

## 5 Further thoughts on the transparency of mental content

---

*Paul Boghossian*

### **Introduction**

In this chapter I want to look further at the question of whether thoughts should be held to obey a principle of transparency: roughly, whether it ought to be a constraint on an adequate theory of concepts, the constituents of thoughts, that a thinker is always able to tell introspectively whether her thoughts exercise the same or distinct concepts (this thesis will be formulated with more care in a moment).<sup>1</sup>

I am spurred to this further examination by a recent book by Mark Sainsbury and Michael Tye (hereafter S&T), *Seven Puzzles of Thought* (2012). In that book, the authors present a novel and intriguing picture of concepts according to which concepts are individuated by their *origins*.

As S&T realize, their theory is committed to denying the transparency thesis. However, unlike most theorists who are similarly committed, they don't merely note this consequence; they try to mitigate its seemingly unpalatable consequences. Their discussion provides us with a valuable opportunity for re-examining the topic of transparency and for reassessing the prospects for doing without it.

### **The transparency thesis**

S&T formulate the transparency thesis as follows:

I am grateful to Mark Sainsbury and Michael Tye for very productive and enjoyable exchanges on these topics, and to Gregory Bochner and Kit Fine for helpful comments on an earlier draft of this chapter. Some of the material was presented at a three-part lecture series at the École Normale Supérieure (Institut Jean Nicod), Paris, in May 2012, and then at an Author Meets Critics session at the Pacific APA in March 2013. I am grateful to the members of those audiences and especially to Michael Murez, Laura Schroeter, and Francois Recanati for helpful feedback.

<sup>1</sup> For my previous discussions of this issue, see Boghossian 1992a, 1994, 2011.

(IKCC) When our faculty of introspection is working normally, we can know a priori via introspection with respect to any two present, occurrent thoughts whether they exercise the same or different concepts. (Transparency)<sup>2</sup>

I will accept this formulation for present purposes. IKCC can be broken up into two distinct and independent theses:

(IKCCS) If S's faculty of introspection is working normally, then if S occurrently thinks two thoughts that exercise the *same* concepts, he will be able to know that fact introspectively. (Transparency of Sameness)

(IKCCD) If S's faculty of introspection is working normally, then if S occurrently thinks two thoughts that exercise *distinct* concepts, he will be able to know that fact introspectively. (Transparency of Difference)

What are some theories under which IKCCS is violated?

Millianism offers one well-known example. According to Millianism (at least as I shall understand it for present purposes), the meaning of (or concept expressed by) a proper name just is its reference. On such a view, it is very easy for there to be cases in which a speaker has two names which mean the same but where that fact is introspectively inaccessible to him. Kripke (1979) describes a now famous example. Peter hears the Polish pianist Ignace Paderewski and comes to believe:

(1) Paderewski is a great pianist.

He also goes to a political rally featuring the Polish prime minister Paderewski. Since he believes that no politician could also be a great musician, he believes of the prime minister

(2) Paderewski is not a great pianist.

The two occurrences of 'Paderewski' have the same reference, but clearly Peter is unable, merely introspectively, to know that. On a Millian view, Peter has two occurrent thoughts that exercise the same concept, but is unable to tell purely introspectively that they do.

As Loar (1988) has shown, IKCCS will also be violated by Fregean views of concepts, if those concepts are individuated in the way advocated by externalist views. Take, for example, the sort of social externalism defended by Burge (1979). On such a view, a thinker may be credited

<sup>2</sup> Some object to using 'a priori' for knowledge gained introspectively. This issue plays no role in the discussion to follow.

with a concept, even though he radically misunderstands that concept, provided he is disposed to defer to the norms adopted by his community. Paul, having been brought up in a sheltered way by his French nanny, learns the French word for cat, namely, 'chat', but without observing the animals themselves. He learns to assent to

(3) All chats have tails.

Because he defers in his use of 'chat' to the community of competent French speakers, by the terms of social externalism, his use of 'chat' expresses the concept *cat*. Later he visits his English cousins, who have a Manx cat, and so he learns to assent to

(4) Not all cats have tails.

Once again, externalist views of concept individuation have it that a thinker is exercising the same concept twice over but is unable to tell introspectively that he is.

As these cases illustrate, if a thinker were to violate the transparency of sameness,

(CT) He would be disposed to affirm contradictory thoughts; and he won't be able to recover from this condition on a purely introspective basis.

(RVI) He would be disposed to refuse to perform inferences that are in fact perfectly valid from a logical point of view, on the grounds that they are logically invalid; and he won't be able to recover from this condition on a purely introspective basis.

What are some theories under which IKCCD is violated?

In his 1988, Burge described the case of a subject, call him "Switched Peter," who has been stealthily switched between Earth and Twin Earth, each time staying long enough to acquire the concepts characteristic of each place *water* on Earth and *twater* on Twin Earth.

In my 1994, I argued that for a subject like Switched Peter, a thesis of *Cohabitation* will be true: earthly and twearthly concepts will commingle in Peter's psychology, without his being aware of that fact. Peter will have both the *water* concept and the *twater* concept, but he will be unaware that he has two 'water' concepts instead of one. Assuming him to be on Twin Earth, and to put it simplistically for present purposes, the *water* concept will get activated when he is recalling 'water' experiences had while on Earth, whereas the *twater* concept will get activated when he is thinking about his current environment. In such a scenario, Switched Peter would clearly violate the Transparency of Difference.

Cohabitation is by no means uncontroversial. In his interesting 1998, Michael Tye denied that switching would lead to cohabitation of twinned concepts. By contrast, Burge (1998) insists that Cohabitation is true. Both Burge and I have presented arguments in favor of Cohabitation, on externalist assumptions. For present purposes, I will simply assume that Cohabitation is correct.<sup>3</sup>

As these cases illustrate, if a thinker were to violate the transparency of difference:

(WCT) She would be disposed to think that certain pairs of thoughts are contradictories when they are not; and she would be unable to recover from this condition on a purely introspective basis.

(PII) She would be disposed to perform logically invalid inferences, thinking that they are perfectly valid; and she would be unable to recover from this condition on a purely introspective basis.

For example, Switched Peter might perform an inference that he would express in public language like this:

(Pavarotti)

- (1) Whoever floats on water, gets wet
- (2) Pavarotti once floated on water
- (3) Therefore, Pavarotti once got wet

According to the cohabitationist, though, the actual contents of his inference may well be this:

(*Pavarotti*)

- (1) *Whoever floats on twater, gets wet*
- (2) *Pavarotti once floated on water*
- (3) *Therefore, Pavarotti once got wet*

However, Peter will not be able to tell introspectively that he is committing a fallacy of equivocation.

Externalist theories of concepts, then, are highly likely to violate both the Transparency of Sameness and the Transparency of Difference. So what? Why is it in any way problematic for a view of concepts that it allows for such violations?

<sup>3</sup> It would be a mistake to think that cohabitation of twinned concepts would arise only as a result of extreme science fiction scenarios involving stealthy removal to far-off planets. If a Burge-style social externalism is true, according to which we possess the concepts that are expressed by the experts in our community, regardless of whether the conceptual roles that we associate with our concepts matches theirs, then cohabitation would arise all the time, as we drift back and forth between different communities subscribing to different linguistic norms, without our realizing that we have done so.

### **The paradox**

The problem can be expressed in the form of a paradox: we are simultaneously attracted to three propositions that form an inconsistent triad.

*(Logical Rationality, LR)* If S is rational, S is able, purely introspectively, to avoid simple contradictions and simple fallacious inferences.

*(Illogical Peter, IP)* Peter is unable to avoid, purely introspectively, simple contradictions or fallacious inferences in his thinking.

*(Rational Peter, RP)* Peter is a fully rational agent.

Externalist theories thus face a challenge. They appear to allow cases in which the conjunction of (IP) and (RP) are true. But these are not consistent with (LR). Something has to give, the only question being what. There is no question of living with all three propositions.

I will focus mostly on the case of Switched Peter, referring to (Unswitched) Peter as needed.

### **Rejecting illogical Peter or rational Peter**

As I mentioned above, Tye (1998) argued against a cohabitationist construal of Switched Peter. He maintained that earthly and tweathly concepts won't cohabit in Peter's psychology. With each slow switch, one set of concepts will completely displace the other.

Burge (1998) also argued against (Illogical Peter). Burge did it not by denying Cohabitation but by arguing that the phenomenon of preservative memory, as it operates in reasoning, will ensure that Peter's inferences are free of undetectable equivocations, despite Cohabitation's being true.

Sorensen (1998) and Faria (2009) have suggested that it is (Rational Peter) that has to go. Being rational, they say, just like being moral, can be a matter of luck. And Peter is unlucky.

### **Rejecting logical rationality – I**

The main line defended by S&T, however, is neither of these, but consists rather in arguing that the thesis of (Logical Rationality) is false: avoidance of simple contradictions, and of simple fallacies of equivocation, is not required for rationality. A thinker can be perfectly rational even while holding occurrent contradictory thoughts or while being inclined to engage in occurrent invalid inferences, provided certain conditions are met.

Before looking at S&T's alternative to (LR), let me briefly describe why (LR) has traditionally seemed attractive. Let us start with the question: What are we after when we hanker after a theory of rationality?

One guiding thought has been that we want to distinguish sharply between two different sorts of cognitive failure that someone may suffer from: between merely *lacking* empirical information about the world, versus *mishandling* the information that one has.

We reserve the word 'irrational' for the second failing; we do not apply it to the first. We want to say that the recently and surreptitiously envatted brain, who is suddenly confronted with a steep loss of accurate information about his actual surroundings, can remain perfectly rational in continuing to believe what he believed previously. Indeed, it would be *irrational* for this unfortunate person to suddenly come to believe that he is envatted, even if so believing meant that he would now have a much higher proportion of true beliefs. The surreptitiously envatted person who continues believing that he is in a normal environment is handling the information at his disposal in a rational way, correctly from the standpoint of the norms governing the handling of evidence, even though this way of handling the information results, in the circumstances, in a high number of false beliefs.

All of this is intuitive. However, now we may want to know: What, exactly, is it to handle information correctly?

Clearly, we may handle the information at our disposal correctly and yet believe something false. Indeed, it looks as though we could be handling the information at our disposal correctly and yet believe two propositions that are inconsistent with one another, that could not be true together in any possible world. But, then, what is it to handle information correctly? What is the substantive content of that idea?

An important part of a traditionally influential answer is that while a rational person cannot be counted upon to avoid falsehoods or even inconsistencies, she can be counted upon to avoid (occurrent, henceforth I will omit this qualification) thoughts that *logically contradict one another*, and inferences that are *logically invalid*.

For this to make any sense, though, there had better be a sharp distinction between a mere inconsistency and a contradiction. What's the difference supposed to be?<sup>4</sup>

A common description of the distinction is this: Two thoughts are *inconsistent* with one another iff it is metaphysically impossible for

<sup>4</sup> Similar remarks apply to the distinction between arguments in which the premises guarantee the conclusion's truth and those that are logically *valid*.

them to be true together. By contrast, two thoughts *contradict* one another if and only if they are inconsistent with one another *and* one consists of *the very same thought* as the other embedded in a concept for negation.

It would be intuitive to illustrate these distinctions with the following examples: (5) and (6) express thoughts that are merely inconsistent with one another:

(5) Water boils at 100 degrees.

(6) H<sub>2</sub>O does not boil at 100 degrees.

However, the thoughts expressed by (5) and (7) *contradict* one another:

(7) Water does not boil at 100 degrees.

It's consistent with a thinker's being rational that she believe (5) and (6) together, but not that she believe (5) and (7) together. But this clarification pushes our question further back. Why should the difference between an inconsistency and a contradiction, as so outlined, *matter* from the standpoint of the theory of rationality?

S&T provide the beginnings of an answer when they say (p. 134): "Being contradictory is supposed to be a specially manifest way of being inconsistent." But this understates the point, which is that, if the traditional way of fleshing out the theory of the rational handling of information is to make sense, contradictions have to be not just especially manifest, but *introspectible*, knowable to obtain by mere reflection, without the benefit of further information about the external world. Otherwise, we will need empirical information to discover that our thoughts contradict one another, which would jeopardize the distinction we are after.<sup>5</sup>

So, if we give the traditional account of what distinguishes a mere inconsistency from a logical contradiction, it had better be that the relation of being the 'same thought as' is introspectively accessible.

This was, of course, guaranteed by Frege's theory: Thoughts are individuated not by their references, but by their *senses*, where two senses a and b are the same iff there are no circumstances under which a rational person would accept both

<sup>5</sup> Of course, some contradictions may involve such syntactically complicated pairs of propositions that the fact that they contradict each other may not be immediately manifest to an ordinary mind, equipped with the usual capacities for memory and computation. But the idea is that they would be introspectively manifest in principle, given enough time, memory, and computational power. In any case, we can restrict our attention here to syntactically simple cases.

Fa

and

Fb.

Frege's theory is designed, in other words, to allow us to take a certain special kind of inconsistency – contradiction – as *constitutive* of someone's irrationality. If we know of a rational person that she is treating two of her thoughts as consistent with one another, then we know to individuate her thoughts in such a way that they don't contradict one another. Conversely, if we know how to individuate her thoughts, we know how to assess her rationality (at least in this basic respect).<sup>6</sup>

However, if we operated with a conception of sameness of thought that is not transparent in this Fregean way, we would have to give up on the idea that conforming to the basic principles of logic is partially constitutive of rationality. For it would then be possible to describe completely rational thinkers who nevertheless fail to conform to the basic principles of logic in even the simplest of cases.

### Rejecting logical rationality – II

Well, if conformity in simple cases to the basic principles of logic is not what rationality (partly) amounts to, what does it amount to?

S&T have an alternative picture in mind. Here, for example, is what they say about the rationality of Peter's (Pavarotti) inference:

... Paul is in no position to recognize the invalidity of the inference he has drawn, but ... this does not make him irrational. Paul, we may suppose, is like the person who upon hearing the following argument takes it to be valid.

All reindeer have four legs  
 Rudolph is a reindeer  
 So, Rudolph has four legs

In reality, the argument is invalid since "Rudolph" in the second premise is used to refer to a reindeer, whereas "Rudolph" in the conclusion is used to refer to a human being. One who hears the argument and assumes that it is valid is not behaving irrationally. Indeed, if the false assumption that the name "Rudolph" is unambiguous is rationally based, rationality *requires* the hearer to draw the conclusion that Rudolph has four legs.

<sup>6</sup> For a discussion of the relation between Frege's cognitive significance test and the Transparency thesis, see Goldberg 2008c.

Paul is a bit like the person deliberating about what to do. An agent, given his knowledge of the situation he finds himself in, may come to a conclusion about the right course of action – a conclusion that is fully warranted by the information available to him – even though in the circumstances it was the wrong thing to do. Evaluating this person from the moral point of view, we may find no fault with his internal reflections, notwithstanding the fact that there was a bad upshot.

Correspondingly, evaluating Paul as a rational subject, we may find no reason for criticism, even though in the circumstances he holds bad beliefs and produces bad arguments.

S&T take a similar view of the ‘Paderewski’ case:

Rational subjects are not always in a position to know whether they are exercising a fresh concept, or the same concept again. Peter thinks at the rally that he has learned a new concept PADEREWSKI different from the PADEREWSKI concept he had earlier acquired at the concert. He is wrong. But if this is a *reasonable* mistake, it is one that makes it reasonable for him to believe that his thoughts are not contradictory. (2012, p. 134)

How should we understand S&T’s view? Their idea seems to be that the (LR) account of the rationality of first-order beliefs that we have been working with is too simplistic. Peter is not necessarily guilty of irrationality just because he is disposed to perform a simple fallacious inference or to embrace a contradiction. Rather, the correct account is disjunctive:

(STR) Peter is rational either if he is disposed to avoid fallacious first-order inferences or contradictions, or by having a rational second-order belief to the effect that his first-order beliefs do not harbor contradictions and his inferences are not fallacious.

S&T want to allow that a thinker can have rational first-order beliefs and inferences without the benefit of any second-order, meta-attitudinal beliefs. So, they don’t want to rest the rationality of first-order attitudes and movements of thought merely on the rationality of second-order beliefs. As a result, their account has the disjunctive character that is captured in (STR).

The question is whether we can make sense of (STR).

### **Rejecting logical rationality with first-order resources**

An experienced eye would hesitate both over the disjunctive and the second-order characters of the account. Trouble looms. Why did S&T go in this direction? Why didn’t they look for an account that tried to solve the problem using only first-order resources? In the Switched Peter case,

for example, why not say that Peter is rational in performing the (Pavarotti) inference because he is tacitly assuming that

(Identity) Twater is water.

To see why, let me begin with the observation that even if we were to grant that Peter is capable of judging (Identity), he would only be able to do so by accident. He wouldn't be able to judge (Identity) *intentionally*.

To be clear, what's at issue for the moment is not Peter's ability to intentionally think second-order thoughts of the type:

(SO-Identity) The concept expressed by the first token of 'water' in (Pavarotti) is the same as the concept expressed by the second token of 'water' in (Pavarotti).

What's at issue is his ability to judge (Identity) with the concepts *water* and *twater* occurring in the thought directly, and not merely under some description that picks them out.<sup>7</sup> Once that's made clear, it requires little argument to show that Peter would not be able to judge (Identity) intentionally. He doesn't know that he has two 'water' concepts; he thinks he has only one such concept.<sup>8</sup>

Now, if this is right, then the limitation in question is extremely interesting. For, by hypothesis, Peter has both the concept *water* and the concept *twater*. By hypothesis, both concepts figure in at least some of his first-order thoughts.

However, on our usual way of thinking about such matters, when we credit someone with possession of two concepts, A and B, we assume that he would be able to intentionally deploy those concepts in all the thought types that he is competent with. For example, since I have both the concept *coriander* and the concept *diamond*, I am able to ask myself:

Is coriander diamond?

however silly the question may be. But this is not true of Peter's ability to deploy the concepts *water* and *twater*.

Does this show that we were wrong to think that full deployment follows automatically from possession? Or does it show that transparency is an important part of the way in which we think about concept possession in the first place, so that there is a question about whether we were right to assume that Peter possesses both concepts? The question demands much more investigation.

<sup>7</sup> Nor is it at issue that he could have a thought of the form "This is that" where the first demonstrative refers to water and the second refers to twater.

<sup>8</sup> I am inclined to think that he Peter may not be able to judge (Identity) at all, whether intentionally or not, but won't argue for the stronger claim here.

Not only is Peter unable to think (Identity) intentionally, even if he were to think it by accident he wouldn't *realize* that he had thought it. He would believe that he had thought (Identity-Trivial).

(Identity-Trivial) Water is water.

As a result, it's clear that his thinking (Identity) would not help vindicate his rationality in performing the Pavarotti inference.

Second, given the *way* in which the two occurrences of 'water' will present themselves to Peter, it will seem absurd to him to think anything like the (Identity) thought. As Schroeter (2007) says of a related case:

You think "Bush smirks" and "Bush swaggers." You then draw the conclusion that "Someone both smirks and swaggers." ... the sameness of the man in question won't strike you as up for dispute: your thoughts simply present themselves to you as co-referential. (598)

In other words, the two occurrences of 'water' in the Pavarotti argument will seem *indisputably* co-referential to Peter. Given that appearance, it will seem absurd to Peter to guard against their possible failure to co-refer by interpolating anything like the (Identity) thought.

Finally, even if we could make sense of all this in the Switched Peter case, it's hard to see how it could be of any use in the Unswitched Peter's 'Paderewski' case.

For all these reasons, then, it looks as though S&T were right to avoid trying to solve our problem using only first-order resources.

### **Rejecting logical rationality with second-order resources**

Let's turn then to a solution, like S&T's, that invokes second-order resources. Peter is to be counted as rational provided he has a rational second-order belief to the effect that his first-order beliefs and inferences are all right.

About any such solution, we need to ask three questions:

- (A) Can we count on there being the appropriate second-order belief?
- (B) What would it be for this second-order belief to be rational?
- (C) Can the putative rationality of such a second-order belief help explain the rationality of Peter's first-order beliefs?

What is the second-order belief to be? How is Peter going to think a substantive identity belief linking his *water* and *twater* concepts when, by hypothesis, he doesn't know that he has two concepts and believes, indeed, that he has only one?

What Peter needs to do is to be able to refer to the occurrences of the concepts without exploiting their particular identities. The most plausible way for him to do that would be *metalinguistically*:

(Water-SO) The concept in the first premise that I would express with the word 'water' is the same as the concept in the second premise that I would also express with the word 'water'.

(Paderewski-SO) The concept expressed by the first occurrence of 'Paderewski' is distinct from the concept expressed by the second occurrence of 'Paderewski'.

But this implies that any second-order solution to our problem is constitutively limited to creatures that are capable of expressing their thoughts through a public language. (Merely assuming that the creatures have a language of thought won't do, since we don't have access to the words of a language of thought in the way that we do to the words of a public language.)

Furthermore, such meta-attitudinal, metaconceptual beliefs are very sophisticated: few thinkers will have them as a matter of course. Having such second-order thoughts requires having the concept CONCEPT, and some view about when concepts are identical and when they are distinct. Yet many thinkers for whom such thoughts will be out of reach may find themselves in Peter's situation. What account will we be able to give of their rationality on S&T's view?

Let's turn next to question (B). What would make beliefs like (Water-SO) and (Paderewski-SO) rational?

Peter's 'Paderewski' belief will clearly have been arrived at by inference. Since we have given up on (LR), we cannot say that his belief (Paderewski-SO) will be rational only if it has been arrived at on the basis of inferences that are not fallacious. Rather, it looks as though we are going to have to offer the following disjunctive condition:

(Paderewski-SO) is rational *either* if it was arrived at on the basis of inferences that conform to logic, *or* Peter has a rational *third-order* belief to the effect that his inferences so conform.

And so on up the hierarchy.

It's not clear that we understand this notion of rational belief. It seems unhappily impredicative.

This seems like a strong reason for thinking that any satisfactory account of a first-order belief's rationality must consist in that belief's conforming to some norm all by itself, and cannot be handed off to the rationality of some higher-order belief.

Finally, in being *thoughts*, such second-order beliefs will themselves involve the exercise of concepts. That means that questions about the sameness and difference of the concepts involved in those second-order beliefs will also arise. How are those questions to be settled?

On pain of vicious regress, they cannot be settled via yet higher-order thoughts. But if they are not settled, then the ability of such second-order thoughts to secure the rationality of Peter's first-order thoughts is highly questionable.

For all these reasons, then, I don't see that S&T's alternative to (LR), (STR), as presented in the body of their book, has much prospect of succeeding.

### **S&T's further thoughts: rational "treatings"**

S&T have responded to these arguments in a postscript to the paperback edition of their book. In that postscript, they concede that an adequate solution to our problem cannot be secured either through a first-order or through a second-order belief strategy.

They maintain, however, that the denial of transparency can be rescued through what is essentially a modification of the second-order belief strategy. They say:

There are two crucial points: one is that a subject may be rational in *not* believing something; the other is that a subject may rationally *treat* concept-uses as uses of the same concept, or uses of a different concept, without having meta-conceptual beliefs. Not believing something does not require the exercise of concepts; and treating uses as uses of the same concept does not require believing that the uses are uses of the same concept (such a belief would require possession of the concept CONCEPT). (2012, p. 184)

In this section I will consider and assess these further thoughts of S&T. I should say that this discussion will of necessity have to be very preliminary since S&T's presentation of their new line of thinking is understandably cursory and the issues at this point are very complex.

I should also point out that those familiar with some of my work on the nature of inference Boghossian (2012) will know that I am sympathetic to the idea that person-level inferences involve something like the subject's "taking it" that the premises support the conclusion. In the case of a deductive argument, this will involve "taking it" that the relevant concept uses are uses of the same concept.<sup>9</sup> This is not very far from

<sup>9</sup> I discussed this point in the lectures I gave at the École Normale Supérieure, in Paris, in May 2012.

S&T's talk of "treating" concept uses as the same. So, I am somewhat sympathetic to the idea that they briefly outline. However, I myself don't see my way clear through the various difficulties that this proposal faces. In this section, I will emphasize those difficulties in the hope that this will help us get clear on the way forward.

Let me start with the question: What does this "treating as the same" or "treating as different" consist in, if it is not to consist in a belief? S&T are quite clear that it cannot be belief:

Boghossian is right that it cannot consist in some further belief which attempts to express sameness or difference of concepts. A belief is a structure of concepts and we still need to know what it is for a subject of the belief to treat the concept-occurrences in that structure as occurrences of the same or different ones. (2012, p. 185)

Of course, it's not just beliefs that would cause a problem. Any construal of "treating" that took it to be a *propositional* attitude would raise similar issues.

So the question is: What sort of *nonpropositional* thing could "treating as the same" or "treating as different" be?

S&T say:

We think that "treating as the same" is revealed in such cognitive behavior as forming the false belief about the validity of the Pavarotti argument. If this were all that could be said, its explanatory value would be small. But a wide variety of behavior can manifest "treating as the same". For example, we and Peter would not only rationally count as valid the following version of the Pavarotti argument, but would also be rational if we judged the rewording to make no substantive difference:

Whoever floats on water gets wet  
Pavarotti once floated on it  
So Pavarotti got wet

The reworded argument is valid (though, as stated by Peter, unsound, assuming Pavarotti never encountered twater). It is as rational for Peter as for anyone else to believe that it is valid. Like us, he treats it as equivalent to the original argument, and has no beliefs that would make it rational for him to make any distinction between the validity or soundness of the one and the validity or soundness of the other. It is as rational for Peter as for us to believe that both arguments are valid. (2012, p. 185)

This is somewhat puzzlingly formulated since the idea was to supply a *nonpropositional* notion of "treating as the same" but the wording suggests that "treating as the same" is to be identified with various dispositions to form such-and-so *belief* or such-and-such *judgment*.

S&T probably meant to say that “treating as the same” should be identified not with the disposition to form the belief that the Pavarotti argument is valid, but with the disposition to behave *as if* one believed that the argument is valid.<sup>10</sup> Let’s grant that this way of understanding the proposal manages to avoid being propositional. Can it help us with our problem?

Our problem, recall, is that we want to vindicate the claim that Peter is rational in performing the Pavarotti inference, despite that inference’s suffering from equivocation.

We assume that not *everyone* who is disposed to behave as if his inference is valid, is *justified* in so behaving. Some subjects may be sloppy and careless and simply do not pay enough attention to whether they are being inconsistent. Jon Stewart, on his *Daily Show* news satire program, had an amusing example of this. A right-wing pundit who wanted to decry what he saw as America’s “culture of dependence,” and to praise personal self-reliance, was shown confidently saying: “I used to be on welfare and food stamps and no one helped me.” A little reflection might have revealed to him that he was contradicting himself. I take it that pundit was not justified in behaving as though he wasn’t contradicting himself.<sup>11</sup>

So, when is a subject *justified* in behaving as though his thoughts are not contradictory, or his reasoning valid, and when not?

S&T seem to think that their first observation – that we can be rational in *not* believing certain things – will help with this question. They say:

Just as it is rational for you not to believe that the book you are holding in your hands will explode in ten seconds time, it is rational for us not to believe we are subject to switching. Suppose the very idea of switching has never entered our heads . . . Without addition to our stock of concepts, we cannot form beliefs about switching, and so our rationality cannot depend on our having rational beliefs about it . . . The argument strikes us as valid and we rationally lack defeaters for that opinion. Hence we are rational in believing it to be valid. (2012, p. 184)

There are two possible ways of reading the view here.

On the first, more minimal reading, if a subject is disposed to make an inference, then she is *ipso facto* justified in making that inference as a matter of default, provided there are no defeaters.

But the case of the right-wing pundit shows that principle to be false.

A different and richer view is suggested in the final sentence of the passage quoted above: if a subject is disposed to make an inference and,

<sup>10</sup> Mark Sainsbury and Michael Tye, in personal correspondence.

<sup>11</sup> It’s true that this case involves a bit of conceptual unpacking – namely, that being on welfare is being helped (by the state) – but not in ways that affect the point.

after reflection on the inference, it continues to strike her as valid, and she lacks a defeater for the opinion that striking generates, then she is justified in making the inference.

Here there are two problems. First, the right-wing pundit continues to be a potential counterexample. Second, this way of putting it clearly relies on the subject's tokening *conceptual structures* the structures involved in the argument's *striking* her as valid, and her subsequently justified *opinion* that it is valid and, as we have seen, that is not on the cards.

One is tempted to say: Well, the contradiction in the pundit's case is *manifest* and so is something that he can be held responsible for. But the contradiction (or equivocation, or what have you) in the case of Peter is not manifest, being the result of switching, and so is not something that he can be held responsible for.

But the very question before us is whether we can make sense of the sameness and difference of concepts not being introspectively manifest. We can't simply *assume* that it does make sense.

S&T's further idea, which follows in the footsteps of Campbell (1987), is that we can make sense of a subject's rationally treating two concept uses of his as the same without this involving the tokening of some conceptual structure on his part. But it is not yet clear whether we have been able to make sense of this.