

## Reply to Bridges

Niko Kolodny

Abstract: Bridges (2008) argues that the ‘Transparency Account’ (TA) of Kolodny 2005 has a hidden flaw. The TA does not, after all, account for the fact that (1) in our ordinary, engaged thought and talk about rationality, we believe that, when it would be irrational of one of us to refuse to *A*, he has, because of this, conclusive reason to *A*. My reply is that this was the point. For reasons given in Kolodny 2005, (1) is false. The aim of the TA is to offer an interpretation of our engaged thought and talk that is compatible with the falsity of (1) and that helps to explain why, when reflecting on our thought and talk, we are so prone to misrepresent what it involves. After making these points, I consider alternative senses in which rationality might be, or be taken by us to be, ‘normative’ and conclude that these alternatives have little bearing on the TA.

In Kolodny 2005, I denied that ‘rationality is normative’ in at least the sense that:

(0) When it would be irrational of one of us to refuse to *A*, he has, because of this, conclusive reason to *A*.

Bridges (2008) believes that I nevertheless accept that:

(1) In our ordinary, engaged thought and talk about rationality, we *believe* that, and express the belief that, (0).<sup>1</sup>

More specifically:

(1a) When one of us, *H*, believes that it would be irrational of him to refuse to *A*, *H* believes that he has, because of this, conclusive reason to *A*.

(1b) When one of us, *S*, tells another of us, *H*, that it would be irrational of him to refuse to *A*, *S* tells *H* that, because of this, *H* has conclusive reason to *A*.

(1c) When one of us, *S*, believes of another of us, *H* (who might be *S* himself at an earlier time, or an arbitrary, abstractly imagined person) that it would be irrational of *H* to refuse to *A*, *S* believes that, because of this, *H* has conclusive reason to *A*.

Indeed, Bridges claims that explaining (1) was the ‘reason for being’ of my ‘Transparency Account’ (TA) (p. [FIRST PAGE OF BRIDGES’S COMMENT]).

I find this puzzling. I do not accept (1). I took the arguments in Kolodny 2005 to tell against not only (0), but also (1). (1a) is implausible, as I argued in Sect. 3, because for *H* to take the prospect of his irrationality as a reason to have *A* would be for *H* to be bizarrely focused on the tidiness of his own mind, rather than on the world. Moreover, this ‘reason’ would be non-evidential, and so *H* could not reason from it to forming or revising a *belief*. A moral of the bootstrapping problem of Sect. 1 was that (1b) and (1c) are implausible as well. *S* may sincerely say to *H*: ‘(i) It is not the case that you have conclusive reason to believe *p*, since there is not a shred of evidence for *p* and a mountain against it. (ii) But given that you believe that you have

---

<sup>1</sup> Or perhaps Bridges thinks that I accept that we (when engaged) believe that ‘rationality is normative’ in some other sense. I discuss other senses below.

conclusive evidence that  $p$ , it would be irrational of you to refuse to.' If what  $S$  really meant by (ii) was that  $H$  had conclusive reason to believe  $p$ , then  $S$  would contradict himself. Suppose  $S$  a decent person and substitute 'commit an atrocity' for 'believe  $p$ ', and one gets the point a different way.

Accordingly, the point of the TA was precisely to suggest an interpretation of our thought and talk about rationality that was compatible with the falsity of (1) (see, for example, p. 558). Bridges's objection is that it is a hidden flaw of the TA that it does not explain (1). My reply is that this was its announced virtue. If the TA explained (1), then it would not be compatible with the falsity of (1).

Why does Bridges nonetheless believe that I accept (1)? Here is one conjecture; I will make another later. Bridges reasons: 'Kolodny says that the TA "explains the apparent normativity of rationality." What else could Kolodny mean by the "apparent normativity of rationality," if not (1): that *when we are engaged* in our ordinary thought and talk about rationality, we think and say that when it would be irrational of one of us to refuse to  $A$ , he has, because of this, conclusive reason to  $A$ ?' What I had in mind instead was: the fact that *when we reflect on* our engaged thought and talk about rationality, we are inclined to accept (0) and (1), thereby *misinterpreting* that ordinary, engaged thought and talk. (It is a truism that we are not transparent to ourselves, and it is, in any event, testified to by the striving of so much of philosophy to spell out the contents of our attitudes and utterances.) The subjects of the 'appearance' or 'illusion,' then, are ourselves *when we step back and try to come to terms with* what we think and say. In fact, the main subject I had in mind was myself for a long time after first reading Broome 1999. My driving question was why it had it seemed so natural to me to accept Broome's original proposal—that when we say that someone is rationally required to  $A$  we mean that he has, because of this, conclusive reason to  $A$ —when on closer inspection, it was clear that this cannot be what we mean.

The negative side of my project was to deny (0) and (1). The positive side, of which the TA was a part, was to explain what, if not (1), our engaged thought and talk about rationality involves, and how this might mislead us, in reflection, to accept (0) and (1).

In Kolodny 2005, I overlooked what I now think is a major source of the tendency to accept (0) and (1). The following is usually, perhaps always, true:

Pattern of Reasons: In any given case, either one has conclusive reason (and so is required by reason) to have  $A$ , or one lacks sufficient reason (and so is required by reason not) to believe that one has conclusive reason to have  $A$ .

In brief, Pattern of Reasons follows from the following two assumptions. The first assumption is that one's reason to have  $A$ , like one's reason to believe that one has conclusive reason to have it, depends on the evidence.<sup>2</sup> Hence (although more needs to be said here) it is usually, perhaps always, true that:

(\*) If one has conclusive reason to believe that one has conclusive reason to have  $A$ , then one has conclusive reason to have  $A$ .

---

<sup>2</sup> I do not think, however, that the evidence has to be that of the agent. Also 'true' in the following sentence may have to be read as 'true at a given context of assessment.' See Kolodny and MacFarlane ms.

The second assumption is:

Epistemic Strictness: There are no mere epistemic permissions. If one does not have conclusive reason to believe  $p$ , then one lacks sufficient reason to believe  $p$ .

Hence:

(\*\*) If one does not have conclusive reason to believe that one has conclusive reason to have  $A$ , then one lacks sufficient reason to believe that one has conclusive reason to have  $A$ .

Together with the fact that:

In any given case, either one has conclusive reason to believe that one has conclusive reason to have  $A$ , or one does not have conclusive reason to believe that one has conclusive reason to have  $A$ ,

(\*) and (\*\*) imply Pattern of Reasons.

To the extent that Pattern of Reasons is true, it will also be true that:

Violation Claim: If one refuses to have  $A$ , but believes that one has conclusive reason to have  $A$ , then one violates a requirement of reason.

Either one has conclusive reason to have  $A$ , which one refuses to have, or one lacks sufficient reason to believe that one has conclusive reason to have  $A$ , which one nonetheless believes. Considering the Violation Claim in isolation, one may think: 'When one believes that one has conclusive reason to have  $A$ , but refuses to have  $A$ , what else is one doing but violating  $C^+$ ? So, the requirement of reason in question must be a requirement not to violate  $C^+$ . Thus, (0) and, given that we know this, (1).' But this is just to overlook that the requirement of reason in question is one of those mentioned in Pattern of Reasons.<sup>3</sup>

Another part of the explanation, which I discussed in Sect. 4, is that philosophers often mistake the evaluative for the normative. None of the arguments against (0) and (1) touch:

(5) When  $S$  judges that  $H$  is being irrational in refusing to have  $A$ ,  $S$  negatively appraises  $H$ , as functioning poorly, or manifesting some vice.

Perhaps we just confuse (5) with (1). This seems fairly plausible with respect to the third-person beliefs in (1c), since the distinction between the normative and evaluative is obscured from the third-person perspective.

I worried, though, that this explanation was too simple with respect to the first-person beliefs in (1a) and second-person charges of irrationality in (1b). For it seems true that:

---

<sup>3</sup> In Kolodny 2007, I suggest how the appeal of other putative requirements of rationality might be explained in a similar way.

(6) When it would be irrational of  $H$  to refuse to have  $A$ ,  $H$  actually believes that he ought to, or must, have  $A$ —and not (or not only) that he would be open to some negative appraisal for refusing to have  $A$ .

(7) Some charges of irrationality look more like advice than a mere appraisal could. For example, Douglass says to Davis: ‘Given that you, Davis, believe that God created all men equal, must not you, on pain of irrationality, believe that He created this man equal?’ in the expectation that Davis, on taking this in, will feel that he, indeed, must believe this.

The TA was meant to explain (6) and (7). Here is how the TA explains (6). If it would be irrational of  $H$  to refuse to have  $A$ , this is because  $H$  satisfies the condition of  $C^+$ . And what it is for  $H$  to satisfy this condition just is for  $H$  to believe that he has conclusive reason to have  $A$ . The misinterpretation arises when we conclude from:

(6) When it would be irrational of  $H$  to refuse to have  $A$ , he believes that he has conclusive reason to have  $A$ .

to:

(1a) When it would be irrational of  $H$  to refuse to have  $A$ ,  $H$  believes that, *because of this*, he has conclusive reason to have  $A$ .

This brings me to my second conjecture as to why Bridges thinks that I accept (1). I wrote that when  $H$  satisfies the condition of  $C^+$ , it will ‘always *seem*’ to him, of the attitude that  $C^+$  requires him to have, that he has conclusive reason to have it (p. 513). This may have suggested to Bridges that I meant that it was  $H$  who is under the illusion that rationality is normative. But  $H$  is under no such illusion (or at least not until he tries, reflectively, to come to terms with his own ordinary thought and talk). What  $H$  takes to be normative is not the rational requirement  $C^+$ , but instead the content of the belief in virtue of which he satisfies the condition of  $C^+$ : a belief to the effect that he has conclusive reason to have  $A$ . What ‘always seems’ to be the case to  $H$  is only is that he has conclusive reason to have  $A$ , not that rationality is normative. (Why write ‘*seem*,’ given that  $H$  is not under the illusion? Because  $H$ ’s belief that he has conclusive reason may be false.) It is only we who reflect on  $H$  who are vulnerable to an illusion: namely, that  $H$  takes the *irrationality* of refusing to have  $A$  as conclusive reason to have  $A$ .

Here is how the TA explains (7). When Douglass observes that Davis would be irrational in refusing to believe that God made this man equal, Douglass may, although he need not, aim to draw Davis’s attention to what makes this so: that, as Davis believes, he has conclusive reason to believe that God made this man equal. If Douglass succeeds in this, then Davis will indeed feel that he must believe it. This looks more like advice than a mere evaluation could, in that it has a feature of advice that a mere evaluation lacks:

(A1) that  $S$  tells  $H$  something about a possible response of  $H$ ’s, with the aim that, if  $H$  accepts what  $S$  tells  $H$ ,  $H$  will believe he has reason to give that response.

Of course, as Bridges correctly emphasizes, it (often) lacks *another* feature of advice:

(A2) that  $S$  says that  $H$ , in fact, has reason to give that response.

As I wrote, it is unlike ‘normal advice,’ in that it (often) lacks (A2) (p. 513). This is crucial. If it (always) had (A2), then we would be saddled with (1b), which we seek to avoid. The mistake is to reason from the fact that:

Douglass’s speech is similar to normal advice in one respect, namely that (A1) is true of it: that Douglass tells Davis something about a possible response of Davis’s, with the aim that, if Davis accepts what Douglass tells Davis, Davis will believe he has reason to give that response,

to the conclusion that:

Douglass’s speech act is normal advice, and so that (A2) is true of it: that Douglass says that Davis, in fact, has reason to give that response.

This mistake misleads us into accepting:

(1b) When one of us, *S*, tells another of us, *H*, that it would be irrational of him to refuse to *A*, *S* tells *H* that, because of this, *H* has conclusive reason to *A*.

I suspect that Bridges believes that our ordinary thought and talk is more deeply committed to the ‘normativity of rationality’ than I have suggested. Indeed, he gives the impression that he believes (1). I have already said why I find (1) untenable. So let us consider whether our ordinary thought and talk might be committed to the ‘normativity of rationality’ in a way different from (1), but still more significant than I have so far suggested. If this commitment is not (1), then either it is a commitment to there being *some kind of normativity other* than that of having conclusive reason that attaches to our giving the responses that it would be irrational of us to refuse to give, or it is a commitment to there being conclusive reason *for something other* than giving the responses that it would be irrational of us to refuse to give.

On the first alternative, there seem to be two possibilities:

(8) In our ordinary, engaged thought and talk about rationality, we believe that when it would be irrational of one of us to refuse to *A*, he has, because of this, *pro tanto* reason to *A*.

(9) In our ordinary, engaged thought and talk about rationality, we believe that when it would be irrational of one of us to refuse to *A*, he *ought*, because of this, to *A* in *some sense other* than that of having reason to *A*.

The arguments against (1) apply, in modified form, to (8). At any rate, it is doubtful that whatever might explain (8) would also explain what the TA explains: namely, (6) and (7). For instance, (6) involves a belief about an ‘ought’ or ‘must’, something with the ‘strictness’ of a requirement, which a merely *pro tanto* reason lacks. So even if (8) is true, we still need the TA. In Kolodny 2005, I conceded too much to (9): to the suggestion that there is an additional normative ‘ought’ of rationality. I doubt that this suggestion is intelligible, let alone plausible. It would amount to the idea that we have two separate, independent registers of deliberation, one focused on reasons, the other on our own rationality, such that the latter obeys none of the

strictures of the former (allowing, for example, deliberation from nonevidential considerations to belief). I suspect that any attraction to (9) derives simply from confusing it with (5): that there is an evaluative ‘ought’ of rationality.

This brings us to the other alternative:

(10) In our ordinary, engaged thought and talk about rationality, we believe that each of us has conclusive reason to do *some X that does not entail giving the responses that it would be irrational of us to refuse to give*.

As it happens, I have argued that something close to (10) is correct, where ‘X’ is replaced by ‘to cultivate (or resolve to cultivate) a *disposition* to give the responses that it would be irrational of us to refuse to give—to become (or resolve to become) the sort of *person* who is not irrational’. Indeed, we in fact *have* reason—*pro tanto*, if not always conclusive—for such a disposition, or for being such a person.<sup>4</sup> Even if some version of (10) is true, it is doubtful that whatever might explain it would also explain what the TA explains. Certainly (10) itself—that we believe that we have conclusive reason to do *something other than give* the responses that it would be irrational of us to give—does not explain (6)—that we believe, of the responses that it would be irrational of us to refuse to give, that we ‘ought to’ or ‘must’ *give them*. So, in any event, we still need the TA.<sup>5</sup>

NIKO KOLODNY

314 Moses Hall #2390  
University of California, Berkeley  
Berkeley, CA 94730-2390  
USA  
[kolodny@berkeley.edu](mailto:kolodny@berkeley.edu)

#### References:

- Bridges, Jason 2008: ‘Rationality, Normativity, and Transparency’. *Mind*, [PAGES].  
Broome, John 1999: ‘Normative Requirements’. *Ratio*, 12, pp. 398–419.  
Kolodny, Niko 2005: ‘Why Be Rational?’. *Mind*, 114, pp. 509–563.  
———2007: ‘How Does Coherence Matter?’. *Proceedings of the Aristotelian Society*, 107, pp. 229–263.  
———2008: ‘Why Be Disposed to Be Coherent?’. *Ethics*, 118, [NO PAGES YET]  
Kolodny, Niko and John MacFarlane ms.: ‘Ought: Between Objective and Subjective’.

---

<sup>4</sup> See Kolodny 2008.

<sup>5</sup> I am grateful to Wai-hung Wong for very helpful discussion and to Jason Bridges for his comment, which helped me to see how unclear I had been.