

What (in)coherence is not¹

Julian Fink
University of Bayreuth
Philosophy
Universitätsstraße 30
D-95447 Bayreuth
Germany

ABSTRACT

Two or more attitudinal states are incoherent if and only if, necessarily, under conditions of full transparency, you are disposed to give up one of them. This is roughly the account of incoherence that has recently been put forward by Alex Worsnip (2018). In this paper, I argue that Worsnip's account of incoherence is incorrect. Not only does it fail to be fully general (i.e., it does not allow us to assess the coherence of all combinations of attitudes), but it also implies that an agent is necessarily incoherent just in virtue of being in one attitudinal state.

Worsnip's account of incoherence

Irrationality, or at least large parts of it, are commonly identified with incoherent patterns of attitudes. As it stands, however, the debate on coherence-based rationality is beset with isolated, "one-by-one" explications of when a pattern of attitudes counts as incoherent (cf. Broome 2013, 130). It would indicate great progress in the debate if we could systematically demarcate coherent from incoherent combinations of attitudes.

Indeed, in a very recent paper, Alex Worsnip (2018) has attempted to rectify this situation by developing an account of what it is "for two or more mental states to be jointly incoherent, such that they are banned by a coherence requirement [of

¹ I would like to thank John Broome, Franzika Poprawe, Andrew Reisner, Benjamin Kiesewetter, and Cristina Borgoni for insightful comments on earlier versions of this paper.

rationality]” (2018, 185). Put roughly, Worsnip claims that two or more of your mental states are jointly incoherent and irrational if you are necessarily prone to abandon at least one of them once you become aware that you hold them at the same time. Or, in Worsnip’s own words:

A set of attitudinal mental states is jointly [sic]² incoherent iff it is (partially) constitutive of the mental states in question that, for any agent that holds these attitudes, the agent is disposed, when conditions of full transparency are met, to give up at least one of the attitudes. (2018, 188)

I will refer to this as “Worsnip’s account”. Worsnip qualifies the account in three important ways (2018, 188). First, as he mentions, the disposition to give up at least one of the attitudes must be sensitive to a constitutive aspect of the states involved. That is, it must be one that comes with what it *is* to be in a particular attitudinal state.³ In the following, I will use “disposed*” to pick out the exact type of disposition Worsnip has in mind here. Moreover, Worsnip also states that the disposition must have considerable force. It must be such that the agent would “(at least normally) not [be] able to (or at least find it difficult to) sustain such combinations of attitudes under conditions of full transparency” (2018, 188).

Second, “attitudinal mental state” is meant to include states in which mental attitudes are *absent*. For example, the combination of a belief that you ought to *a* and

² It is the attitudes within the set that are jointly incoherent, not the set itself.

³ Here is an example of a disposition that is *not* sensitive to a constitutive aspect of a set of attitudinal states. Suppose you are in a state of being scared of becoming obese. This disposes you to abandon either your intention to consume sugar or your intention to consume butter if you become aware that you have both intentions. However, this is not the kind of disposition envisioned by Worsnip. Instead, the disposition must be one that is sensitive to a constitutive aspect of the attitudes involved.

the absence of an intention to *a* counts as a combination of two attitudinal states. Third, subscribing to *s* under “conditions of full transparency” means being maximally aware that you subscribe to *s*. This entails that “the agent knows, and explicitly and consciously believes, that she has the states in question, without self-deception, mental fragmentation, or any failure of self-knowledge (pertaining to those attitudes)” (2018, 188).

Worsnip illustrates his account by showing how it is able to identify instrumental incoherence, akratic incoherence, and the incoherence that stems from intransitive preferences (2018, 190–94 and 198–200). To save space, I will restrict my illustration to one example.

Suppose you (*i*) intend to go shopping, (*ii*) believe that you will go shopping only if you leave your house, yet (*iii*) have no intention to leave your house. Worsnip posits that “part of *what it is* to intend an end [is to be] disposed, under conditions of full transparency, to form corresponding intentions to intend the means that one believes to be necessary to that end (or to give up the intended end)” (2018, 191). Consequently, (*i*), (*ii*), and (*iii*) are such that, because of an essential disposition that comes with these states, at least once you become aware of them, you will be strongly disposed* to give up at least one of (*i*), (*ii*), or (*iii*).⁴ And this is precisely what constitutes their incoherence, according to Worsnip. Moreover, it is precisely this type

⁴ I have followed the structure of the example offered by Worsnip (2018, 190–91). That is, (*i*) you intend *a*, (*ii*) you believe that [*a* only if *b*], yet (*iii*) you do not intend *b*. However, there is a problematic glitch in the structure of the example. Obviously, you will not be necessarily disposed to abandon one of these attitudinal states (even under full awareness) if you also believe that *b* will obtain even if you do not intend *b*. A better strategy for Worsnip would therefore be to replace (*ii*) with (*ii'*), i.e. you believe that [*a* only if *you intend b*].

of disposition that constitutes all incoherent combinations of attitudes (that violate a coherence-based requirement of rationality). Or so Worsnip argues.

A dilemma

I welcome Worsnip's attempt to develop "a general account as to what it is for mental states to be jointly coherent" (2018, 186). I doubt, however, that the account he presents can achieve this. In the remainder of this paper, I argue that Worsnip's account faces a fatal two-horned dilemma. On the first horn, the account does not qualify as fully general. It cannot assess *all* possible patterns of attitudes for coherence. As it stands, and as I will explain, this is mainly a consequence of Worsnip's restricted notion of awareness. On the second horn of the dilemma, however, I demonstrate that restoring the full generality of Worsnip's account makes it prone to a further set of defects. I will discuss two distinct proposals in this direction. Although both restore generality, they make the account even less defensible, or so I argue.

Here is the first horn of the dilemma. Worsnip claims that a set of jointly held attitudes is incoherent if and only if, necessarily, you would be disposed* to abandon at least one of the attitudes if you were to become aware of the entire set. Note that Worsnip, as explained above, construes awareness as *knowledge* and *explicit and conscious belief*. That is, you are aware of an attitude if and only if you explicitly and consciously believe and know that you are in a particular attitudinal state. I will refer to this conception as "awareness*".

This construal of awareness* comes with a significant problem, however. There are combinations of attitudinal states of which, conceptually, we cannot become aware*.

In “Dispositions and Conditionals”, C.B. Martin (1994) famously objected to a standard counterfactual analysis of dispositions by pointing to so-called “finkish” dispositions. These dispositions are “finkish” because their stimulus conditions are identical to a condition in which an object (or person) loses or gains that disposition. In the context of Worsnip’s account, a somewhat analogous objection can be made. I argue that there are attitudinal states that are “finkish” regarding Worsnip’s conception of awareness. That is, there are attitudinal states that exclude awareness; you cannot be in these states while simultaneously being aware of them.

There are countless examples of this phenomenon. Suppose you are in an attitudinal state in which you have no genuine belief. While it is possible to be in such a state, it is not possible to remain in that state while becoming aware* of it. Awareness* entails believing and knowing. So, once you become aware* of the state, you are no longer in it. Hence, you cannot know that you are in it.

Likewise, suppose you know that you cannot add another higher-order belief to your state of mind. Again, you cannot become aware* of the state, for if you know that you cannot add another higher-order belief to your state of mind, then it is true that you cannot add another higher-order belief to your state of mind. Thus, you cannot come to believe that you know that you cannot add another higher-order belief to your state of mind, for that would be the addition of another higher-order belief.

As a result, Worsnip's account is unable to establish the coherence of all combinations of attitudinal states. It is thus less than fully general, as I shall put it. It is therefore also less than fully convincing.

Of course, Worsnip may not worry about these two particular examples. He may argue that an account of incoherent attitudinal states should only apply (i) to *believers* (i.e., subjects who hold beliefs) and not to (ii) subjects who are cognitively impaired to the extent that they cannot adopt a higher-order belief. Even so, however, I believe that Worsnip will and should worry about what I take to be the most unfavourable example for his account.

Suppose you intend to fix your car, but it is not the case that you believe that you intend to fix your car. Suppose you become aware* of these two attitudinal states. One consequence of this is clear. You now believe that you intend to fix your car. And so, you no longer *do not* believe that you intend to fix your car. In the very process of your becoming aware* of the attitudinal states we wish to examine for coherence, they necessarily disappear.

This is a devastating result for Worsnip's account. If Worsnip is right, then we can never assess whether lacking a belief about an attitude of yours is (in)coherent. Recall that what we want to determine is whether you are *psychologically* disposed* to sustain [an attitude and the absence of a second-order belief about that attitude] under awareness*. But we cannot determine this; conceptually, you cannot be in a state where you lack the relevant second-order belief under awareness*. This

limitation of generality and explanatory scope is unacceptable.⁵ Worsnip's account suffers from a design failure.

Before I turn to the second horn of the dilemma, let us look at a quick objection to the effect that this design failure may not really harm Worsnip's account. For one could argue that this design failure is *de facto* significant only if (a) there is a combination of attitudes, say *M*, that is (suspected to be) incoherent and irrational, and (b) it is impossible to adopt *M* while being aware* of *M*.

Indeed, as I will argue below, I believe that intending to fix your car while not believing that you intend to fix your car does *not* qualify as an irrational combination of attitudes. Hence, the fact that Worsnip's account does not allow us to assess this combination may not turn out to be a significant defect. It could be true that any combination of attitudes of which one cannot become aware* is simply excluded from being irrational.

This is doubtful, however. It seems that there are combinations of attitudes that Worsnip's account does not allow us to test for rationality but for which it is at least an open question whether they are irrational.

Here is a simple example that shows this. Suppose you believe *a* and believe that if you believe *a*, then you believe that you believe *a*. Yet you do not believe that you

⁵ So far, I have said that the consequence of this is that some attitudinal states are not up for assessment on Worsnip's account. Intuitively, this is correct. Strictly or technically speaking, however, the situation may be worse, for if it is necessarily impossible to be aware of a given attitudinal state, then this means that the antecedent of the conditional statement that forms the right side of Worsnip's account turns out to be necessarily false. On most standard semantics of necessary (or counterfactual) conditionals, this implies that the conditional that forms the right side of Worsnip's account turns out to be *true* – which in turn implies that as soon as one is in an attitudinal state of which one cannot be aware*, one is automatically incoherently irrational. This is of course absurd.

believe *a*. No doubt, this is a combination of attitudes the rationality or irrationality of which a theory of irrationality should be able to establish; *prima facie*, your beliefs seem to violate a plausible *modus ponens* requirement of rationality. However, since you can never become aware* of this combination of attitudes, Worsnip's account does not allow us to establish this. This underlines that the design failure Worsnip's account suffers from is significant and should not be ignored.

I now turn to the second horn of the dilemma. Recall that as it stands, Worsnip's account does not qualify as fully general. The account cannot assess *all* possible patterns of attitudes for coherence. On the second horn of the dilemma, however, I demonstrate that restoring the full generality of Worsnip's account makes it prone to a further set of objections. I will discuss two sets of distinct proposals in this direction.

Here is the first proposal for saving Worsnip's account from finkish states (no pun intended). Let us assume that (irrational) incoherence requires the possibility of awareness*. If, conceptually, you cannot become aware* of a set of attitudes, then the set cannot be incoherent. Worsnip's account would thus read as follows: a set of attitudes is incoherent if and only if, under awareness*, you are disposed* to give up at least one of the set's attitudes *unless* it is conceptually impossible to become aware* of those attitudes. In that case, a set of attitudes is *not* incoherent.⁶

Indeed, the addendum restores the full generality of Worsnip's account. It can now assess the in/coherence of all possible sets of attitudes. Contra Worsnip's original intention, however, the in/coherence of attitudes is no longer *exclusively* a question

⁶ I am grateful to an anonymous referee for suggesting this point to me in such a clear and intelligible manner.

of a disposition* to sustain them under awareness*; it now also depends on the *possibility* of becoming aware of a combination of attitudes.

Nevertheless, one may still be able to construe a philosophical justification of the amendment. One may argue, for instance, that irrationality implies accountability. Yet you can only be held accountable for attitudes of which you can become aware*. Thus, a set of attitudes of which you cannot become aware* cannot be irrationally incoherent.

However, I argue that this proposal remains *ad hoc*. I offer two reasons for this. First, the added condition may make the account prone to neglect core conditions of irrational incoherence. I doubt, for instance, that you can become aware* of a set of contradictory beliefs. Likewise, I doubt that you can become aware* of a set of contradictory intentions. That is, you cannot consciously entertain both a belief (or intention) that *a* and a belief (or intention) that not-*a*. If that is correct, the amended account fails to pick out these attitudes as incoherent.

Second, suppose you intend to fix your car, you do not believe that you intend to fix your car, *and* you intend [to believe that you intend to fix your car if you intend to fix your car]. I assume that this combination of attitudes is incoherent. You fail to display a type of *strength of will* that rationality requires. Or alternatively, suppose you believe *a*, you believe that *a* implies that you believe that you believe *b*, you believe *b*, yet you do not believe that you believe *b*. I likewise assume that this

combination of attitudes is incoherent. You fail to close your beliefs under *modus ponens*.⁷

In both cases, however, it is conceptually impossible to become aware* of these mental states. As soon as you become aware* that you intend to fix your car, you also believe that you intend to fix your car; as soon as you become aware* that you believe *b*, you also believe that you believe *b*, and thus you inevitably exit the incoherent state. It is thus less than convincing to make the possibility of awareness* a prerequisite of irrational incoherence. Consequently, the amended account genuinely fails to identify incoherent combinations of attitudes. It restores the generality of Worsnip's account in an *ad hoc* fashion.

I now turn to a second attempt to restore the generality of Worsnip's account. As it will become apparent below, this response adds to the second horn of the dilemma confronting Worsnip's account. Instead of adding that the impossibility of awareness* implies coherence, Worsnip could abandon his specific notion of awareness*. He could operate with a notion of awareness that does not require knowledge or belief. On such a notion, you could possibly be aware of an attitudinal state *s* without knowing or believing that *s*.⁸

Strictly speaking, this move would not fully eliminate the revealed design failure, for even on such a thin notion of awareness, awareness would still be a *factive* state.

⁷ If you do not believe that rationality requires closing your beliefs under *modus ponens*, you could, for example, add that you *care* about whether you believe that you believe *b* (Broome 2013, 157).

⁸ Frankly, I do not know how and whether such a notion of awareness could be developed. Perhaps it could be akin to the concept of perception that does not entail belief (which is also controversial). At any rate, the onus would be on Worsnip to present us with such a conception of awareness.

That is, you can be aware of s only if s . Hence, Worsnip's account would still not be able to assess the (in)coherence of a state that includes lack of awareness *tout court*. But perhaps this is a bullet one should be prepared to bite. (After all, being aware of something could arguably be a minimal condition for having some degree of rationality, as opposed being *arational*.)

However, I argue that resorting to this relaxed (and almost inconceivable) notion of awareness represents another implausible *ad hoc* attempt to restore the generality of Worsnip's account. (I will refer to Worsnip's account under this relaxed notion of awareness as "Worsnip's relaxed account".) Relaxing the notion of awareness again introduces more problems than it solves.

Recall the situation where you intend to fix your car yet you have no belief that you intend to fix your car. Since awareness no longer requires belief, you can become aware of being in that attitudinal state. But what does Worsnip's relaxed account say about the coherence of these jointly held attitudinal states?

Suppose you are aware that you intend to fix your car and that it is not the case that you believe that you intend to fix your car. More generally: you are aware of a state s and you are aware that you do not believe s . Can you sustain both attitudinal states? While conceptually conceivable, this would be psychologically paradoxical. Indeed, this situation mimics Moore's paradox. Suppose you were to express the contents of your awareness. You would declare "I intend to fix my car, yet I do not believe that I intend to fix my car". It is hard to conceive of how, after such a declaration, you could psychologically maintain these two attitudinal states, or not even be disposed* to abandon one of them, at least without mental fragmentation or

in the absence of some cognitive defect. Moreover, that disposition would indeed be tied, in the right sort of way, to a constitutive aspect of the involved belief (i.e. being regulated by truth or evidence).

The result is clear. On Worsnip's relaxed account, if you intend to fix your car and do not believe that you intend to fix your car, you are incoherent. Likewise, to give an analogous example with an absent attitude, if you lack the belief that water can be turned into wine and also have no belief that you lack this belief, you are again incoherent. Or in general, for any attitudinal state *s*, if you are in *s* and do not believe that you are in *s*, you are incoherent.

This is implausible, if not absurd. Coherence cannot require us to back up, as it were, every attitude we have with a belief that we have it. In fact, I fail to see what is *per se* incoherent about believing, intending or desiring *a* and not believing that you believe, intend or desire *a*. Any of these attitudes may be perfectly successful or functional even if you do not have a higher-order belief about them. (Of course, do not confuse this with having a particular attitude and believing that you do not have it. That, I would argue, is a clear case of incoherence.)⁹

⁹ Here is a brief summary of my argument so far and the dilemma I have drawn up in this paper. Suppose, for example, you intend to *a*, yet you do not believe that you intend to *a*. Are you irrationally incoherent? On the first horn of the dilemma, I argued Worsnip's account is unable to answer this question. Worsnip's conception of transparency does not allow us to test if an agent could ever sustain this combination of attitudes under transparency. I showed why this is significant flaw of Worsnip's account.

On the second horn of the dilemma, having now adjusted and relaxed Worsnip's notion of transparency to deal with the problem above, I show that Worsnip's account is indeed in the position to establish whether intending to *a* and not believing that you intend to *a* is irrational. Yet, it does not get it right. The relaxed account implies that intending *a* while not believing that you intend to *a* is irrational – though it is clearly not. In sum, Worsnip's account faces the following dilemma: for some combination of attitudes *C* that are *not* irrational, Worsnip's account is either unable to establish that *C* is not irrational or (once adjusted to deal with this problem) it falsely implies that *C* is irrational.

However, suppose you still believe that this does not derail Worsnip's relaxed account. Suppose you endorse the view that lacking a belief about an attitudinal state of yours involves some kind of incoherence. In this case, consider the following implication of the argument I have just offered.

Suppose you are incoherent by dint of not believing that you are in an attitudinal mental state, say *A*, that you are actually in. It is an illusion to think that you can save yourself from being incoherent by coming to believe that you are in state *A*. That won't do, for if you come to believe that *A* is a mental state of yours, you are now in another attitudinal state – call it *B(A)* – for which it holds, again, that as long as you do not believe that this is a state of yours, you are incoherent. However, believing *B(A)* will not save you from being incoherent, for if you come to believe that *B(A)*, you are yet again in another state – call it *B(B(A))*, for which it holds once more that, as long as you do not believe *B(B(A))*, you are incoherent. And so on, *ad infinitum*.

The consequence of this is clear. Worsnip's relaxed account in fact implies that necessarily, as soon as you are in *one* attitudinal state (which could even be the state of not having an attitude), you are *incoherent*. And no matter what states you add to your situation, you will never become coherent. Arguably, the only realistic way¹⁰ to

¹⁰ I am not considering here the possibility of adding an *infinite* number of higher-order states to your mental states. I assume that our minds are limited and do not allow us to add an infinite number of attitudinal states. But suppose you think otherwise. Even so, Worsnip's relaxed account still comes with a set of implausible implications. First, it implies that there is an infinite set of rational coherence requirements that require us to clutter our minds with an infinite number of pointless beliefs (cf. Harman 1986, 12). Second, it implies that we are necessarily subject to at least one rational coherence requirement that we cannot satisfy. Suppose you are in an attitudinal mental state *s*, yet you do not believe that you are in that state. Then you violate a requirement of rationality. However, suppose you now form a belief that you are in an attitudinal mental state *s*. Then you satisfy that requirement of rationality, yet you will violate another, unless you also believe that you believe that you are in *s*. However, even if you have *that* belief, you will again violate another requirement, unless you have the next higher-order belief, and so on. Note too that exiting the attitudinal mental state *s* would not stop

avoid being incoherent is to be *dead* – and even that works only if being dead means that no attitudinal state can be ascribed to you.

On that daunting and morbid note, I conclude that Worsnip's account, as well as the alternative versions I have discussed in this paper, cannot resolve the question of whether (and when) a combination of attitudes is incoherent.

the infinite iteration of requirements, for then rationality would require you to believe that you are *not* in *s*. As just described, satisfying that requirement would trigger another requirement, whose satisfaction would trigger another requirement, and so on. So, even by forming an infinite number of beliefs about the states you are in, you would still always be violating at least one rational coherence requirement. Or put succinctly: Worsnip's relaxed account implies that for every number of requirements you satisfy, say *n*, you are at least subject to *n*+1 requirements. I take this as further evidence that Worsnip's relaxed account cannot capture the essence of attitudinal (in)coherence.

References

- Broome, John 2013. *Rationality Through Reasoning. The Blackwell/Brown lectures in philosophy*. Malden, Mass.: Wiley Blackwell.
- Harman, Gilbert 1986. *Change in View: principles of reasoning*. Cambridge, Mass.: MIT Press.
- Martin, Charles, Burton, 1994. "Dispositions and Conditionals." *Philosophical Quarterly* 44, 1–8.
- Worsnip, Alex 2018. "What is (In)Coherence?", In: *Oxford Studies in Metaethics 13*, edited by Russ Shafer-Landau, Oxford: Clarendon Press, 184-206.