have them, aware of anything. To attribute

⁵ Kriegel is not unaware of the third possibility for the denier of reflexive intentional relations; he briefly addresses something akin to the option in reply to an objection (2003: 129). The option, as formulated below, is less easily dispatched, for reasons stated in note 14.

consciousness to the states themselves is rather to attribute to them intransitive consciousness. Intransitively conscious states are, at a minimum, those whose subject is aware of being in them. They are contrasted with unconscious states, states whose subject is unaware of being in them. Intransitive state consciousness implies transitive creature consciousness (awareness) of the state. A mental state is intransitively conscious only if the creature which has the state is transitively conscious (aware) of it. Transitive creature consciousness, in turn, implies representation. Creatures are transitively conscious (aware) of something, only if they are the subject of a mental state which represents (is about) that thing.⁶ Letting x and y range over the mental states of a single subject (for the sake of simplicity), it follows that:

$$\forall x(\text{IntransitiveStateConscious}(x) \rightarrow \exists y(\text{Represents}(y,x))).$$

This is the first background premise—call it the *simple representational thesis*.

The simple representational thesis is carefully worded to avoid the entanglements of its more controversial kin. First, the thesis makes the representation of a mental state a necessary condition of its being intransitively conscious, not a necessary and sufficient condition. The thesis is therefore not committed to Rosenthal's representationalism about consciousness—or anyone else's, including Kriegel's. Representationalism is a reductive program, an attempt to state the necessary and sufficient conditions for intransitive state consciousness in representational terms.⁷ The simple representational thesis claims only

⁶ Acquaintance theorists may dispute the representational necessary condition on awareness. This issue is addressed below.

⁷ As an anonymous referee noted, 'representationalism' does not always denote a reductive program; see Chalmers (2004). For an overview of the reductive programs the

that intransitively conscious states are represented, not that their being conscious consists in their being represented.

Second, the simple representational thesis does not claim that the representation of an intransitively conscious state is the only representational condition necessary for intransitive state consciousness. Perhaps intransitive consciousness is conceptually connected with the representation of other things as well—e.g., the subject of the intransitively conscious state, the fact that the subject has the intransitively conscious state. The simple representational thesis requires only that the intransitively conscious state itself is represented, not that the pain alone is represented.

Finally, though the simple representational thesis requires intransitively conscious states to be represented, it does not place any requirements on the type of representation necessary. The thesis is therefore satisfied by the most primitive forms of mental representation. It does not require the representation to be reflective, conceptualized,

term denotes, see Lycan (2006). The higher-order variety of representationalism is examined in Gennaro (2004). Tye (1999) and Dretske (1995) present a first-order variety. The self-representational variety is explored in Kriegel and Williford (2006). Those committed to a self-representational analysis of consciousness are *ipso facto* committed to reflexive intentional relations, but not vice-versa. Proponents of higher-order and first-order versions of representationalism are not *ipso facto* opponents of reflexive intentional relations. The first-order representationalist's claim that a state is conscious iff it represents something else in just the right way does not entail that no state represents itself; nor does the higher-order representationalist's claim that a state is conscious iff it is represented by something else in just the right way.

discursive, or explicit. The claim is simply that you have transitive creature consciousness (awareness of), and hence a representation, of your intransitively conscious states. Transitive creature consciousness is common to human beings and creatures (e.g., animals) who lack complex conceptual repertoires.

This last point provides a working answer to the first of two objections to the simple representational thesis in the literature. Carefully worded though it is, the simple representational thesis is open to an apparent counter-example raised by Dretske (1993). He displays two complicated arrangements of shapes, which differ only minutely (one has a small spot the other lacks). When comparing the arrangements, most observers initially fail to notice the difference. This means some observers have an intransitively conscious visual experience of the difference maker (the spot), but are not transitively creature conscious (aware) that the difference making feature is present. Dretske presents this as evidence that awareness of a state is not a necessary condition of its being intransitively conscious. But the objection supposes that conscious experience of the difference maker (the spot) requires awareness of the fact that this aspect of one's experience differentiates it from others.⁸ As the last comment emphasizes, the simple representational thesis requires only that the difference making aspect of one's experience is represented in some way or other, not that it is represented as such, i.e., not that it is represented as the difference maker.

A second, more recent, objection accepts that awareness of a mental state is a necessary condition of its being intransitively state conscious, but denies that

⁸ Bryne (1997: 113-114) presses this response; Seager (1994) develops the point at some length.

representation is a necessary condition on awareness.⁹ If the representational state invoked by the simple representational thesis is representational only in the sense that it is directed at an object, there is nothing to dispute. But if the representational state is representational in the sense that it somehow encodes conditions which its object may or may not satisfy, some acquaintance theorists may demur (e.g., Hellie, 2007: 299-301). To ground an objection, ‘acquaintance’ must be used in this context to denote a form of non-representational awareness. If awareness comes in two forms (representational and non-), then the simple representational thesis is false: we may be aware of our intransitively conscious states without representing them; i.e., we may be aware of them by acquaintance.

The possibility of non-representational awareness of our intransitively conscious states need not detain the argument for reflexive intentional relations, for at least two reasons. First, acquaintance functions (at least for normal human beings) as a prelude to representational modes of awareness of its objects. Acquaintance relations earn their keep by securing our most primitive objects of thought and representational abilities; acquaintance that is not accompanied by representational modes of awareness is of very limited cognitive significance.¹⁰ Appealing to acquaintance to explain how we are aware

⁹ Thanks to an anonymous referee for bringing this objection to my attention.

¹⁰ Chalmers (2003), for example, characterizes acquaintance in terms of its ability to ground the formation of direct phenomenal concepts and the justification of phenomenal belief. Hellie (2007: 293-294) also recognizes the close connections between non-representational modes of awareness (he calls them NIRAs for nonintentional relations of awareness) and the representational (which he calls IRAs).

of intransitively conscious states, then, only delays the claim that intransitively conscious states are represented. The need for a coherent account of their representation therefore survives the appeal to acquaintance. Second, acquaintance is typically understood as a *sui generis* relation,¹¹ and is therefore already a problem for causal theories of mental content. Whether acquaintance is, in the final analysis, *sui generis*, it is almost certainly not causally analyzable. For states of acquaintance and their objects bear a closer relation than any causal *relata*. Because casual *relata* are distinct, it is possible for the cause to exist without the effect and vice-versa. But it is not possible for a state of acquaintance to exist in the absence of its object—the content of states of acquaintance is constitutively dependent upon the existence of the object of acquaintance.¹² Acquaintance relations, therefore, may suppress one problem for causal theories of mental content (i.e., reflexive intentional relations) only by raising another (i.e., acquaintance).

Though no argument may be needed beyond this explication and clarification of the simple representational thesis, one may be given from what Kriegel (2005) calls the subjective character of intransitively conscious states. Nothing that undergoes an

¹¹ Fumerton (1995: 74) makes this point explicitly.

¹² Hellie (2007: 299) cites this close connection between acts of acquaintance and their objects as one of the primary advantages of appealing to acquaintance, rather than representational modes of awareness, to account for our awareness of intransitive state consciousness. He even suggests that a natural explanation of the connection “is that the phenomenal character is not distinct from the awareness” (318 n. 20). An acquaintance theory developed in the direction of this suggestion is entirely compatible with reflexive intentionality, broadly construed.

intransitively conscious state, like a pain, is unaware of whose state it is. Pain is experienced as one's own. To experience a mental state as one's own implies awareness of the experience as one's own. Awareness of the experience as one's own is *a fortiori* awareness of the experience. Awareness of the experience implies representation of the experience. So the experience of pain requires the representation of pain.

The Argument. Suppose there are no reflexive intentional relations. It follows that nothing is about or represents itself, not even partly. In the context of the simple representational thesis, this consequence is demonstrably false. For the representational state by virtue of which you are aware of an intransitively conscious state either is or is not identical to the intransitively conscious state itself. If the state by virtue of which you are aware of an intransitively conscious state is identical to the intransitively conscious state itself, then, contrary to our supposition, a mental state represents itself and there are reflexive intentional relations. If, on the other hand, nothing represents itself, then the representational state is not identical to the intransitively conscious state itself. Instead, episodes of intransitive state consciousness are a complex of two states: an intransitively conscious state and a distinct representational state. This means that if you pound your thumb with a hammer you experience two states: your pain and a representational state by virtue of which you are aware of the pain. Call the conjunction of the simple representational thesis and the non-identity of the representing and represented states the *representational appendage thesis*.¹³ The thesis claims that every intransitively conscious state is accompanied by a distinct representational state which represents it; i.e.:

¹³ The terminology here is due to Natsoulas (1993); see also Rosenthal (1993).

$\forall x(\text{IntransitiveStateConscious}(x) \rightarrow \exists y (\text{Represents}(y,x) \ \& \ y \neq x)).$

This thesis is to be contrasted with the conjunction of the simple representational thesis and the identity of the represented and representing states, i.e., the *representational aspect thesis*:

$\forall x(\text{IntransitiveStateConscious}(x) \rightarrow \exists y (\text{Represents}(y,x) \ \& \ y = x)).$

According to this thesis, when you pound your thumb and experience pain and awareness of pain, you experience a single state with two aspects. Your pain has intentional properties in addition to its familiar phenomenal properties, and these intentional properties are (at least) partly reflexive.

The supposition that there are no reflexive intentional relations entails the representational appendage thesis. The representational appendages posited by the thesis are, themselves, either all intransitively state conscious or not. If they are not all conscious, then either none are or some are and are not. Each of these three cases—all the representational appendages are conscious, none are, or some are and some are not—contradicts plausible background premises.

Suppose the appendages are all intransitively state conscious (henceforth simply called conscious), i.e., suppose *representational appendage*_{conscious}:

$\forall x(\text{Conscious}(x) \rightarrow \exists y (\text{Represents}(y,x) \ \& \ y \neq x \ \& \ \text{Conscious}(y))).$

Kriegel (2003) presses a familiar regress argument against this supposition, pitting it against the very plausible background assumption that human minds are finite, i.e., not the subject of an infinite number of conscious mental states. Take the case of the pain in your pounded thumb, and call it M. M is conscious and therefore represented; call the state representing M, M*. Since we are assuming there are no reflexive intentional

relations M is not M*. Since we are assuming all the representational appendages are conscious, M* is conscious. Because M* is conscious, it falls within the scope of *representational appendage_{conscious}* and is therefore represented. By parity of reasoning, M* must be represented by a distinct state, M**, which, again, must be conscious and therefore represented by a distinct state, M***, and so on *ad infinitum*.

Now suppose the representational appendages are not all conscious. As noted, the case breaks into two sub-cases: either none of the appendages are conscious or some are and others are not. Suppose none are, i.e., suppose *representational appendage_{unconscious}*:

$$\forall x(\text{Conscious}(x) \rightarrow \exists y (\text{Represents}(y,x) \ \& \ y \neq x \ \& \ \sim \text{Conscious}(y))).$$

This option conflicts with two other very plausible background premises, as Kriegel explains (2003: 120-123): experienced mental states are conscious, and awareness of our mental states is (at least sometimes) experienced. If the representational appendages are all unconscious, then, in the case of your pounded thumb, M* is unconscious. But M* constitutes your awareness (transitive creature consciousness) of M, your pain. If M* is unconscious, then your awareness of M is unconscious. If your awareness of M is unconscious, then awareness of M is not something you experience—since all experienced mental states are conscious. It follows that you would experience pain but not awareness of your pain. Kriegel neatly illustrates the absurdity of this result by noting that if awareness of your pain is not something you experience, it is not something you could learn (become aware of) by experience. To become aware of the pain you would have to posit the existence of your pain for the sake of its explanatory power vis-à-vis observed behavior. But we learn of our pain by experience, not inference. To become aware of pain by experience, awareness of pain must be something we

experience. Since all experienced mental states are conscious, awareness of our pain must be conscious. In the case of the pounded thumb, awareness of your pain just is M^* . So M^* must be conscious. But M^* is one of the representational appendages. So it is not the case that all the representational appendages of consciousness are unconscious; some, e.g., M^* , are conscious.

The supposition that all the representational appendages of consciousness are conscious leads to an infinite regress; the supposition that none are is inconsistent with the fact that awareness of our mental states is experienced. The only viable option, then, is that the appendages are not homogeneous with respect to consciousness, that some of the appendages are and others are not conscious.

*representational appendage*_{mixed}: $\forall x(\text{Conscious}(x) \rightarrow \exists y (\text{Represents}(y,x) \ \& \ y \neq x \ \& \ (\text{Conscious}(y) \vee \sim\text{Conscious}(y))))$.

This position avoids the problems generated by the earlier iterations of the appendage thesis. Since we no longer assume that all the representational appendages are unconscious, we may concede that M^* , your awareness of your pain, is conscious, and therefore experienced. Since we no longer assume that all the representational appendages are conscious, the concession does not generate an infinite regress. It generates a regress, but a regress of only a step. Since your awareness of your pain, M^* , is something you experience, it is conscious. Since it is conscious, it is represented by M^{**} . Since we are assuming no mental states represents itself, M^* is not M^{**} . But since there is nothing requiring M^{**} to be conscious, the regress may end here. If you introspect and experience awareness of M^{**} , this would push the regress back one more step. Presumably, there will be a point past which it is not reasonable to claim that your

awareness of your awareness (of your awareness of your awareness) is something you experience, and is therefore conscious, and therefore represented. At that point the regress ends in an unconscious representation of the immediately preceding state. The regress thus goes only as far as experienced awareness of our mental states requires. Since experienced awareness of our mental states does not go on to infinity, neither does the regress.¹⁴

¹⁴ As Gennaro (2004b) explains, most higher-order theorists of consciousness handle regress concerns in the way opened by the mixed appendage thesis; see for example Rosenthal (1986: 336). Higher-order theorists take the regress not to be a problem, but an indication of something important about the structure of viable higher-order analyses. The moral is not that the states constituting our awareness of a lower-order state are never conscious, but rather if they are conscious they are represented by a still higher-order state. Since the second-order states constituting our awareness of first-order intransitively conscious states are, as Kriegel shows, typically conscious, they are typically represented by a third-order state, according to any viable higher-order theory. Since the third-order state need not be conscious, however, it need not be represented. Kriegel (2003: 129) suggests that a regress still arises. This is because he formulates what I am calling the mixed version of the appendage thesis so that it posits two appendages, one conscious and the other not, at each step of the regress. This way of framing the mixed thesis still has the appendage thesis positing a conscious appendage at each step of the regress, and therefore is still subject to the regress. The version of the mixed thesis I have offered, however, is not subject to this problem; it posits one appendage which may be either conscious or unconscious.

The problems for *representational appendage*_{conscious} and *representational appendage*_{unconscious} are not, however, the only problems facing appendage theses. The distinction between represented and representing states is problematic for several other reasons as well. The most serious additional problems concern what is possible for the represented and representing states, given that they are distinct. Given the distinction between representation and target, both non-representation and misrepresentation are possible—the penultimate background premise. The mental state which represents the paper you are reading is distinct from the paper, and you might not have read the paper (non-representation) or you might be suffering a paper-reading illusion (misrepresentation). According to any version of the representational appendage thesis, the state representing your pain and the pain itself are just as distinct from one another as a representation of the paper and paper. It should therefore be just as possible for the awareness of your pain, M*, to occur in the absence of your pain, M. But this is absurd. If M existed without M*, then there would be a pain where there is no awareness of a pain. But pain is an intransitively conscious state, and intransitively conscious states are states whose subject is aware of them. Similarly, if M* existed in the absence of M, then there would be awareness of pain where there is no pain; i.e., there would be an illusion of pain. As Brock and Raymond (2006) put the point, things will seem to you just as they would if the represented item (your pain) had been real. But for pain to be real just is for things to seem a certain way to you. If things seem to you just as they would if you were in pain, then you are in pain and not merely under the illusion that you are. Since unconscious pains and pain-illusions are impossible, it is not possible for M and M* to exist in the absence of one another.

Horgan, Tienson, and Graham (2006) make a similar point. If one's pain and the mental state by virtue of which one is aware of the pain are distinct, then radical internal world skepticism should be just as genuine an epistemic possibility as radical external world skepticism. Skepticism about the external world can be very radical, indeed; it extends not only to the way in which mental states represent the objects of experience, but to the very existence of the represented objects. We recognize that things in the external world could vary wildly while the way things seem to us (the way our mental states represent the world to be) remains constant, because the mental states which represent the world to us and the states of the world they represent are distinct, and therefore can exist separately. According to the representational appendage thesis, the mental states representing the internal world are just as distinct from their targets as the mental states representing the external world are distinct from theirs. By parity of reasoning, radical internal world skepticism should be just as genuine an epistemic possibility for us. But it is not—this is the final background premise.¹⁵ It may indeed be

¹⁵ An anonymous referee noted that this premise has detractors. Some philosophers of perception entertain fairly radical forms of internal world skepticism; see, for example, William Fish's (2008) disjunctivist account of hallucinatory visual experiences. Fish allows that visual hallucinations may "have no phenomenal character whatsoever" and yet be "indistinguishable" from mental states that do have a phenomenal character (159). The position Fish describes raises many questions, which can not be settled here, concerning, for example, the truth of disjunctivism and the extend-ability of the account of hallucinatory visual experience to experiences like pain. Anything indistinguishable from a pain is, arguably, a pain; and it is unclear that even disjunctivists

possible to doubt the accuracy of the way mental states represent things within the mind. Perhaps things may seem just as they do when one believes that *p*, when in fact one only hopes that *p*; or perhaps things may seem just as they do when one is experiencing a mild itch, when in fact one is experiencing a mild burn. But it is not possible for things to seem just as they do when one is experiencing a mental state when one is in fact not experiencing anything at all. For if things seem any way at all, then one is indeed experiencing some mental state. In the case of the pounded thumb, for example, you can not doubt that you are experiencing something.

This completes the *reductio*. The supposition that there are no reflexive intentional relations, together with the background premises, entails one of three absurdities.¹⁶ The cost of avoiding the conclusion is therefore to reject one of these

have a reason to deny this. Pain in a phantom limb, for example, still hurts. Phantom limb pain involves the hallucination of a limb, not a pain. For present purposes, however, it is sufficient to note that Fish's position, in his own words, "bites the bullet" (159)—by which he presumably means that it denies something plausible. Plausibility is all that is claimed on behalf of the background premises in this paper.

¹⁶ The costs catalogued by the *reductio* are not the only costs of the supposition that there are no reflexive intentional relations. Other problems do not amount to inconsistencies, and so do not bear directly on the *reductio*. They suggest, rather, that even if some version of the representational appendage thesis proves coherent, it is not the best explanation of the representation of conscious states required by the simple representation thesis. See for example Kriegel's discussion of the immediacy of one's awareness of one's conscious experiences; Kriegel (2006: 153-158).

premises: intransitively conscious mental states are represented, human minds are finite, experienced mental states are conscious, awareness of mental states is (at least sometimes) experienced, misrepresentation and non-representation are possible for representations and distinct targets, and radical internal world skepticism is not a genuine epistemic possibility. Since each of these background premises are hard to deny, the conclusion that there are reflexive intentional relations is hard to avoid.

IV. Troubles.

The main argument of this paper is valid, and its premises true. What havoc does its conclusion wreck on causal theories of mental content? The answer, in brief, is that the argument refutes the stronger and eviscerates the weaker versions.

Causal theories of mental content are exercises in naturalistic metaphysics. They attempt to answer affirmatively Dretske's memorable question: "Can you bake a mental cake using only physical yeast and flour?" More precisely, causal theories attempt to identify the conditions under which things exhibit one mental property, intentionality (aboutness), in strictly non-mental (i.e., causal) terms. Ideally causal theories would offer necessary and sufficient conditions, and thus offer a reduction of intentional to causal relations. But, as naturalistic programs go, causal theories of mental content are unusually circumspect. Some versions aim only to identify sufficient conditions, with the hope that such conditions can be developed in various ways to account for the full range of cases (Fodor 1993: 98-99). Call any reductive causal theory—that is, any causal theory proposing necessary and sufficient conditions of intentionality in broadly causal terms—strong; all others, weak.

The argument of this paper refutes strong theories. For a strong causal theory is inconsistent with a consequence of true premises. The point may be made formally. Let A stand for the aboutness relation, and C for whatever that causal relation that best serves the causal theorist's purposes. The following statements form an inconsistent triad:

$$(1) \forall x \forall y (A(x,y) \leftrightarrow C(y,x))$$

$$(2) \forall x \forall y (C(y,x) \leftrightarrow x \neq y)$$

$$(3) \exists x \exists y (A(x,y) \ \& \ x = y)$$

The inconsistency, here, is easily demonstrated. Let a be some mental state that satisfies (3). It follows from (1) and (2) that a is distinct from the thing it is about; it follows from (3) that it is not. So we must reject one of (1)-(3). But we have shown that (2) is a conceptual truth, and the denial of (3) is absurd. So we must reject (1), which is just to say that we must reject the strong theory.

A thought about itself is thus a devastating counter-example to any causal analysis of aboutness. For it shows that causal relations are not necessary conditions of aboutness; *a fortiori*, they are not necessary and sufficient conditions. Weaker causal theories are not straightforwardly refuted by reflexive intentional relations. For weaker theories claim only sufficiency for their analysis. They are not committed to (1), but to

$$(1^*) \forall x \forall y (C(x,y) \rightarrow A(y,x))$$

the claim that a causal relation, C , of just the right sort is sufficient for an intentional relation, A . (1*) is not inconsistent with (2) and (3). For (1*) and (2) entail only that the intentional relations which have causal relations as sufficient conditions also have distinct *relata*. It does not entail that all intentional relations have causal relations as sufficient conditions. (1*) can thus accommodate the truth of (3). But at what cost?

Accommodating reflexive intentional relations puts two theoretical desiderata permanently out of reach for the weak causal theories of content. The first cost is any hope of expanding the causal theory into a comprehensive account of intentionality. The philosophical interest of weak theories presumably lies in the hope that their ability to account for the intentionality of some core cases can be extended, by the addition of further clauses, to account to cover the full range of cases. An account of the intentional properties of thought, for example, might hope to be extended to account for the intentionality of things like words and animal tracks. A weak theory which grants the truth of (3), however, gives up all hope of being extended in this way. For there is no way to extend the causal account of intentionality to cover cases where there is no causal relation. To cover such cases we need a non-causal account of intentionality; and no such account is an extension of a causal account. Furthermore, the cases of reflexive intentionality we have argued for are not peripheral cases of intentionality. They are as central a case of intentionality as there is. So the weak theory can not be extended to cover even all the core, un-derived cases.

The second cost is any hope of offering a unified, simple account of intentionality. The weak causal theory allows that the conditions that naturalize one type of intentional relation may not be precisely the same conditions that naturalize another. That is why it claims to offer sufficient but not necessary conditions. But it ceases to be a unified theory if it allows the conditions which naturalize one type of intentional relation to be completely unrelated to the conditions that naturalize another. A weak causal theory of content which grants (3) does just this. The conditions that naturalize reflexive intentionality can not be related to the causal conditions which supposedly naturalize

other cases. But this means the weak causal theory can not offer a unified theory of intentional relations. At best, it can offer only two unrelated theories, one causal, the other non-. It will have to posit two mechanisms to account for a single type of phenomena. A weak causal theory therefore accommodates reflexive intentional relations at the expense of its theoretical unity and simplicity.

Simplicity, unity and comprehensiveness; these are desiderata of any theory. To accommodate reflexive intentional relations, weak causal theories must give up on them all. Weak causal theories thus prove aptly named.

Jerry Fodor (1993: 97) generated considerable interest in the causal theory of mental content by noting that aboutness is not likely to be an entry in the physicists final catalogue of the primitive properties of matter. He therefore recommended that realists about intentionality be reductionists. “If aboutness is real,” he quipped, “it must really be something else.” Intentionality is no doubt real. But the argument of this paper shows that it is not really a causal relation.¹⁷

References

- Block, Ned, Owen Flanagan, and Güven Güzeldere (eds.) 1997: *The Nature of Consciousness: Philosophical Debates*. Cambridge, Massachusetts: MIT Press.
- Brook, Andrew and Paul Raymont. 2006: “The Representational Base of Consciousness.” *Psyche* 12, pp. 1-12.

¹⁷ Thanks to anonymous referees, Uriah Kriegel, Rob Rubert, and John Dilworth for comments which improved this paper. The faculty colloquium in the Baylor Philosophy Department devoted two sessions to the material in this paper; I am grateful for the careful attention and helpful comments my colleagues.

- Bryne, Alex. 1997: "Some Like it HOT: Consciousness and Higher-Order Thoughts." *Philosophical Studies* 86, pp. 103-129.
- Caston, Victor. 2002: "Aristotle on Consciousness." *Mind* 111, pp. 751-815.
- Chalmers, David. 2003: "The Content and Epistemology of Phenomenal Belief," in Smith and Jokic 2003, pp. 220-272.
- Chalmers, David. 2004: "The Representational Character of Experience," in Leiter 2004, pp. 153-181.
- Dretske, Fred. 1995: *Naturalizing the Mind*. Cambridge, Massachusetts: MIT Press.
- Dretske, Fred. 1988: *Explaining Behavior: Reasons in a World of Causes*. Cambridge, Massachusetts: MIT Press.
- Fish, William. 2008: "Disjunctivism, Indistinguishability, and the Nature of Hallucination," in Haddock and Macpherson 2008, pp. 144-167.
- Fodor, Jerry. 1993: *A Theory of Content and Other Essays*. Cambridge, Massachusetts: MIT Press.
- Fodor, Jerry. 1987: *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*. Cambridge, Massachusetts: MIT Press.
- Fumerton, Richard. 1995: *Metaepistemology and Skepticism*. Boston, Massachusetts: Rowan and Littlefield Publishers.
- Gennaro, Rocco, ed. 2004a: *Higher-Order Theories of Consciousness: An Anthology*. Philadelphia, Pennsylvania: Johns Benjamins Publishers.
- Gennaro, Rocco. 2004b: "Higher-Order Theories of Consciousness: An Overview," in Gennaro 2004a, pp. 1-13.

- Harman, Gilbert. 2006: "Self-Reflexive Thoughts." *Philosophical Issues* 16, pp. 334-345.
- Haddock, Adrian, and Fiona Macpherson, (eds.) 2008: *Disjunctivism: Perception, Action, Knowledge*. New York: Oxford University Press.
- Hellie, Benj. 2007: "Higher-Order Intentionality and Higher-Order Acquaintance." *Philosophical Studies* 134, pp. 289-324.
- Horgan, Terry, John Tienson, and George Graham (eds.) 2006: "Internal-World Skepticism and the Self-Presentational Nature of Phenomenal Consciousness," in Kriegel and Williford (2006).
- Kriegel, Uriah. 2005: "Naturalizing Subjective Character." *Philosophy and Phenomenological Research* 71, pp 23-57.
- Kriegel, Uriah. 2003: "Consciousness as Intransitive Self-consciousness: Two Views and an Argument." *Canadian Journal of Philosophy* 33, pp. 103-132.
- Kriegel, Uriah. 2006: "The Same-Order Monitoring Theory of Consciousness," in Kriegel and Williford (2006), pp. 143-170.
- Kriegel, Uriah, and Kenneth Williford, (eds.) 2006: *Self-Representational Approaches to Consciousness*. Cambridge, Massachusetts: MIT Press.
- Leiter, Brian (ed.) 2004: *The Future of Philosophy*. New York: Oxford University Press.
- Levine, Josheph. 2001: *Purple Haze: The Puzzle of Consciousness*. New York: Oxford University Press.
- Lewis, David. 1986: "Poscripts to Causation," in *Philosophical Papers*, vol. II. New York: Oxford University Press.

- Lycan, William. 2006: "Representational Theories of Consciousness," in *The Stanford Encyclopedia of Philosophy (Winter 2006 Edition)*, Edward N. Zalta (ed.), URL = <http://plato.stanford.edu/archives/win2006/entries/consciousness-representational/>
- Natsoulas, Thomas. 1993: "What's Wrong with Appendage Theory of Consciousness." *Philosophical Psychology* 6, pp. 137-145.
- Seager, William. 1994: "Dretske on HOT Theories of Consciousness." *Analysis* 54, 270-276.
- Smith, Quentin, and Aleksander Jokic, (eds.). 2003: *Consciousness: New Philosophical Perspectives*. New York: Oxford University Press.
- Rosenthal, David. 1997: "A Theory of Consciousness," in Block, Flanagan and Güzeldere (1997), pp. 729-754.
- Rosenthal, David. 1993: "Higher-Order Thoughts and the Appendage Theory of Consciousness." *Philosophical Psychology* 6, pp. 155-166.
- Rosenthal, David. 1986: "Two Concepts of Consciousness." *Philosophical Studies* 49, 329-359.
- Stampe, Dennis. 1977: "Toward a Causal Theory of Linguistic Representation." *Midwest Studies in Philosophy* 2, 42-63.
- Tye, Michael. 1999: *Ten Problems of Consciousness*. Cambridge, Massachusetts: MIT Press.
- Williford, Kenneth. 2006: "The Self-Representational Structure of Consciousness," in Kriegel and Williford 2006, pp. 111-142.