

# A solution to the problem of hijacked experience

Jill is not sure what Jack's current mood is, but she fears that he is angry with her. Then Jack steps into the room. Jill gets a good look at his face. What she sees are various facial features that are in fact characteristic of a blank stare, and in good conditions she would recognize them as such. However, because Jill is afraid, she jumps to conclusions *at the experiential level*: she *experiences* Jack's face as angry. That is, she experientially takes the features that in fact indicate a blank stare to indicate anger. Finally, on the basis of this experience, Jill comes to believe that Jack is in fact angry with her.<sup>1</sup>

Here are two contradictory claims about Jill's final belief – her belief that Jack is angry with her:

- (1) Jill's final belief is rational.
- (2) Jill's final belief is not rational.

Which of these claims is true?

On the one hand, it is tempting to accept (1), on the grounds that once Jill has a perceptual experience as of an angry face, there is no other response that Jill could rationally have given. For instance, suppose that Jill had responded to her perceptual experience by forming the belief that Jack was confused. It is hard to see how this could be a rational response to an experience that has already classified Jack's face as angry. Or suppose that Jill had responded to her perceptual

---

<sup>1</sup> I draw this case, and the problem that it generates, from Siegel (2017, especially pp. 117-119).

experience by suspending judgment about what emotion Jack was feeling. Given that Jill has gotten a good look at Jack's face and has no reason to think that her perceptual experience is misleading, this response would be irrationally timid.

On the other hand, it is tempting to accept (2), on the grounds that Jill's final belief is based on a perceptual experience, which is in turn formed via something akin to jumping to conclusions. For Jill has a perceptual experience which successfully detects various facial features that are indicative of Jack's confusion, but, because she is afraid that Jack is angry, she perceptually takes Jack to be angry. This perceptual taking seems not to be rational, and thus her belief that Jack is angry – which is based on her perceptual taking – also seems not to be rational.

The fact that (1) and (2) both seem to be plausible, even though these claims are inconsistent, constitutes the *problem of hijacked experience*.

In her recent book *The Rationality of Perception*, Susanna Siegel claims that she can solve the problem. Her proposal is to reject (1) while keeping (2). And she defends this solution by offering a variety of arguments for the novel and intriguing claim that perceptual experiences can be the products of inferences and thus can themselves be rational or irrational. This allows her to say that Jill's perceptual experience is formed on the basis of a bad inference, so this perceptual experience is irrational; the irrationality of the experience is then transmitted to her final belief.

Suppose, for the sake of argument, that these claims of Siegel's are all correct. I submit that this does not yet solve the problem. For all that Siegel has done is buttress the argument in favor of (2) – but nowhere in her book does she rebut the argument for (1). Until we have such a rebuttal, we will not understand *why* (1) is false, and the problem stands.

In this paper, I will attempt to fill this lacuna on Siegel's behalf, while supposing, just for the sake of argument, that Siegel's defense of (2) is correct. My focus will be on the central

intuition that is supposed to support (1): the intuition that no matter what the subject's epistemic situation is, there must be *some* rational way for her to respond. I will show that we can accommodate this intuition while replacing (1) with a nearby claim in a similar spirit. This nearby claim is, in fact, compatible with (2).

Thus, by the end of this paper, I will offer a full resolution of the problem of hijacked experience.

## 1. The intuition intended to support (1)

Jill has a perceptual experience that classifies Jack's face as angry, and she is considering the question of what emotion that face in fact expresses. Once Jill has landed herself in this epistemic situation, it seems that she must choose from the following responses:

*Response A.* Believe that Jack is not angry.

*Response B.* Suspend belief about whether Jack is angry.

*Response C.* Believe that Jack is angry.

Now, it is plausible that Response A would be irrational, given that Jill already perceptually takes Jack to be angry. To be sure, we are assuming, with Siegel, that this perceptual taking is irrational. But Jill does not know, and indeed *cannot* know, that this is the case. And once Jill perceptually takes Jack to be angry, while having no evidence that this perceptual taking is irrational, it is hard to see how she could rationally arrive at the belief that Jack is not angry.

For similar reasons, it is plausible that Response B would be irrational. Again, we are supposing that Jill perceptually takes Jack to be angry, while having no evidence that this

perceptual taking is irrational. Thus, if Jill were to suspend belief, she would have to ignore the way that she perceptually takes Jack to be, while having no good epistemic reason to ignore this. And that seems to be irrational.

With this in the background, return to:

(1) Jill's final belief is rational.

We are now in a position to appreciate the full force of the argument for (1). The argument relies on the following claim:

*No Hopeless Situations*: In any epistemic situation, there is at least one rational way for the subject to respond.

Apply this idea to Jill's case. Jill has only three possible responses, and we have just seen that the first two are not rational. Thus, the third response must be a rational one – or so the argument goes.

I find No Hopeless Situations to have a high degree of immediate plausibility. For one thing, rationality is supposed to be a guide, even in cases where the subject has already made a rational mistake. So surely it will tell subjects like Jill what they ought to do, rationally speaking. Moreover, it will emerge in §2 that No Hopeless Situations is supported by a plausible picture of why it is worth making assessments about epistemic rationality in the first place. Of course, it is not *impossible* to reject No Hopeless Situations. But it would be better, all else equal, to find a way to accommodate it.

The trouble is that we are assuming, for the sake of argument, that the following claim is true:

(2) Jill's final belief is not rational.

But how can we square (2) with No Hopeless Situations?

## 2. The problem solved

My suggestion is that we can answer this question, and thereby solve the problem, by making a distinction between a rational *belief-forming process* and a rational *belief*. The idea is that when Jill forms the belief that Jack is angry, Jill is using a rational belief-forming process, but she thereby arrives at an irrational belief. Let us call this the *two-factor view*.<sup>2</sup>

Is the two-factor view even coherent? To see that it is, notice that a single belief can be the end result of many different belief-forming processes. For example, I might form the belief that  $p$  immediately on the basis of perception (first process). I might then form the belief that if  $p$ , then  $q$  via an inference to the best explanation from various background pieces of evidence (second process). Finally, I might conclude that  $q$  via modus ponens (third process). In this case, my belief that  $q$  is the end result of at least three distinct processes, each of which might contribute in some way to the rationality of the belief itself. Thus, it is perfectly coherent to say that one of the processes – say, the final process of using modus ponens – is rational, while also saying that the belief formed by this process is irrational. In particular, this is coherent because we can say that

---

<sup>2</sup> For discussion of a very similar distinction, see Kolodny (2007).

the irrationality of the belief is due to some irrationality *elsewhere* in the *total* process of forming that belief.

So the two-factor view is coherent. But is it correct? I offer three reasons to think that it is.

First, the view gives us an attractive picture of why it is worthwhile to make assessments about epistemic rationality in the first place. Perhaps an analogy will help. Imagine a factory that contains a series of machine: the first machine is supposed to fold sheets of metal into open cans, the second machine is supposed to take these open cans and drop fruit into them, the third machine is supposed to pour preservative juices over the fruit, and the fourth machine is supposed to weld a lid over the filled cans. But when you arrive on site, you observe that after some time, the second machine goes haywire: after dropping fruit into the open cans, it knocks the cans off the conveyor belt. As a consequence, the third machine starts pouring preservative juices directly onto the belt, and the fourth machine drops metal lids onto the belt.

Now imagine that your job is to assess this factory on behalf of the canning company. Then there will be at least two things that you will want to do. First, you will want to evaluate the *quality of the final product*, which in this case is obviously poor. But, second, you will want to assess *precisely what went wrong* in the process that yielded the bad product, so that you can effectively intervene to correct the process. And in this factory, it is only the second machine that is out of order. *Even though* the third and fourth machines are not producing the correct products at the relevant stages of the process, they are still functioning properly.

I submit that it is promising to theorize about epistemic rationality in a similar way. As epistemologists, we want to assess the rationality of various cognitive *products* – beliefs, credences, and (if Siegel is right) perceptual experiences. But, if a cognitive product is irrational, we should also care about precisely which cognitive *processes* went wrong, so that we can

effectively intervene on just those processes. That is why we should not classify as irrational a process that is functioning well epistemically, *even if* that process has an irrational product: such processes are not ordinarily the right ones to intervene on.<sup>3</sup>

My claim is not that you can *never* effectively intervene on a cognitive process that is epistemically well-functioning. Sometimes you can: you might modify an epistemically good process to make it more likely to yield epistemically good products even if it is given epistemically bad products as input. For instance, you might add a double-checking procedure to an already epistemically good process. My claim is just that it is very often useful to know precisely what processes will tend to produce epistemically bad outputs *even when* they are given epistemically good inputs. Thus it is useful to have a label – “irrational” – to distinguish these processes from others.

The first reason to accept the two-factor view, then, is that it helps us to explain why it is worth making assessments about epistemic rationality in the first place: such assessments are frequently useful for telling us how to make effective epistemic interventions.

A second reason to accept the two-factor view is that it gives independently plausible verdicts about non-perceptual cases. Return to the case in which the subject infers that  $q$  from (i) her belief that  $p$  and (ii) her belief that if  $p$ , then  $q$  by using (iii) *modus ponens*. Suppose that the subject’s belief that  $p$  is rational, while the subject’s belief that if  $p$ , then  $q$  is formed through a

---

<sup>3</sup> What I say here is broadly in the spirit of Kolodny (2007). But there is one important difference. Kolodny claims that “rational requirements ... are process requirements” (p. 371), by which he seems to mean that *all* rational requirements are process requirements. And Kolodny takes his opponent Broome to hold that all “rational requirements are state [i.e., product] requirements” (ibid). By contrast, my suggestion is that there are rational requirements of both kinds – process requirements and product requirements – and it is important to consider them separately.

faulty attempt at inference to the best explanation and so is not rational. Finally, suppose that when the subject then forms the belief that  $q$ , her use of modus ponens is fully competent.

The two-factor theorist says that it is perfectly rational for the subject to infer from the premises that  $p$  and that if  $p$ , then  $q$  to the conclusion that  $q$ . That is independently plausible, since it is independently plausible that it is rational to competently make a deductively valid inference. The two-factor view also has it that the subject's belief that  $q$  is not rational. This, too, is plausible, for her belief that  $q$  relies essentially on her irrational belief that if  $p$ , then  $q$ .

Finally, return to the problem of hijacked experience, which is generated by the following claims:

- (1) Jill's final belief is rational.
- (2) Jill's final belief is not rational.

And recall that (1) is supported by:

*No Hopeless Situations*: In any epistemic situation, there is at least one rational way for the subject to respond.

A third and final reason to accept the two-factor view is that it solves the problem of hijacked experience in a satisfying way. In particular, it lets us reject (1) while maintaining the plausible claim underlying (1) – namely, *No Hopeless Situations*.

To begin with, the two-factor view lets us reject (1). Before we can appreciate why this is true, we should recall that we are assuming for the sake of argument that Siegel has the correct

view of hijacked experience. In particular, we are assuming that Jill accurately perceives Jack's facial features, and that she then irrationally *perceptually* takes Jack to be angry. And we are also assuming that, on the basis of this irrational perceptual taking, Jill forms the belief that Jack is angry, which inherits the irrationality of the perceptual taking.

The two-factor view is compatible with rejecting (1) in this way. For the view says that when we evaluate the rationality of Jill's final belief, we should look at the *overall* process that resulted in that belief. And, since the overall process essentially included an irrational inference – the irrational inference that led to Jill's perceptually taking Jack to be angry – Jill's final belief is irrational.

At the same time, however, the two-factor view lets us retain:

*No Hopeless Situations*: In any epistemic situation, there is at least one rational way for the subject to respond.

In particular, consider Jill's epistemic situation once she has – irrationally – perceptually taken Jack to be angry. Jill *does* have a rational response in this situation, and it is precisely the response that she makes: to believe, on the basis of this experience, that Jack is angry. This *response* – to infer from the experience to the belief – is rational. It is just that the *product* of that response – Jill's final belief – is not rational, due to an earlier irrationality in Jill's cognitive processes.

Siegel's proposed solution to the problem of hijacked experience is incomplete insofar as it does not reveal where the argument for (1) goes wrong. The two-factor view fills this lacuna: it lets us reject the inference from No Hopeless Situations to (1). Thus, we now have a genuine

solution to the problem. And, since there are independent reasons to accept the two-factor view, there are independent reasons to adopt this solution.

### 3. Some exegetical remarks

As an aside, I will make two brief exegetical points about how the two-factor view relates to what Siegel herself says. The first point is that Siegel seems to be officially committed to rejecting the two-factor view. For she considers cases in which an otherwise good inference operates on an epistemically bad input, and she repeatedly describes *the inference itself* as “bad” or “poor.”<sup>4</sup> But, on the two-factor view, there is nothing (rationally) bad or poor about such inferences. There is just a rational flaw in the input to the process, which results in a rationally flawed output, but *the inference itself* is perfectly rational.

My second point, however, is that Siegel’s rejection of the two-factor view is entirely unnecessary for her central philosophical purposes. As far as I can tell, if Siegel were to adopt the two-factor view, she would need to retract only the handful of claims that I described in the previous paragraph – claims which play no further role in her book. She could continue to make all of her other central claims. For example, as we have already seen, she could continue to claim that Jill’s perceptually taking Jack to be angry is the product of a bad inference; that Jill’s overall perceptual experience is therefore irrational; and that Jill’s final belief based on this perceptual experience is also irrational. Siegel would only need to add that the inference from the perceptual experience to the belief is rational.

---

<sup>4</sup> See, e.g., Siegel (2017, p. 109 and p. 111).

Thus, I consider the two-factor view to be a friendly amendment to Siegel's view, one that would fill out, in a plausible and independently-motivated way, her solution to the problem of hijacked experience.

## 4. Conclusion

The problem of hijacked experience is that the following contradictory claims both seem to be plausible:

- (1) Jill's final belief is rational.
- (2) Jill's final belief is not rational.

In this paper, I have proposed a true *solution* to this problem – a view that not only rejects one of these claims, but that also reveals where the argument for that claim goes wrong. In particular, my proposal is to reject (1). But, at the same time, I have shown how to retain the claim that motivates (1), namely:

*No Hopeless Situations*: In any epistemic situation, there is at least one rational way for the subject to respond.

The two-factor view accommodates this view by allowing that the subject might make a rational *response*, such as an inference, that – due to some earlier irrationality – still yields an irrational *product*, such as a belief or a perceptual experience. Besides solving the problem of hijacked

experience, the two-factor view also gives plausible verdicts about particular cases and offers an appealing explanation for why it is worth assessing epistemic rationality in the first place.

This may not be the only solution to the problem of hijacked experience. But it is a very attractive one.

Moreover, it so happens that the two-factor view is relevant to another philosophical debate: the debate over whether or not certain requirements of rationality have *wide scope*. To illustrate, consider someone who has overwhelming evidence that *not-p*, but who nevertheless believe that *p*. And suppose that she notices that *p* entails *q*, and she is considering whether or not to believe that *q*. Is this subject rationally required to believe that *q*?

A *wide-scope* theorist will think that she is not. For, according to the wide-scope theorist, there is only a rational requirement for her *either* to believe that *q* *or* to abandon her belief that *p*. And, since the subject's evidence overwhelmingly supports the belief that *not-p*, she in fact is rationally required to abandon her belief that *p*.<sup>5</sup> By contrast, a *narrow-scope* theorist will think that the subject is rationally required to believe that *p*. For the narrow-scope theorist will hold that if you believe that *p* and notice that *p* entails *q*, then you are rationally required to believe that *q*.<sup>6,7</sup>

The two-factor view is a narrow-scope theory. For a core idea of the view, as applied to the problem of hijacked perception, is that there is a narrow-scope rational requirement when it comes to responding to perceptual experiences. In particular, there is a requirement of something like the following form: if you notice that you are having a perceptual experience as of *p*, and you

---

<sup>5</sup> Some representative arguments for a wide-scope theory can be found in Broome (2005), (2007), and (2008).

<sup>6</sup> Some representative arguments for a narrow-scope theory can be found in Kolodny (2005a), (2005b), (2007a), (2007b), and (2008).

<sup>7</sup> More precisely, this is just an illustration of what the rational requirement might look like.

are considering whether or not to believe that  $p$ , while having every reason to believe that your perceptual experience is reliable, then you are rationally required to believe that  $p$ .

I have made no attempt in this paper to address this broader debate between wide- and narrow-scope theories. I will simply note that insofar as a particular narrow-scope theory – namely, the two-factor view – offers an attractive solution to the problem of hijacked experience, that is a reason to favor narrow-scope theories over wide-scope ones.

## REFERENCES

- Broome, J. (2005). "Does rationality give us reasons?" *Philosophical Issues* 15 (1): 321-337.
- Broome, J. (2007). "Wide and narrow scope." *Mind* 116 (462): 359-370.
- Broome, J. (2008). "Is rationality normative?" *Disputatio* 23 (2): 161-178.
- Kolodny, N. (2005a). "Why be rational?" *Mind* 114 (455): 509-563.
- Kolodny, N. (2005b). "Does rationality give us reasons?" *Philosophical Issues* 15: 321-337.
- Kolodny, N. (2007a). "State or process requirements?" *Mind* 116 (462): 371-385.
- Kolodny, N. (2007b). "How does coherence matter?" *Proceedings of the Aristotelian Society* 107 (1): 229-263.
- Kolodny, N. (2008). "Why be disposed to be coherent?" *Ethics* 118 (3): 437-463.
- Siegel, S. (2017). *The Rationality of Perception*. Oxford: Oxford University Press.