

# Undefeated dualism

Tomas Bogardus

Published online: 24 May 2012  
© Springer Science+Business Media B.V. 2012

**Abstract** In the standard thought experiments, dualism strikes many philosophers as true, including many non-dualists. This ‘striking’ generates prima facie justification: in the absence of defeaters, we ought to believe that things are as they seem to be, i.e. we ought to be dualists. In this paper, I examine several proposed undercutting defeaters for our dualist intuitions. I argue that each proposal fails, since each rests on a false assumption, or requires empirical evidence that it lacks, or overgenerates defeaters. By the end, our prima facie justification for dualism remains undefeated. I close with one objection concerning the dialectical role of rebutting defeaters, and I argue that the prospects for a successful rebutting defeater for our dualist intuitions are dim. Since dualism emerges undefeated, we ought to believe it.

**Keywords** Dualism · Intuition · Phenomenal conservatism · Defeaters · Mind–body problem · Phenomenal concepts · Explanatory gap · Physicalism · Materialism · Reduction

## 1 Dualist intuitions

Many philosophers believe that for any mental state type *M*, there exists some type *T*, such that *T* is a naturalistically-acceptable reductive state type, and *M* is identical with *T*. Let’s call those philosophers *non-dualists*. It’s difficult to say precisely what would count as a “naturalistically-acceptable reductive state type,” in no small part because it’s difficult to say just what naturalism is and what it would accept. But

---

T. Bogardus (✉)  
University of St. Thomas, 2115 Summit Avenue, St. Paul, MN 55105, USA  
e-mail: tbogardus@gmail.com

paradigmatic examples will suffice to secure the relevant concept in our minds: non-dualists are apt to identify any mental state type—for example pain—with a physical brain state type,<sup>1</sup> or some physical-functional state type,<sup>2</sup> or some purely formal functional state type,<sup>3</sup> or some physical representational state type.<sup>4</sup>

Other philosophers believe that there is at least one mental state type that is not identical with any naturalistically-acceptable reductive state type. These are the *dualists*. Many of these philosophers believe dualism because, they say, in the light of well-known thought experiments, propositions that clearly entail dualism seem obviously true.<sup>5</sup>

But don't just take their word for it. Try it for yourself. Suppose that, while you experience a real fireball of a pain, you also have an autocerebroscope ready to hand. It's arranged so that you can observe the states of your own brain in perfect detail. Let the type-demonstrative concept THIS—deployed introspectively—pick out that hurtful feature of your painful experience. Let THAT refer to whichever of the candidate reductive state types you care to demonstrate. Now summon the following proposition before your mind:

**Possibility:** This and that are not necessarily coextensional.

Many philosophers—even many non-dualists—have reported the seeming truth of Possibility with respect to at least some reductive type-identity claims. For example, Saul Kripke (1972, p. 154) says “the correspondence between a brain state and a mental state seems to have a certain obvious element of contingency.” And many non-dualists agree: David Papineau (2002, p. 85) reports that “...it certainly *seems* possible that [conscious properties and the material properties they are identical with] should come apart,” and admits that zombies and ghosts seem possible (ibid., p. 87). Christopher Hill (1997, p. 65) accepts the “apparent separability of pain and C-fiber stimulation.” Speaking of Kripke's intuition, Colin McGinn (2003, p. 153) says that our correct theory “must deactivate the intuitions of contingency that surround our thinking about the relation between the mind and the body.” And Thomas Polger (2004, p. 42) says “It certainly seems that my pain now could have been other than, say, activation of C-fiber #237 now... Mind-brain identity claims have the appearance of contingency.”

<sup>1</sup> Empirical candidates for such a physical type are *displaying activity in the pyramidal cells of layer 5 of the cortex involving reverberatory circuits* (cf. Block and Stalnaker 1999), or *being cortio-thalamic oscillation or being C-fibers firing*.

<sup>2</sup> I have in mind here a chauvinistic functionalism. An empirical candidate for such a physical-functional state type would embed neurophysiological information into our Ramsey sentence. (See e.g. Block 2006).

<sup>3</sup> This is what Block calls the “deflationary” view: phenomenal properties are identical to some purely formal functional type, i.e. our Ramsey sentence eschews neurophysiological information and uses only logical (e.g. “and,” “if,” etc.), causal, or statistical terms (e.g. “typically”).

<sup>4</sup> For example, a PANIC state realized in the brain that represents tissue damage as bad (Tye 1995, 2000, 2006). Or perhaps a representational state realized in the brain that represents a cluster of properties nonconceptually, which properties are suitably poised to bring about cognitive responses (Tye 2007). Presumably, properties in this cluster are not irreducibly non-physical.

<sup>5</sup> Similarly, most of us believe that knowledge is not justified true belief because it sure *seems* like Gettier's Smith, who has a justified true belief, doesn't know. That's the sense of “seem” relevant to this dualist claim.

Possibility has been widely discussed in the literature. Whichever instance of it is of interest to you—that is, whichever reductive state type you demonstrate with THAT—the resulting proposition entails but is not entailed by a much less-discussed non-modal proposition<sup>6</sup>:

**Non-Identity:** This is not identical with that.

Many philosophers—and again even many non-dualists—have reported the apparent truth of Non-Identity with respect to at least some of the reductive type-identity claims above.

- David Papineau (2002, p. 3): “We find it almost impossible to free ourselves from the dualist thought that conscious feelings must be something *additional* to any material goings-on... the compelling intuition that the mind is ontologically distinct from the material world... we feel it is *obvious* that conscious states are not material states.”
- Daniel Dennett (1992, p. 27): “It does seem as if the happenings that *are* my conscious thoughts and experiences cannot be brain happenings, but must be *something else*, something caused or produced by brain happenings, no doubt, but something in addition...”
- Christopher Hill (2005, p. 153): “[W]hen I am attending introspectively to a pain, I am aware of something that appears to resist characterization in terms of neuroscientific concepts. To apply neuroscientific concepts to it would be like applying them to a patch of blue sky.”
- Peter van Inwagen (2009): “[The Hard Problem is] the question: ‘How could this collection of molecules actually have this kind of awareness that is my feeling of pain or orange?’ And indeed I don’t see how it could. In fact, it looks to me as if it couldn’t, except for the fact that it does.”<sup>7</sup>

Somewhat surprisingly, Non-Identity seems true even to hardcore eliminative materialists. According to Stephen Stich (1991, 1996), the motivating idea of eliminative materialism is that some theoretical terms fail to refer due to a *high degree of mismatch* between reality and the supposed nature of this theoretical posit. So, for example, Richard Rorty (1965) and many after him suggest that the folk theoretical expression “demon possession” fails to refer since the reality of the situation—manifestations of hallucinatory psychosis or epilepsy—is very different from the supposed nature of demon possession. Due to this high degree of mismatch (and our preference for the neuroscientific theory), Rorty and others say that “demon possession” fails to refer: really, there isn’t any demon possession.

<sup>6</sup> Non-Identity does not entail Possibility so long as we do not accept a modal-separability criterion as necessary for property distinctness, as we should not. After all, there are examples of properties—ways things could be—that are distinct even though they could not fail to be coinstantiated. Triangularity and trilaterality, for example.

<sup>7</sup> Of course, here Van Inwagen just reports an intuition that this collection of molecules couldn’t have phenomenal states. But since presumably he also thinks that this collection of molecules could token the relevant naturalistically-acceptable reductive types mentioned above, I take it that Non-Identity also seems true to him.

Analogously, according to Stich, the idea behind Rorty's "disappearance theory" and later eliminativist views seems to be that very many, if not all, folk-psychological terms (such as "belief") fail to refer due to a high degree of mismatch between the supposed nature of these theoretical entities and reality. That is, Non-Identity seems true to these theorists, and this seeming does not relent even under the weight of neuroscience. But given their commitment to materialism, they opt to deny that folk-psychological terms refer. For example, Paul Feyerabend (1963a, p. 295) says that the "usual" or "ancient" sense of the term *mental* is essentially non-materialistic, and (1963b, p. 54) that, on the basis of introspection, it appears that thoughts (if they exist) are very different from material processes. But he embraces reductive materialism. He therefore advocates saying there are no mental processes, and that there are no thoughts. And so it is in large part the seeming truth of Non-Identity that drives him—and other eliminativists—to this conclusion.

So, Possibility and Non-Identity seem true to wide variety of philosophers, including many non-dualists.<sup>8</sup> Of course, in that respect these propositions are not unique: countless claims concerning mental states seem obviously true in this way. For example, it seems obvious that pain is not pleasure, and that a scenario could include pain but no *pain aux raisins*. Yet such propositions do not straightforwardly entail dualism. By contrast, Non-Identity clearly entails dualism,<sup>9</sup> and Possibility clearly entails Non-Identity.<sup>10</sup> And, to repeat, these propositions seem true even to many non-dualists. They also seem true to me. Perhaps they seem true to you as well. So why not believe that things are as they seem? Why not adopt dualism?

## 2 Motivating dualism

The way in which philosophers standardly go about justifying their beliefs (cf. Bealer 1992, 1996) includes counting all the following as *prima facie* evidence: experiences, observations, testimony, and—importantly—so-called *intellectual* seemings such as those elicited by dualist thought experiments. Examples of this procedure abound: Gettier's refutation of  $K = JTB$ , Chisholm's perceptual relativity refutation of phenomenalism, Putnam's Spartan-prettender refutation of behaviorism, all the various twin-earth examples, Burge's arthritis example,

<sup>8</sup> Eliminativists count as non-dualists on my definition, insofar as they claim, with Feyerabend, that there are no mental state types, and so it's false that there is some mental state type that fails to be identical with some reductive type. That is to say that dualism, as I've defined it, is false.

<sup>9</sup> More precisely, the proposition that *this* is not *that*, or *that*, or *that*... (where each deployment of THAT refers to a distinct member of the domain of naturalistically-acceptable reductive state types and every member of that domain is referred to by an instance of THAT) both seems obviously true and clearly entails dualism.

<sup>10</sup> Non-Identity by itself entails that non-dualism—as I have defined it—is false, but is consistent with a Nagelian-style primitivism, according to which mental phenomena supervene on physical phenomena in virtue of some metaphysically necessary relation which is not identity. As I have defined dualism, Nagelian primitivism is a form of dualism. Possibility is stronger, entailing that there is a metaphysically possible world in which *this* and *that* come apart, and therefore it entails that Nagelian primitivism is false.

multiple realizability, etc. These examples all involve the evidential use of intellectual seemings, which some philosophers call “intuitions.”

So widespread philosophical practice and methodology supports the view that intuitions confer at least *prima facie*, defeasible justification. I will take it that this widespread practice is correct.<sup>11</sup> And so, the striking truth of Possibility and Non-Identity generates defeasible justification for dualism. How are intuitions defeasible? Pollock (1974) distinguishes between two types of defeaters: rebutting and undercutting. Rebutting defeaters in this case would be any argument for the conclusion that dualism is false.

An undercutting defeater, on the other hand, is more difficult to characterize. Pollock’s classic analysis would have it that an undercutting defeater, in this case, is a reason to think that the following subjunctive conditional is false: dualism would not seem true unless it were true. This analysis is not uncontroversial, however.<sup>12</sup> Yet that controversy need not delay us here. Whatever the right analysis of undercutting defeat turns out to be, an undercutting defeater attacks the relation between one’s belief and the grounds on which one holds the belief. And a successful undercutting defeater diminishes the strength of those grounds to such a degree that rationality requires the subject to lower her confidence in her belief conditional on those grounds. That much is clear. So, in this paper, we will rely only on that necessary condition for a successful undercutting defeater:

**Undercutting Defeat:** Evidence D is a successful undercutting defeater for a subject’s belief that *p* held on evidence E only if rationality requires that the subject’s credence in *p* conditional on (D&E) be lower than the subject’s credence in *p* conditional on E alone.

Below, we will run the following rough-and-ready test on proposed undercutting defeaters: Does this *type* of consideration, in general, require that a subject lower her confidence in her belief? If the answer is “no,” the proposed undercutting defeater is not successful.

Given that intuitions provide *prima facie* justification, those of us to whom Possibility and Non-Identity seem obviously true are *prima facie* justified in believing them. *Ultima facie* justification results from searching for defeaters to an extent that satisfies our relevant epistemic obligations, and finding none. Furthermore, not only are we *permitted* to believe Possibility and Non-Identity in the absence of defeaters, but we epistemically *ought* to. For it is plausible that if someone is justified in believing that *p* and (i) it’s not the case that she is permitted to suspend belief that *p*, and (ii) it’s not the case that she is permitted to disbelieve *p*, then she *ought* to believe that *p*. It’s always nice, after all, to gain one more true belief. (The goal is to collect them all.) And if you find yourself in a situation in

<sup>11</sup> Note that this claim is weaker than Michael Huemer’s (2007, p. 30) phenomenal conservatism, which entails that *every* kind of seeming confers *prima facie* justification.

<sup>12</sup> What’s questionable, at least to me, is Pollock’s claim that if P is a *prima facie* reason for some other proposition Q, then an undercutting defeater is a reason to doubt the subjunctive conditional (P → Q). For it’s plausible that a successful defeater must target a condition necessary for knowledge, and it’s plausible that the truth of that subjunctive conditional is not required for a subject to know a proposition on the basis of P. Or so I argue in Bogardus (forthcoming).

which you are justified only in believing that  $p$ , then you *ought* to believe that  $p$ . As we've said, the seeming truth of Possibility and Non-Identity justifies us in believing them. And we'd be permitted to disbelieve or suspend judgment on Possibility and Non-Identity only if we encountered successful defeaters for them. Therefore, we epistemically ought to believe Possibility and Non-Identity unless we have or gain access to successful defeaters for them. In the absence of defeaters, we ought to be dualists.

And so progress on the mind–body problem might be made by establishing whether there are any defeaters for the seeming truth of Possibility and Non-Identity. In the following section, our project is to examine several proposed undercutting defeaters. We will ask whether each proposal entails that a subject must lower her confidence in her belief. We will find that, in each case, the answer is “no”: each proposal fails as an undercutting defeater. Then, in the final section, we will consider whether these proposals might play some other role in the dialectic, and we'll explore the prospects for a successful *rebutting* defeater.

### 3 Undercutting dualism

#### 3.1 Unreliability with respect to a certain kind of intellectual seeming?

Suppose I examine the track record and learn that, when it comes to the mathematics of infinity, many propositions that strike me as true are actually false. (For example, that the set of natural numbers has more members than the set of even numbers, or that 1 is not equivalent to the infinite decimal 0.999...) I may thereby gain a defeater for any belief I currently hold about the mathematics of infinity on the basis of an intellectual seeming. Similarly, suppose I examine the track record and learn that, when it comes to non-trivial identity statements involving natural kind terms, many propositions that strike me as true are actually false. I may thereby gain a defeater for any belief I currently hold about such identity statements on the basis of intellectual seemings.

Some philosophers believe that this is our exact situation. For example, they say, water = H<sub>2</sub>O. And yet, they say, there is an appearance of contingency here. It seems possible for water and H<sub>2</sub>O to come apart, and (so) they just seem different. Yet “they” are not different, and cannot come apart. The seeming is an illusion. Therefore, they conclude, considerations like these furnish us with an undercutting defeater for many beliefs we hold regarding non-trivial identity statements involving natural kind terms. Among these beliefs that are undercut, they insist, are those regarding Possibility and Non-Identity.

#### 3.2 Response

What some *report* as the seeming possibility of a scenario with water but no H<sub>2</sub>O is, indeed, an intellectual seeming. Yet this intellectual seeming is not misleading, it's merely misreported. Let me explain.

The *sentence*

(1) “Water is H<sub>2</sub>O”

is true. Since “water” and “H<sub>2</sub>O” are rigid designators, (1) expresses a proposition that says, of one natural kind (variously called), that it is self-identical. This proposition is necessarily true. The objector urges that the proposition expressed by (1), despite being necessarily true, seems possibly false. That is, the objector urges that the following proposition seems possibly true:

(2) Water is not H<sub>2</sub>O

But since the proposition expressed by (1) is necessarily true, (2) is necessarily false. And so the objector may conclude that when it comes to the modal status of non-trivial identity claims involving natural kind terms, intellectual seemings are unreliable guides to truth: (2) *seems* possibly true, but it’s *really* necessarily false, and so intuition is prone to illusion when it comes to non-trivial identity statements involving natural kind terms. Since my belief in Possibility and my belief in Non-Identity are based on intellectual seemings, the objector concludes that I have an undercutting defeater for each.

But I and many others think the objector suffers from proposition-confusion.<sup>13</sup> She mistakenly takes the sentence “Water is H<sub>2</sub>O” to express something like one of these propositions:

(3) The watery stuff of our acquaintance is H<sub>2</sub>O

(4) “Water” and “H<sub>2</sub>O” corefer

And *these* propositions seem possibly false. But (3) and (4) are indeed as they seem: they are (metaphysically) possibly false.

A person who thought that the sentence or utterance “Water is H<sub>c</sub>O” expresses (3) or (4) would be apt to issue the report that “it seems possible that water isn’t H<sub>2</sub>O.” Perhaps such a person thinks that natural kind terms merely abbreviate non-rigid definite descriptions, and so she thinks that the sentence “Water is H<sub>2</sub>O” expresses (3). What seems true to her isn’t that the proposition expressed by (1) is possibly false, but that proposition (3) is possibly false. Or perhaps the person is confusing word and object, and what seems true to her isn’t that the proposition expressed by (1) is possibly false, but that proposition (4) is possibly false. But in each case, then, intuition has not led this person astray. She has not erred with respect to the modal statuses of propositions. Rather, she is misunderstanding the semantics of the sentence, thinking that a *sentence* (namely “Water is H<sub>2</sub>O”) expresses a contingent proposition [namely (3) or (4)] when really it expresses a necessary proposition. In each case, propositions have the modal status that they intuitively seem to have, and so we have no defeater for intellectual seemings generally.

<sup>13</sup> See especially Kripke (1972) and Plantinga (2006, pp. 25–28). Michael Tye (1986, p. 5) says “If...a man without scientific knowledge claims to be imagining that gold has atomic number 80 (rather than its actual 79) what I think we would say he *really* imagines is that some substance with the superficial observable qualities of gold has atomic number 80 (rather than 79), and that is something quite different.”

To put it another way, Kripke did *not* teach us that certain propositions that appeared contingent were really necessary. He did not point out a *modal* illusion that we suffer from. Rather, Kripke taught us that proper names and natural kind terms function—at least in the relevant cases—as rigid designators, not as disguised non-rigid definite descriptions. He drew our attention to certain sentences or utterances that we had thought expressed contingent propositions. Upon closer examination we came to see that, in those sentences at least, the proper names and natural kind terms involved function as rigid designators, and hence that these sentences or utterances express propositions that are necessarily true if true at all. Thus we are subject to a *semantic* mistake, not a *modal* mistake, an illusion as to what sentences express, not as to the modal status of propositions. But then this is not a case in which intellectual seemings have led us astray with respect to identity statements, and therefore this cannot be marshaled in support of non-dualists who claim that in the case of dualism we are suffering from a modal illusion. And having secured the proper semantics, the threat of this illusion disappears. In conclusion, then, the *prima facie* justification of Possibility and Non-Identity remains undefeated.

### 3.3 Fallacious operator shifts, perhaps?

René Descartes thought he saw, “clearly and distinctly,” that his essence did not include spatial extension, and so that he—unlike his body—could exist while no material objects existed. Hence, dualism. Antoine Arnauld gently suggested that perhaps Descartes merely failed to see the *impossibility* of his existing while no material objects exist, and mistook that for successfully seeing a possibility. By way of example, Arnauld pointed out (cf. Descartes 1984, pp. 141–142) that someone may see for certain that the angle in a semi-circle is a right angle, and yet may doubt, or not be certain, or not understand that the square on the hypotenuse is equal to the squares on the other two sides. And yet it hardly follows that having the square of the hypotenuse equal to the squares of the other two sides is not essential to this right triangle. I take the suggestion from Arnauld to be that Descartes may have been the victim of a fallacious operator shift, moving hastily from the truth that *p does not seem impossible* to the conclusion that *p seems possible*, or at least mistaking the former for the latter.

Michael Tye (1986) offers a similar suggestion:

Where I suggest we go wrong in our thought experiments is in the belief that if it seems to us that we have imagined things *A, B, C, ...* occurring together in some possible world  $W_n$  it automatically follows that we have really done so. ...[W]e may have succeeded in imagining all of *A, B, C, ...* but not together in a single possible world

Tye’s worry, I take it, is that one may mistake the possibility of *A* and the possibility of *B* for the possibility of *A&B*, thereby committing an elementary modal fallacy.

Similar charges can be laid at the feet of Possibility and Non-Identity. In the case of Possibility, an Arnauldian proposal would be that perhaps we mistake our failure to see the impossibility of *this* without *that* for successfully seeing the possibility.



And Tye's suggestion would be that perhaps we mistake our seeing that a possible world includes *this* and another possible world doesn't include *that* for seeing that a single possible world includes *this* but not *that*.

As for Non-Identity, a non-dualist who takes this line has two options. She may say that, after fallaciously arriving at Possibility, the dualist goes on to validly infer Non-Identity. Alternatively, the non-dualist may say that perhaps Non-Identity is itself the product of a fallacious operator shift. David Armstrong (1968, pp. 48–49) puts one such proposal this way: "It can... be suggested by the Materialist that we tend to pass from something that is true: *I am not introspectively aware that mental images are brain processes* to something that is false: *I am introspectively aware that mental images are not brain processes*."

And so perhaps, as these authors suggest, I am the victim of some fallacious operator shift, hastily passing from a proposition of the form *I can't see that p is impossible* to a proposition of the form *I can see that p is possible*. Or from a proposition of the form *I can see that p is possible and that q is possible* to a proposition of the form *I can see that (p&q) is possible*. Or from a proposition of the form *I am not aware that p* to a proposition of the form *I am aware that not-p*. The question is: do any of these proposals deliver an undercutting defeater for our dualist intuitions?

### 3.4 Response

No, these suggestions do not provide any undercutting defeaters for Possibility and Non-Identity, for two reasons. First, merely pointing out the *possibility* of a fallacious operator shift—merely suggesting that this *might* occur—is not in general sufficient to undercut beliefs held on the basis of intuition. That type of consideration does not in general require a subject to lower her confidence in her belief. And so these proposals fail to meet the necessary condition on undercutting defeat mentioned above.

Consider the seeming truth of the proposition that the Prime Minister is not a prime number. What could be more obvious than that? But now consider this suggestion from a niggling skeptic: "*Perhaps* you mistake failing to see that the Prime Minister *is* a prime number for successfully seeing that he isn't." Surely this bare possibility—this unsubstantiated suggestion—does not require that you lower your confidence in your belief that the Prime Minister is not a prime number.

Secondly, if the *mere possibility* of a fallacious operator shift were sufficient to defeat beliefs held on the basis of intellectual seemings, then all such beliefs would be vulnerable to quick and decisive defeat, including Arnauld's, Tye's, and Armstrong's beliefs that the operator shifts they mention are fallacious. After all, for example, Tye *may* unconsciously mistake his failure to see the validity of  $((\diamond p \& \diamond q) \rightarrow \diamond(p \& q))$  for successfully seeing its invalidity. But if those beliefs are so easily undercut—if neither they nor we can justifiably believe that these operator shifts are as fallacious as they seem—then their proposals collapse upon themselves.

And so it seems perfectly rational to maintain one's confidence in Possibility and in Non-Identity on the basis of their seeming truth, even upon learning that this seeming *may* be the result of an unconscious fallacious operator shift. Things would

be much different if we had solid empirical data that we actually *are* victims of a subconscious fallacious operator shift. But, as the proposals stand, they do not provide us with an undercutting defeater for the seeming truth of Possibility and Non-Identity. So let us now turn to a more empirically-motivated proposal for an undercutting defeater.

### 3.5 Dual-process cognition

In their 2011 paper, Fiala et al. (2011) propose that dual-process cognition spawns our misleading dualist intuitions. In contrast to the previously discussed suggestions, this proposal has substantial empirical evidence in its favor. The idea is that humans have a “low-road” cognitive process for attributing mental states. This process is “quick, automatic, unconscious, associative, heuristic-based, computationally simple, evolutionarily old, domain-specific and non-inferential.” It is triggered by simple, surface level features, e.g. having eyes, appearing to behave in a contingently interactive manner, and displaying distinctive (non-inertial) motion trajectories. Humans also come equipped with a “high-road” cognitive process for attributing mental states. This process is “relatively slow, controlled, introspectively accessible, rule-based, analytic, computationally demanding, inferential, domain-general, and voluntary.”

Usually, these two processes issue harmonious verdicts. Not so, however, in dualist thought experiments. There, we consider the mass of gray matter that composes the human brain. If we are non-dualists, we might deduce from our internalized theoretical beliefs that the right kind of reductive state type, tokened in the brain, is also a certain type of conscious experience. Or, if we are dualists, we presumably use this high-road cognitive process merely to entertain reductive type-identity claims.

At the same time, our low-road cognitive process does *not* categorize the eye-less, behavior-less, motionless brain as an agent. And so this low-road cognitive process falls silent as the grave, and fails to make any attributions of consciousness. We lack any quick, automatic ‘gut-feeling’ that the reductive state type in question is a certain type of conscious experience. But it would be hasty to conclude on this basis that Possibility is true: from a failure for type A to seem identical with type B, it hardly follows that A and B are not necessarily coextensional. Neither should we conclude on this basis that Non-Identity is true: from the fact that type A fails to seem like type B it doesn’t follow that A is not identical with B.

Let’s grant that this is all correct. Would it follow that we have an undercutting defeater for our dualist intuitions?

### 3.6 Response

No, it would not follow, since the proposal overgenerates defeaters. If this suggestion from Fiala et al. (2011) required that we be suspicious of our dualist intuitions, it would also cast suspicion on perfectly mundane intuitions that we know are above reproach. For example, I look down and notice that a couple of my floorboards are misaligned. I consider the proposition that *being a misaligned*

*floorboard* is identical with the feeling of pain. I immediately dismiss the suggestion out of hand. *Of course* the property of being a misaligned floorboard is not identical with pain: the two properties are not coextensive, and they are just obviously non-identical. These ‘dualist’ intuitions, I take it, are completely... above board. We should not be suspicious of them.

Yet whatever disharmony occurs between my dual cognitive processes in the standard dualist thought experiments also occurs in the case of the misaligned floorboards. My “high-road” cognitive process entertains a certain identity claim: being a misaligned floorboard just is pain. My “low-road” cognitive process does not consider the eye-less, behavior-less, motionless floorboards to be agents. That low-road process therefore does not attribute consciousness to the floorboards; it’s silent. If this type of disharmony between my dual cognitive processes is sufficient to undercut my dualist intuitions, it should also undercut my floorboard intuitions. But since I can rationally maintain my confidence in my floorboard intuitions, this proposal from Fiala et al. (2011) does not undercut my dualist intuitions.

This hints that there is more to the story that Fiala et al. (2011) recognize. Perhaps I have some other cognitive process that issues verdicts about what types of states *could not be* conscious states. For all Fiala et al. tell us, this third process might be perfectly reliable, and it may be operative in the standard dualist thought experiments.

### 3.7 Duped by our concepts?

Let’s now discuss what I believe to be the most popular and plausible attempt to undercut dualist intuitions. Many philosophers believe that a close investigation of the *concepts* that figure in Possibility and Non-Identity will furnish us with an undercutting defeater for each.<sup>14</sup> These philosophers typically claim that the culprit is a lack of a priori entailment relations between some relevant concepts. Typically, they say that there are no substantive a priori ties between our phenomenal and physical (or physiological) concepts. That is, there are no non-tautologous a priori knowable inferences from thoughts containing phenomenal concepts to thoughts containing physical (or physiological) concepts, and vice versa. You will find this thought in Block (2006, p. 53), Hill (1997, p. 75), Loar (2003, pp. 115–116), Nagel (1998, Sect. 4), Papineau (2002, p. 87), Tye (1999, p. 715), and many others.

So the suggestion is that the concepts that figure crucially into Possibility and Non-Identity are, from the armchair at least, really just silent with respect to each other. It is not Possibility itself that intellectually seems true. Rather, we really only *fail* to see the impossibility of this without that. And it is not Non-Identity itself that intellectually seems true. Rather, we really only *fail* to see that this is identical with

<sup>14</sup> In what follows, the philosophers I discuss typically do not construe dualist intuitions as crucially involving only demonstrative concepts like THIS and THAT. However, I believe that they would mean what they say about specifically phenomenal and physical concepts—like PAIN and C-FIBERS FIRING—to apply, for even stronger reasons, to stripped-down demonstrative concepts. After all, demonstrative concepts stand in fewer substantive entailment relations than full-blooded concepts like PAIN and C-FIBERS FIRING.

that. And of course, as Arnauld and Armstrong pointed out, these are bad grounds on which to believe Possibility and Non-Identity.

Now, in order to improve on the mere *suggestions* from Arnauld and Armstrong, and in order get a successful undercutting defeater, it must be the case that, in the dualist thought experiments, we are in fact (and not merely *possibly*) disposed to believe dualism on these poor grounds. This, I take it, is why Nagel (1998, Sect. 4) says that “our concepts fail to reveal a necessary connection, and we are tempted to conclude to the absence of any such connection.” Elaborating on Nagel’s proposal, Hill (1997, pp. 75–78) asserts that we are *in fact* endowed with an unreliable psychological mechanism the function of which is to churn out the belief that the referents of any two concepts could come apart, whenever there are merely no substantive a priori ties between the concepts (and no immediately accessible sufficient a posteriori reasons to think that the concepts corefer).

I interpret Nagel and Hill as asserting something like the following:

(5) For any concepts C1 and C2 and normal human subject S, if C1 and C2 have no substantive a priori ties, and if S lacks sufficient a posteriori reason to think C1 and C2 corefer, then it won’t intellectually seem metaphysically impossible to S for the referents of C1 and C2 to come apart, and it will seem to S *on this basis* that it’s metaphysically possible for the referents of C1 and C2 to come apart.

If (5) is true and its antecedent is met, as Nagel and Hill believe, then the seeming truth of dualism is poorly based, and so it is not a reliable indicator of or does not warrant belief in its actual truth. This proposal, therefore, would give us an undercutting defeater for Possibility (and Non-Identity, which, presumably on this view, we infer from Possibility). However, if (5) is false, or if its antecedent is not met, then this proposal fails to deliver an undercutting defeater.

### 3.8 Response

(5) is false, and so this proposal fails to deliver an undercutting defeater. The second “seem” in the consequent of (5) may refer to an intellectual seeming, or it may refer to a disposition to believe. Here’s a counterexample to (5) on both interpretations: Something somewhere is named “Chomolungma,” but I am not telling you what or where it is. You now have the concept CHOMOLUNGMA in your cognitive economy. CHOMOLUNGMA has no substantive a priori entailment relations with your concept MOUNT EVEREST, and you lack sufficient a posteriori reason to think that these two concepts corefer.

So, in this case, the antecedent of (5) is met, and we are halfway to a counterexample. How about the consequent? Well, it doesn’t intellectually seem metaphysically impossible for their referents to come apart, as (5) predicts. Yet, importantly, neither does it intellectually seem *metaphysically* possible for the referents to come apart. For if it really struck us as genuinely possible for the referents to come apart, then it would strike us as obvious that the referents are distinct (since actual identity in this case clearly entails necessary identity). But it doesn’t seem obviously so. Nor are you inclined to judge that it is genuinely

possible for the referents to come apart. You are just agnostic about the identity claim—neither it nor its negation intellectually seems true, and you are not inclined to believe either. And so (5) is false no matter how we take the verb “to seem” in the consequent.

Also, (5) overgenerates undercutting defeaters, and this is an additional reason to believe it is false. The Nagel–Hill proposal would have us be suspicious of some unassailable intuitions of distinctness. For example, consider your phenomenal concept PAIN and your physiological concept ANGIOGENESIS. (Angiogenesis is the formation and development of blood vessels.) I take it to be obvious that the felt quality type of pain is not (and metaphysically could not be) the formation of blood vessels (that event-type). Clearly and uncontroversially, something *could* be a token of one type without being a token of the other.

But, if Nagel and Hill are right about phenomenal and physiological concepts, these two concepts have no substantive a priori entailment relations between them. And I take it that you don’t have *any* a posteriori reason to think that ANGIOGENESIS and PAIN corefer, let alone sufficient reason. And so the antecedent of (5) is true in this case. And, as (5) predicts, it does seem metaphysically possible for the referents of PAIN and ANGIOGENESIS to come apart. But then it would follow from (5) that I should be suspicious of my belief that the felt quality type of pain is not necessarily coextensional with the formation of blood vessels (that event-type). For, on these suppositions, it seems metaphysically possible for pain and angiogenesis to come apart on bad grounds: it seems possible not because it really *is* possible, but only on the basis of a contingent fact about my concepts. But clearly I shouldn’t be suspicious of that intuition. So this proposal overgenerates undercutting defeaters, and so we shouldn’t accept the proposal.

Apart from (5), I can find no reason within a bull’s roar to think that actually, our belief in Possibility and our belief in Non-Identity are based on bad grounds. And recall from the discussion of Arnauld, Tye, and Armstrong that it is insufficient merely to point out that *perhaps* these beliefs are based on bad grounds. So I conclude that, as it stands, Nagel and Hill offer us no undercutting defeater.

Furthermore, Nagel and Hill provide us with an undercutting defeater only if the antecedent of (5) is met with respect to our dualist intuitions. But there are good reasons to think that the antecedent of (5) is not met here, since there are good reasons to believe that there *are* substantive a priori ties between phenomenal and physical (or physiological) concepts. Consider PAIN and ANGIOGENESIS again. Can’t we know a priori that the referents of these concepts are not necessarily coextensional? But then there *are* a priori entailment relations between at least some of our phenomenal and our physical (or physiological) concepts. And consider the a priori ties between THIS and THAT as deployed in the autocerebroscope case. In Possibility and Non-Identity we see substantive a priori entailment relations between THIS and THAT (as deployed during the autocerebroscope case), and the intuition is only strengthened when we swap out the demonstrative concepts for non-demonstrative phenomenal and physical (or physiological) concepts. So here we find good reason to believe that the antecedent of (5) isn’t met, and therefore additional reason to believe that Nagel and Hill do not provide us with an undercutting defeater.

Here's another reason to believe that there are substantive a priori ties between our phenomenal and our physical (or physiological) concepts. If there are no substantive a priori entailment relations between any two concepts for someone, then it won't seem to her metaphysically impossible for the referents of these concepts to come apart, but it *also* won't seem metaphysically impossible to her that the referents be *identical*. The concepts will just be silent with respect to each other. (Consider CHOMOLUNGMA and MOUNT EVEREST again.) Yet in the dualist thought experiment, it *does* seem metaphysically impossible that the referents of THIS and THAT be identical (since their referents seem actually distinct and actual distinctness clearly entails necessary distinctness).

Finally, if there are no substantive a priori entailment relations between any two concepts for some normal human subject S, then sufficient empirical information will lead S to believe with no compunction that the concepts corefer. For example, if I gave you reason to believe that CHOMOLUNGMA refers to Mount Everest (as it does), then you would have no problem believing that the relevant concepts corefer. However, no amount of the relevant empirical information closes the explanatory gap in the philosophy of mind—dualist intuitions persist even in light of all the relevant empirical data. As Papineau (2002, p. 161) says: “even given all the arguments, intuition continues to object to mind-brain identity.” And Tye (1999, p. 706) says that the explanatory gap remains open “even for those who understand full well the relevant phenomenal terms and who know the underlying physical and functional story.” Therefore, again, the antecedent of (5) isn't met, and therefore Nagel and Hill do not provide us with an undercutting defeater.

Let's now leave Nagel and Hill behind and move to another type of proposal for an undercutting defeater from Thomas Polger.

### 3.9 Insufficient grasp of relevant concepts?

Thomas Polger (2004, p. 49ff) offers a similar account of our dualist intuitions, which he believes are misleading. According to Polger, the fault lies with our insufficient understanding of brain states. Polger thinks that we have, in an important way, failed to fully grasp the concept of a brain state, since we do not know the identity conditions of brain states. And he thinks that, in general, if we are uninformed about the identity conditions of either Xs or Ys, then even if  $X = Y$  it might seem that it could be otherwise. “Thus arises the appearance of contingency,” he says. This is a conceptual failure on our part. Our anemic grasp of the concept of brain states lets us deem them candidates for identity with phenomenal states, and yet may also cause the relevant identity claim to appear contingent even if it isn't.

This proposal bears a family resemblance to the previous one. The idea, I take it, is that our concept of a brain state is, due to our weak grasp on it, really just silent with respect to our phenomenal concepts. Since we don't fully understand the identity conditions of brain states, we fail to see that they just are phenomenal states. And, Polger believes, dualists mistake their *failure* to see that brain states *are* phenomenal states for successfully seeing that they are not. This idea is not unmotivated, as Polger shows with the example of Thingamajigs and Whatchamacallits. If—as suggested by their names—you are unclear on the nature of

Thingamajigs and Whatchamacallits, you may mistake the *epistemic* possibility of their distinctness for the *metaphysical* possibility. You may mistake, that is, your inability to rule out their distinctness for your ability to rule it in. This is a tempting slip, at least in the case of Thingamajigs and Whatchamacallits. And perhaps we are making the same mistake when it comes to dualism. If we learn that we are making this mistake, or even that we easily might be, this would successfully undercut our dualist intuitions.

### 3.10 Response

But there are at least two reasons to think that we're not making this mistake when it comes to our dualist intuitions. As with the proposal from Nagel et al., Polger's proposal overgenerates undercutting defeaters. We hold many beliefs about brain states that are obvious, uncontroversial, and based on intellectual seemings, and yet which would fall under a cloud of suspicion were Polger correct. Therefore, Polger is not correct.

For example, it seems obviously true that brain states are not numbers, and that brain states are not earthquakes. But of course if, as Polger says, we are in the dark about the identity conditions of brain states and therefore at risk of mistaking our inability to see the truth of identity claims involving brain states for our ability to see their falsity, then it would follow that we should doubt the obvious truth that brain states are not earthquakes. But since that seeming truth is clearly above reproach, something must have gone wrong with Polger's proposal.

Polger may reply that we are not *completely* in the dark when it comes to the identity conditions of brain states. We know enough about their identity conditions to see that brain states are not earthquakes or numbers, but not enough to see that brain states are not phenomenal states. This is a promising line, but Polger needs to say more to support this claim. How might we distinguish—in a principled way—between earthquakes and numbers on the one hand and phenomenal states on the other, so it comes out that we know enough about brain states to rule out identity with earthquakes or numbers but not phenomenal states? (It can't be that numbers and earthquakes are better known to us than the contents of our own minds.) Unless he answers that question, Polger's proposal does not provide us with an undercutting defeater for our dualist intuitions.

There is also a second problem with Polger's proposal. Consider his Thingamajigs and Whatchamacallits example again. Since we are clueless about the identity conditions of these things, we may feel a pull toward thinking that they are possibly distinct. But note well that we feel an equally strong pull toward thinking that they are possibly *identical*. We'd naturally say "Maybe they're identical, but maybe they're not," and we'd be reporting epistemic possibilities, i.e. our inability to rule out their identity as well as our inability to rule out their distinctness. We're on the fence, as disposed to accept the identity claim as to reject it.

And so, if dualist thought experiments play out on this same stage, with a similarly weak grasp on the concept of a brain state, then we should be just as tempted to conclude that brain states are phenomenal states as we are to conclude that they are distinct. They should be candidates for identity as well as candidates

for distinctness in the standard thought experiments. But we are not so tempted: brain states seem like the wrong kind of thing to be conscious states. The standard thought experiments don't sit us on the fence with respect to whether phenomenal states are, for example, brain states. Rather, it clearly seems only that they *aren't*. And so Polger's proposal does not capture the data; it issues a prediction that is falsified by the data. And so I again conclude that Polger fails to provide us with an undercutting defeater for our dualist intuitions. Let's now turn to a final proposed undercutting defeater, this time from David Papineau.

### 3.11 Papineau's proposal

Papineau's strategy for undercutting our dualist intuitions goes like this: call the way in which we think about the phenomenal character of pain from the inside a *phenomenal* concept. (I'll use "PAIN<sub>p</sub>" to mention that phenomenal concept.) When we think about conscious experiences in this phenomenal way, when we "deploy" or "exercise" phenomenal concepts, the concepts themselves exemplify or stimulate versions of their respective conscious states, according to Papineau. So when we think about pain in a phenomenal way, when we deploy PAIN<sub>p</sub> imaginatively, Papineau (2002, p. 170) says that "we activate a 'faint copy' of the experience referred to. (You know what he's talking about. Close your eyes and think about the color red. You'll get a faint image of red "in your mind's eye," as they say. This phenomenon is hard to describe, but all too familiar.) When we deploy a phenomenal concept introspectively, according to Papineau (*ibid.*), we amplify the experience referred to into a 'vivid copy' of itself."

On the other hand, according to Papineau, non-phenomenal concepts do not do this, and so we feel that non-phenomenal concepts "leave out" the feelings themselves. However, the mere fact that non-phenomenal concepts "leave out" these faint or vivid copies does not for a moment suggest that they do not refer to sensations. My non-phenomenal concept C-FIBERS FIRING and my phenomenal concept PAIN<sub>p</sub> may still corefer, even though the former "leaves out" the phenomenology that the latter activates. Similarly, the concepts LAUGHING GAS and N<sub>2</sub>O may corefer (in fact they do), even if that seems incredible since these concepts activate radically different mental images in us.

According to Papineau, we succumb to the fallacy of thinking that the reason Possibility and Non-Identity seem true has anything to do with the *referents* of our concepts. Rather, he says, we are committing a use-mention fallacy. A third-person, non-phenomenal way of thinking might not *use* conscious experiences in the way that a first-person, phenomenal way of thinking does. But this fact does not imply that the non-phenomenal concept does not *mention* the same thing that the phenomenal concept does. So although it may seem to us that Possibility and Non-Identity are true, this in no way warrants belief in the truth of Possibility and Non-Identity, since the seeming is caused by facts about the functional roles of our concepts, not by facts connected to the *truth* of Possibility and Non-Identity. Possibility and Non-Identity would seem true whether or not they were true. We are being fooled by a contingent feature of our concepts, and this is a poor ground on



which to believe Possibility and Non-Identity. And so, if Papineau is right, our dualist intuitions are undercut.

### 3.12 Response

Let's think about Papineau's argument for the conclusion that we have an undercutting defeater in the case of Possibility and Non-Identity. First, Papineau asserts this proposition:

- (6) For normal human subjects, deployment of non-phenomenal concepts "leaves out" something that deployment of non-phenomenal concepts doesn't.

In addition, Papineau points out the following truth:

- (7) For any concepts C1 and C2, and subject S, the fact that *S's deployment of C1 "leaves out" something that S's deployment of C2 doesn't* does not render probable or warrant belief that *C1 and C2 are not necessarily coextensional (and therefore have distinct referents)* for S.

So Papineau points out a bad basis on which one might judge the truth of Possibility and Non-Identity, namely this contingent feature on our concepts. However, in order to provide an undercutting defeater, Papineau must give us some reason to think that it is not merely *possible* that Possibility and Non-Identity seem true on this bad basis, but that they *actually* do. Here's an unpromising strategy:

- (8) For any concepts C1 and C2, and normal human subject S, if S's deployment of C1 "leaves out" something that S's deployment of C2 does not, then it will seem true to S *on this basis* that C1 and C2 are not necessarily coextensional and therefore have distinct referents.

If true, (8) would [in combination with (6) and (7)] furnish us with an undercutting defeater, a reason to think that the basis on which we judge Possibility and Non-Identity to be true does not warrant belief in Possibility or Non-Identity.

However, this strategy would overgenerate undercutting defeaters, i.e. it would give us reason to doubt the seeming truth of propositions that we rightly take to be indubitable. For example, suppose my friend, hung up on Ockham's razor and seeking to scale down his ontology, proposes that the felt quality type of pain is identical with the felt quality type of euphoria. I consider the identity, deploying the relevant phenomenal concepts. Deployment of each concept "leaves out" something that deployment of the other does not. And it seems obviously true to me that the referents are not necessarily coextensional, and (therefore) distinct. But, if (7) and (8) are true, I should be suspicious of this seeming. After all, if, as (7) and (8) say, distinctness of referent and non-necessary coextension will seem true on a bad basis (viz. a contingent feature of the concepts), then I ought to refrain from judging that things are as they seem. But clearly this judgment is not suspicious—it is as obviously true as anything can be. Therefore Papineau's proposal overgenerates defeaters, and so it's false. There must be some other, good basis on which I non-fallaciously judge distinctness and non-necessary coextension, and Possibility and Non-Identity may be held on this other, good basis.

(8) is also plagued by straightforward counterexamples. Suppose I overhear some friends discussing Smith's favorite color. I don't know what color they're talking about, but a desire to name it "Kevin" surges up within me and I submit to the urge. I then wonder "Is Kevin identical with red?" In considering the identity, I deploy my phenomenal concept RED<sub>p</sub> and my non-phenomenal concept KEVIN. Deployment of the latter concept leaves out something that deployment of the former concept does not, namely a "faint copy" of red. And yet it doesn't at all seem to me that the referents are distinct or not necessarily coextensional; I'm completely agnostic on the question. But (8) predicts otherwise. Therefore, (8) is false, and—as it stands—Papineau fails to provide us with an undercutting defeater.

#### 4 Might these proposals play another dialectical role?

I take myself to have shown that the proposals we considered in Sect. 3 do not require that we lower our confidence in our dualist intuitions, and therefore they are not successful undercutting defeaters. I construed these proposals as attempts to undercut Possibility and Non-Identity, and thereby—either on their own, or in combination with rebutting defeaters—deconvert dualists. But one might worry that I have mischaracterized the role that the proposals are meant to play, and that I am judging them by too high a standard. This worry has three parts, which I will number (i)-(iii).<sup>15</sup>

The worry is that (i) these proposals are *not* in fact intended as undercutting defeaters. Instead: (ii) they are merely meant to explain why we ever had these intuitions, and why the intuitions may persist even after one has accepted non-dualism. Perhaps, that is, the proposals are meant to explain—in the wake of a convincing argument for non-dualism or an objection to dualism—why we were ever misled.<sup>16</sup> *Pace* me, one might insist that (iii) it is up to the standard *rebutting* defeaters to deconvert dualists, and the proposals of Sect. 3 are merely attempts to diagnosis the origin of our error only *after* deconversion.

In closing, I'll have a word with each of (i)-(iii). First, consider the suggestion that (i) these proposals are *not* in fact intended as undercutting defeaters. On the contrary, many philosophers do explicitly intend to explain *away* our dualist intuitions, and not merely explain why they arise. For example, Hill (1997, p. 70) says:

I have not yet indicated how it might be true that our Nagelian account explains these intuitions *away*. One explains a set of intuitions by describing the mechanisms that produce them. To explain the members of the set *away* one must in addition provide evidence which calls the reliability of the relevant mechanisms into question, i.e., one must produce evidence which

<sup>15</sup> I thank an anonymous referee of this journal for encouraging me to consider this worry.

<sup>16</sup> As Nagel (1974, p. 446) says as a preface to his original proposal, "A theory that explained how the mind-brain relation was necessary would still leave us with Kripke's problem of explaining why it nevertheless appears contingent." It may be that the proposals I have discussed are aimed merely at that second problem, and are not intended to lower our confidence in dualism.

makes it reasonable to doubt that the intuitions produced by these mechanisms are quite likely to be true.

Hill later attempts to provide that evidence—examined in Sect. 3—and so he explicitly intends to undercut our dualist intuitions. He shares this project with Polger (2004, p. 62), who explicitly considers it a virtue of his account that it explains *away* the intuition that my pain might not be a C-fibers firing, as well as the intuition that my C-fibers firing might not a pain. We have here, then, two prominent philosophers who explicitly intend their proposals to lower our confidence in dualism. But then we need not worry that, in general, we have so far misconstrued the dialectical role that the proposals of the previous section are intended to play.

Secondly, consider the suggestion that (ii) these proposals are intended not to lower our confidence in dualism, but rather only to satisfy a certain curiosity in the wake of successful rebutting defeaters. In response, I grant that some of philosophers I have discussed do not *explicitly* intend to explain away our dualist intuitions. Let's suppose for the moment that these philosophers intended, as (ii) claims, for their proposals to play no role in the dialectic except in combination with a cogent argument for non-dualism. Even if that's right, this explanatory project requires that their proposals reduce our confidence in dualism on their own. And so if their proposals fail to reduce our confidence in dualism, as I have shown in this paper, so too fails their explanatory project. Let me explain.

Even if the proposals of Sect. 3 explicitly intend only to explain where our dualist intuitions come from, naturally the proposals must do this *in a way consistent with non-dualism*. After all, it would be odd indeed for a non-dualist to explain the origin of our dualist intuitions in a way that entailed the truth of dualism. And so consider for example Papineau (1993, p. 169) who, before presenting his diagnosis of dualist intuitions, says: "My aim will be to explain how these intuitions arise, and why they do not discredit physicalism." Suppose Papineau succeeds in this explanatory enterprise, and consider René who accepts dualism at least in part on the basis of dualist intuitions. René thereby takes these dualist intuitions to count against physicalism. Should René encounter Papineau's explanation—which we're granting for the moment successfully shows that dualist intuitions do *not* count against physicalism—René must lower his confidence in dualism. Therefore, to be successful, Papineau's explanation of the origin of our dualist intuitions must deliver a defeater *en passant*. And this holds generally of anyone interested in explaining the origin of dualist intuitions in a way consistent with non-dualism: a successful explanation of this kind must also defeat our dualist intuitions. But since I have shown in this paper that not one of the proposals discussed in Sect. 3 requires that we lower our confidence in our dualist intuitions, it follows that they cannot successfully play the alternative dialectical role that the above worry suggests.

Finally, consider the claim that (iii) the proposals of §3 have no defeating power on their own, but rather serve only to explain, in the wake of cogent rebutting defeaters, the origin of our dualist intuitions. This suggestion would, if true, seriously weaken the case for non-dualism. Without the aid of undercutting defeaters for our dualist intuitions, it should be difficult for many of us to get a successful rebutting defeater. After all, a rebutting defeater in this case is (or would be easily translated into) a

deductively valid argument for the truth of non-dualism or the falsity of dualism. And deductively valid arguments are merely invitations to compare subjective probabilities of an inconsistent set of propositions. On the one hand, we consider the conjunction of the argument's premises together with each of the argument's inferences in the form of conditionals. On the other hand, we consider the negation of the argument's conclusion. If the negation of the conclusion of a deductively valid argument deserves sufficiently higher credence than the conjunction of the premises and inferences, then the right thing to do is to reject the conjunction and maintain the negation of the conclusion. One should suspect some catch in the argument, even if one cannot identify the catch.<sup>17</sup>

Now, suppose someone—call her “Smith”—accepts dualism at least partly on the basis of the striking truth of either Possibility or Non-Identity. Suppose, for example, that Smith accepts dualism in part because, as Papineau (2002, 3) says “we feel it is *obvious* that conscious states are not material states.” Her credence in dualism on the basis of these intuitions is therefore quite high—perhaps as high as Cartesian certainty, though likely a few notches down. As explained in the previous paragraph, to rationally deconvert Smith without the aid of undercutting defeaters, a rebutting defeater would require a conjunction of premises and inferences that seemed *more* obvious to her than dualism. This would be a herculean task, given how quickly subjective probabilities diminish when we conjoin (and thereby multiply the probabilities of) propositions that are less than absolutely certain. To rationally deconvert Smith without the aid of undercutting defeaters, then, *every single step* of a rebutting defeater would have to be at least as obvious to her as her dualist intuitions. I, for one, have never witnessed an argument against dualism (or for non-dualism) soar to such Olympian heights. And I don't expect to.

Therefore, to accept (iii)—that the proposals of §3 have no defeating power on their own but rather merely sweep up behind cogent rebutting defeaters—is to demand that the non-dualist tie one hand behind his back and then jump rope. Since successful rebutting defeaters should be hard to come by without the aid of undercutting defeaters for our dualist intuitions, it's unclear how much value these proposals will have, if any. Better for the non-dualist to construe the proposals of §3 as undercutting defeaters, as I have in this paper.

And yet, if I am right that the proposals discussed in Sect. 3 do not successfully undercut our dualist intuitions, then all the burden of deconverting dualists really does fall on rebutting defeaters. And if dualism strikes a person as sufficiently obvious, then, for reasons given in the previous paragraph, it is unlikely that any rebutting defeater can rationally require the dualist to change her mind. So, if the arguments in this paper are sound, dualism is more epistemically respectable than its

<sup>17</sup> For example, consider this argument for the conclusion that  $2 = 1$ : First, let  $a = b$ , where  $a$  and  $b$  are non-zero quantities. Then, multiply both sides by  $a$ , to get  $a^2 = ab$ . Now subtract  $b^2$  from both sides, to get  $a^2 - b^2 = ab - b^2$ . Next, factor both sides to get  $(a - b)(a + b) = b(a - b)$ . Then, divide both sides by  $(a - b)$  to get  $(a + b) = b$ . Since we were given that  $a = b$ , we can infer that  $b + b = b$ . By combining like terms on the left, we get that  $2b = b$ . Finally, if we divide both sides by the non-zero  $b$ , we get the conclusion that  $2 = 1$ , Q.E.D.

But, of course, since the conclusion is so obviously false, the proper response is to reject the argument and maintain that  $2 \neq 1$ , even if one cannot identify the error in the argument.

current popularity in the academy suggests. Dualism strikes us as true, dualist and non-dualist alike. Our tour of the landscape has turned up dim prospects for successful rebutting defeaters and not one successful undercutting defeater. As things stand, then, dualism is well-motivated and undefeated, so rationality demands that we believe it.

**Acknowledgments** This paper benefited greatly from helpful conversations with Nathan Ballantyne, David Barnett, George Bealer, John Bengson, Dan Bonevac, Rob Koons, Tim Pickavance, Christopher Shields, David Sosa, and Malte Willer. Special thanks are due to Adam Pautz and Michael Tye, two unfailing sources of insightful comments and objections, and to J.P. Moreland, for first introducing me to these issues.

## References

- Armstrong, D. (1968). The headless woman illusion and the defense of materialism. *Analysis*, 29(2), 48–49.
- Bealer, G. (1992). The incoherence of empiricism. *Proceedings of the Aristotelian Society Supp.*, 66, 99–138.
- Bealer, G. (1996). A priori knowledge and the scope of philosophy. *Philosophical Studies*, 81(2–3), 121–142.
- Block, N. (2006). Max black's objection to mind-body identity. In D. Zimmerman (Ed.), *Oxford studies in metaphysics* (Vol. III). Oxford: Oxford University Press.
- Block, N., & Stalnaker, R. (1999). Conceptual analysis, dualism, and the explanatory gap. *Philosophical Review*, 108(1), 1–46.
- Bogardus, T. (forthcoming). Knowledge under threat. *Philosophy and Phenomenological Research*.
- Dennett, D. (1992). *Consciousness explained*. New York: Back Bay Books.
- Descartes, R. (1984). In J. Cottingham, R. Stoothoff, & D. Murdoch (Eds.), *The philosophical writings of descartes* (Vol. II). Cambridge: Cambridge University Press.
- Feyerabend, P. (1963a). Mental events and the brain. *Journal of Philosophy*, 40, 295–296.
- Feyerabend, P. (1963b). Materialism and the mind-body problem. *Review of Metaphysics*, 17(1), 49–66.
- Fiala, B., Arico, A., & Nichols, S. (2011). On the psychological origins of dualism. In E. Slingerland & M. Collard (Eds.), *Creating consilience: Issues and case studies in the integration of the sciences and humanities*. New York: Oxford University Press.
- Hill, C. (1997). Imaginability, conceivability, possibility, and the mind-body problem. *Philosophical Studies*, 87, 61–85.
- Hill, C. (2005). Remarks on David Papineau's Thinking about consciousness. *Philosophy and Phenomenological Research*, 71(1), 147–154.
- Huemer, M. (2007). Compassionate phenomenal conservatism. *Philosophy and Phenomenological Research*, 74, 30–55.
- Kripke, S. (1972). *Naming and necessity* (lecture III). Cambridge, MA: Harvard University Press.
- Loar, B. (2003). Qualia, properties, modality. *Philosophical Issues*, 13(1), 113–129.
- McGinn, C. (2003). What constitutes the mind-body problem? *Philosophical Issues*, 13(1), 148–162.
- Nagel, T. (1974). What is it like to be a bat? *The Philosophical Review*, 83(4), 435–450.
- Nagel, T. (1998). Conceiving the impossible and the mind-body problem. *Philosophy*, 73(285), 337–352.
- Papineau, D. (1993). Physicalism, consciousness, and the antipathetic fallacy. *Australasian Journal of Philosophy*, 71(2), 169–183.
- Papineau, D. (2002). *Thinking about consciousness*. Oxford: Oxford University Press.
- Plantinga, A. (2006). Against materialism. *Faith and Philosophy*, 23(1), 3–32.
- Polger, T. (2004). *Natural minds*. Cambridge, MA: MIT Press.
- Pollock, J. L. (1974). *Knowledge and Justification*. Princeton: Princeton University Press.
- Rorty, R. (1965). Mind-body identity, privacy and categories. *Review of Metaphysics*, 19, 25–54.
- Stich, S. (1991). Do true believers exist? *Aristotelian Society Supplement*, 65, 229–244.
- Stich, S. (1996). *Deconstructing the mind*. New York: Oxford University Press.
- Tye, M. (1986). The subjective qualities of experience. *Mind*, 95(377), 1–17.
- Tye, M. (1995). *Ten problems of consciousness*. Cambridge, MA: MIT Press.

- Tye, M. (1999). Phenomenal consciousness: The explanatory gap as cognitive illusion. *Mind*, 108(432), 705–725.
- Tye, M. (2000). *Consciousness, color, and content*. Cambridge, MA: MIT Press.
- Tye, M. (2006). Another look at representationalism about pain. In M. Aydede (Ed.), *Pain: New essays on its nature and the methodology of its study*. Cambridge, MA: MIT Press.
- Tye, M. (2007). Intentionalism and the argument from no common content. *Philosophical Perspectives*, 21, 589–613.
- Van Inwagen, P. (2009). “Why a mind-body problem?” video interview with Robert L. Kuhn for the television program *Closer to Truth*. <http://www.closetotruth.com/video-profile/Why-a-Mind-Body-Problem-Peter-van-Inwagen-/150>