

## *Neural Materialism, Pain's Badness, and A Posteriori Identities*

IRWIN GOLDSTEIN

Materialists say sensations and other kinds of mental states are physical events. Today, most materialists are neural materialists. They think mental states are neural events or material properties of neural events.

Orthodox neural materialists think mental states are neural events or orthodox material properties of neural events. Orthodox material properties are defining properties of the physical. A *defining property* of the physical is a type of property that provides a necessary condition for something's being correctly termed 'physical' (a conjunction of all defining properties provides a sufficient condition.) Defining properties of the physical include *spatial* and *temporal* properties and *causal propensities* and *sensitivities*. A particle is an *electron*, for instance, by having a particular set of spatiotemporal properties and causal sensitivities and powers.

In this paper I give an argument against orthodox neural materialism. If successful, the argument would show at least some properties of some mental states are not orthodox material properties of neural events. The argument is not, however, a self-sufficient refutation of orthodox materialism. To support one of the argument's premises I appeal to views of pain, pleasure, and value I do not defend here (though defend in detail elsewhere). My thesis in this paper is this: If these views are correct, we refute orthodox neural materialism with the argument presented here.

The contentions about pain, pleasure, and value are the following. First, ethical realism is true. 'Good' and 'bad' name properties.

An object is correctly called 'good' or 'bad' if it has the property the word names. Second, pain's intrinsic badness is an "objective" property of pain in that it does not hinge on our disliking and avoiding pain. Rather, being intrinsically bad, pain has an intrinsic character that grounds these reactions and so gives agents reason to dislike and avoid pain. (Good and bad are analyzable. Pleasure is good, approximately, by having a character that gives us reason to seek pleasure. Pain's badness is analyzable in a parallel way.) Third, intrinsic badness is a defining property of 'pain'. When 'pain' is used in the broad sense in which it encompasses all unpleasant experiences, and so both localized and non-localized pain, pain's badness is a property every pain has and a property by which localized pain sensations and non-localized emotional pain are, in qualitative character, a single kind of mental state. Finally, pleasure's link to intrinsic goodness parallels pain's link to intrinsic badness. Given this analysis, pain and pleasure are opposites through their intrinsic axiological character.<sup>1</sup>

Many people both (1) affirm causal role analyses of mind, and (2) claim token mental states are neural states in sentient beings on Earth. People who both affirm (1) and accept (2) commit to orthodox materialism. This conjunction of views is one target of my argument against orthodox materialism.

Causal theorists say a mental state's defining or essential properties lie in its causes and effects or "causal role." An event is pain, for instance, by having a particular set of causes and effects. While causal theorists say there could be sentient beings in which immaterial substances are the bearers of a mental state's causal role, theorists regularly say *neural states* are the bearers in sentient beings on Earth. Causal theorists who say (token) mental states are neural states in Earthlings commit to orthodox materialism for Earthlings. (If they reject orthodox materialism in their view of Earthlings, their thinking has internal inconsistencies.) For Earthlings, causal role theorists commit to an ontology of neural states and their causal roles, i.e., sets of their causes and effects. (Some causal theorists identify mental

---

1 The views of pleasure and pain, and the realist objectivist analysis of intrinsic value, I defend in various places. These include Goldstein 1980, 1989, 1994 (50–52), 2000, 1980 and 2002.

states with the *causal roles*. Others identify (token) mental states, in Earthlings, with the *neural events* (they say) have these causal roles. Either way, their ontology is neural events and causal roles.) A neural event's causal role, its causes and effects, is an *orthodox material property*: It is a defining property of the physical – a kind of property that gives a necessary condition for being physical.<sup>2</sup>

What I call *orthodox materialism* is not what others call *reductive materialism*. (I do not use the reductive–anti-reductive distinction.) Many people insist causal theorists are *anti-reductive*. I call them orthodox materialists. My argument against orthodox materialism targets both what people consider reductive materialism (*viz.*, the identity theory) and views people consider anti-reductive.

Eliminativism is not a form of orthodox neural materialism, and so my argument does not oppose it. Orthodox neural materialists identify mental states with elements in the neural orthodox material. Eliminativists do not. They deny the existence of mental states.

## I. The Argument

Here is my argument: 1. If orthodox neural materialism is true, the exhaust and dispense principle is true. 2. The exhaust and dispense principle is false. Therefore, 3. Orthodox neural materialism is false.

Both premises mention *the exhaust and dispense principle*. The principle has two components: the *exhaust* and *dispense* theses.

- i. The *exhaust* thesis: It is possible, in principle, to identify every kind of mental state, and to specify every immediate property every mental state has, using the vocabulary of neuroscience and other physical sciences, i.e., with the kinds of words physical scientists use to identify neural

---

<sup>2</sup> Some causal theorists mention a mental state's causal links to other mental states in their analyses. (They may mention 'desire' when defining 'pain'. They might partially define 'pain' by saying that what it names "causes a desire to flee.") Even so, causal theorists who "inter-define" psychological words so, do not intend to affirm non-causal, irreducibly psychical properties. Emphasizing causal roles and token mental-neural identities, they work from an ontology of neural states and causal roles.

events and specify their orthodox material properties. While declaring it is possible *in principle* to specify properties of mental states with a physical science vocabulary, proponents of the exhaust thesis need not suppose it is possible *in practice* for people to do so.

- ii. The *dispense* thesis: People might dispense with psychological words. They would still be able, in principle, to identify every mental state and to specify every property every mental state has. They might do this with physical science words.

Premise 1 is true. We have words with which we can identify neural events and specify their spatiotemporal and other orthodox material properties. Were mental states neural events or their orthodox material properties, it would be possible in principle to identify mental states with the physical science words we might use to identify neural events and their orthodox material properties.<sup>3</sup> Thus, the exhaust thesis is true if orthodox materialism is true. Further, if the exhaust thesis is true, the dispense thesis also is. If we could identify mental states using physical science language, we could identify them without using psychological words. Thus, both the exhaust and dispense theses are true if neural materialism is true.

Some people will think the exhaust and dispense thesis is true, and hence that premise 2 is false. Part II contains my defense of premise 2.

## II. Premise 2 Defended

Suppose we dispense with psychological words. We restrict ourselves to words we might use to specify neural events and their orthodox material properties. We use words for spatiotemporal properties and causal sensitivities and propensities. We use the vocabulary of neuro-

---

<sup>3</sup> To say this is not directly to comment on the meanings of words. It is not directly to claim, for instance, that if pain were neural event *n* or physical property *p* of *n*, 'pain' would have the same meaning as a set of physical science words we might use to specify *n* or *p*.

science. When we are constrained so, there are some defining properties of 'pain' and 'pleasure' – pain's intrinsic badness and pleasure's intrinsic goodness – we cannot specify. By 'pain's intrinsic badness' I intend pain's second-order property of being intrinsically bad, not pain's qualitative character. (The qualitative character is the first-order property of pain that *has* the property of being intrinsically bad. Pain is intrinsically bad *because of* its qualitative character.) The property which makes both localized and emotional pain a single kind of mental state, which unites qualitatively diverse pleasures into a single kind of mental state, and which makes pleasure and pain opposites, we could not specify. The exhaust and dispense doctrine is false. Premise 2 is true.

People may reply as follows: "Premise 2 is false; the exhaust and dispense principle is true. Pain's intrinsic badness and pleasure's intrinsic goodness are specifiable in a physical language and without the use of psychological words. We can specify pain's intrinsic badness by referring to a property in pain – in both localized pain sensations and non-localized emotional pain – of disposing a person to avoid the occurrence intrinsically, as an end in itself. (Analyzed so, pain's badness does not hinge on our reactions to pain but causes these reactions.) Similarly, we can specify pleasure's intrinsic goodness by referring to pleasure's property of disposing a person to seek pleasure as an end."

With these proposals people do not succeed in upholding the exhaust and dispense doctrine. The proposals do not enable us to specify pain's badness and pleasure's goodness with physical science language and without psychological words. The proposals contain 'person', 'avoid', and 'seek'. 'Person' is not a physical science word. People have minds. In principle, mindless automata might have a person's physical properties. They fall short of being people through lacking minds and so mental properties. 'Avoid' here denotes a human action and so is partly psychological. In avoiding Sarah's dog, people *intentionally* move in ways they *think* will lead them not to face her dog. Intentional pain avoidance is an act and so an event with a cognitive component. Hence, when we try to analyze it by reference to pain "avoidance," we do not depict pain's badness in physical science terms. A parallel attempt to portray pleasure's goodness in physical science terms fails for a similar reason.

Pain's badness and pleasure's goodness are not specifiable in physical science language. They are not locations or other spatial properties, nor are they temporal properties or any other orthodox material properties of neural events. Orthodox neural materialism is false.

Were pain's badness an orthodox material property, the effects pain has through being bad would differ from what they are. A physical object's spatiotemporal properties produce their effects in non-teleological, brutally mechanistic ways. They have effects without anyone being aware of those properties or the objects that have them. Pain's badness does not have its effects in a brutally mechanistic way. It has its impact teleologically – through grounding or justifying the cognitively saturated responses it prompts.<sup>4</sup> In being bad, pain prompts our avoidance of pain by grounding or justifying that avoidance and so by giving us reason to avoid pain. For an object's badness to have effects we must be *aware* of what is bad. Thus pain's badness is not an orthodox material property.

Claiming pain's badness is an orthodox material property, materialists may argue as follows. "Some identities – between water and H<sub>2</sub>O, for instance – are opaque and knowable only *a posteriori*." In the spirit of a new wave equation of goodness with some "natural" property, people might propose, "Pain's badness may be identical to some orthodox material property of a neural event. The identity may be knowable only *a posteriori*. Neuroscientists may discover it and thereby explain what pain's badness is." In Part III, I explain why no such discovery is possible.

### **III. There is No *A Posteriori* Identity between Pain's Badness and an Orthodox Material Property**

How should we formulate the *a posteriori* identity materialists entertain? *Some normative properties are identical to some non-normative properties?* Expressed so and interpreted straightforwardly, this formulation of the identity is self-contradictory. We discover *o* both has and lacks a particular property (normativeness). We need not gaily anticipate a discovery of this form.

---

<sup>4</sup> See Goldstein 2002.

Materialists might avoid self-contradiction in formulating the imagined discovery. For some *m* they might say, "Pain's intrinsic badness is identical to orthodox material property *m* of neural event *n*." For this proposal to be true: 1. It must represent an identity, and 2. There must be *a posteriori* identities – identities knowable only *a posteriori*. Neither condition is satisfied.

1. The standard proposals for mental-physical *a posteriori* identities are not viable candidates for identities.

Suppose pain's badness were identical to material property *m* of neural event *n*. If it were, we could not *explain* what pain's badness is, or analyze pain's badness, by reference to that material property. (We could no more do this than the reverse and explain what the material property is, or analyze it, by reference to pain's intrinsic badness.) Yet, people who propose *a posteriori* identifications of mental properties with neural properties regard the proposals as explanatory. They treat them as analyses. Materialists want to subsume pain's intrinsic badness and other intrinsic properties of mental states under the orthodox material. They expect thereby to explain what properties of mental states are by reference to the orthodox material. However, subsumption is not identity. In formulations of such alleged identities, proponents offer one-directional proposals, not reversible ones. (Neural materialists are *materialists*. They do not see themselves as claiming "Physical states are mental states." People who embrace the latter, reversed claim seem to endorse idealism, not materialism.) One-directional proposals are not identity statements.

2. There are no *a posteriori* identities.

Every object is identical to itself and only to itself. (No object is identical with an object having even one property it lacks.) This is knowable *a priori*. Thus, for any object it is knowable *a priori* what the object is identical with. In no case is what an object is identical with knowable only *a posteriori*. In no case is there a brute "explanatory gap" between something and what it is identical to.

If there are no *a posteriori* identities, there is no *a posteriori* identity between pain's badness and something. Hence, there is no *a posteriori*

identity between pain's badness and some orthodox material property of a neural event. (Hence, in no ("token") pain is there an *a posteriori* identity between pain's badness and some orthodox material property.) In every pain, what pain's badness is identical with – viz., itself – is knowable *a priori*.<sup>5</sup>

With this reasoning, aimed at proving there are no *a posteriori* identities, I speak *de re* not *de dicto*. I refer to objects and with what they are identical, and I speak of our ability to know with what something is identical. I do not directly refer to *sentences* said to report identities, or comment about the epistemic status of these sentences.

3. Some errors that facilitate belief in *a posteriori* identities.

One reason people believe in *a posteriori* identities is they conflate talk of 1. the epistemic status of identities with talk of 2. the epistemic status of *sentences* they assume represent identities. However, when I say this cat's self-identity is knowable *a priori*, I speak about the epistemological status of a particular fact – this cat's self-identity. I do not directly refer to some *sentence* about the cat or claim the truth of that sentence is knowable *a priori*.

To say every identity is knowable *a priori* is not to state or assume every statement that reports an identity is analytic. When rejecting *a posteriori* identities, I speak about objects and what they are identical to, and not about sentences about those objects.

In discussions of identity, people regularly assume identity concerns *two objects* identical to each other. (On these occasions people are not speaking *de dicto* and claiming, what is very different, that identity concerns two names for a single object.)<sup>6</sup> The assumption:

---

5 People who identify mental properties with neural properties offer *a posteriori* identities. Suppose there are no *a posteriori* identities. Then no psychological property – not a sensation's intrinsic qualitative character or any other property of a sensation – has an *a posteriori* identity with a neural property.

6 "If two things, *x* and *y*, are identical, then they have exactly the same properties," David Armstrong writes (1999, 23).

"For any objects *x* and *y*, if *x* is *y*, then it is necessary that *x* is *y*," Saul Kripke writes (1971, 137). When using the plural 'objects' to classify "*x* and *y*," Kripke regards *x* as one object and *y* another. When people assign Leibniz a belief in "the identity of indiscernibles," they use the plural term 'indiscernibles' to refer



facilitates a belief in *a posteriori* identities. Between *two* objects there are relations that are non-transparent and knowable only *a posteriori*. However, identity does not concern two objects. For every identity, there is *one* object identical to itself. That an object is identical to itself, and only to itself, is transparent and knowable *a priori*.

People believe in *a posteriori* identities primarily because they think they have examples of them. They assume certain kinds of *a posteriori* 'is'-sentences represent identities. Their analyses of these sentences and the discoveries people use the sentences to represent requires refinement. Here are four kinds of 'is'-sentences people present:

i. Cicero is Tully.

As people ordinarily intend the sentence, it is shorthand for 'The person named "Cicero" is also named "Tully".' (Somewhat similarly, in saying 'I am Irwin' I ordinarily announce my name and so say, in effect, 'My name is "Irwin"'.) The *a posteriori* knowledge I gain in learning the truth of 'Cicero is Tully' is linguistic – that the words 'Cicero' and 'Tully' name the same person. The linguistic knowledge I have in knowing that people use both 'Cicero' and 'Tully' to refer to a particular person is indeed attainable only *a posteriori*. (Indeed, the linguistic knowledge I have in knowing that people use only *one* name – 'Einstein' – to refer to my hamster also is attainable only *a posteriori*.) This *a posteriori* knowledge is distinguishable from the knowledge, attainable *a priori*, that a person is identical to himself and to no one else.

ii. The morning star is the evening star (Phosphorus is Hesperus).

What philosophers treat as a discovery of a single, simple fact (a fact they think they crisply identify with the single, short 'is'-sentence 'The morning star is the evening star') is analyzable into discoveries of distinguishable spatiotemporal and linguistic properties a particular

---

to the two objects. When people represent an identity as  $a = b$ , the presumption is that identity concerns two objects – that '*a*' represents one object and '*b*' a second. In this formula, '*a*' and '*b*' represent objects, not words for those objects.

heavenly body – Venus – has. Further knowledge of astronomical history might allow us to explain the advance in more detail.

When labelling a planet *the morning star*, the *Encyclopedia Americana* states, people deem it a “bright planet that becomes visible in the eastern sky shortly before sunrise.” When calling a planet *the evening star*, this work states, people identify it as a “bright planet appearing in the western sky, near the horizon, shortly after sunset.” There are five planets – Venus, Mercury, Mars, Jupiter, and Saturn – people label ‘the morning star.’ Each of the five, at times, satisfies the conditions for being labelled ‘the morning star’. Each planet at times also meets the conditions for being designated ‘the evening star’. For both morning and evening designations, Venus meets the conditions more often than do the other four planets. Venus is visible at the appropriate time and place more often than any of the other planets is.

With the words ‘The morning star is the evening star’ we do not crisply pick out a single fact. This ‘is’-sentence is elliptical, and, expressed without qualification and detail, is misleading in various ways in contexts in which philosophers use the sentence. The sentence does not accurately capture the astronomical advance that underlies the philosophers’ words. The following description of the advance is more accurate and informative. Scientists realized: “1. Any of five planets can be visible in the morning in the east, though Venus is visible at this time and place more often than the other four planets. 2. Though Venus is the planet most often visible in the evening in the west, any of five planets can be visible at that time and place. 3. More often than the other four planets, Venus has the (linguistic) properties of being labelled ‘the morning star’ and ‘the evening star’. 4. When people assign the morning star’s properties to one planet and the evening star’s to a second, they err or think crudely. Rather, both sets of properties belong to a single planet – Venus (given the most common referent of ‘the morning star’ and ‘the evening star’).” Formulating the astronomical advance with more precision and detail than philosophers usually give, though still somewhat crudely, we might say, “Astronomers discovered that the spatial, temporal, and linguistic properties people assign what they call ‘the morning star’, and those that people assign what they call ‘the evening star’, belong to a single object – Venus.” We might then say astronomers discovered these facts.

That a planet is identical to itself is knowable *a priori*. That cer-

tain planets have the properties of being visible from Earth at certain times of the day, and of not being visible at certain other times, and of appearing in various directions when visible, and of being identified with various names and expressions in human languages; is knowable only *a posteriori*.

People say astronomers discovered *Phosphorus* is identical to *Hesperus*. The example is akin to the 'The morning star is the evening star.' The *Phosphorus-Hesperus* discovery remains analyzable into spatiotemporal and linguistic components.

iii. Napoleon is the loser of Waterloo.

The *a posteriori* knowledge people have in knowing the truth of 'Napoleon is the loser of Waterloo' is that Napoleon has a particular property: having lost at Waterloo.

Suppose we convert the statement and say '*The loser of Waterloo is Napoleon*'. We say something different. We *identify* the loser of Waterloo, i.e., we state who lost at Waterloo.

iv. Water is H<sub>2</sub>O.

This idiom, as philosophers ordinarily intend it, is a formula for water's composition. It does not represent an identity and so does not represent an *a posteriori* identity. (Philosophers habitually report the molecular analysis as 'Water is H<sub>2</sub>O,' i.e., with a three-word set in which 'is' stands between 'water' and 'H<sub>2</sub>O'. This habit reinforces the belief that scientists use this idiom to present an identity.) The statement is not reversible: 'H<sub>2</sub>O is water' is not equivalent to 'Water is H<sub>2</sub>O'. The latter does not parallel 'Cicero is Tully'. 'Water' names a substance, 'H<sub>2</sub>O' (in this idiom) does not. To learn the truth of 'Water is H<sub>2</sub>O' is to learn water's composition — that hydrogen-oxygen molecules make up water.<sup>7</sup>

---

7 'H<sub>2</sub>O' has a secondary sense in which it is synonymous with 'water' and does not signify water's composition. ('H<sub>2</sub>O' has this sense in talk of 'a bottle of H<sub>2</sub>O'.) A person could intend 'H<sub>2</sub>O' in this sense in 'Water is H<sub>2</sub>O'. His sentence would then parallel 'The person called "Cicero" is also called "Tully"'. Understood so, the sentence would be reversible. However, in this sense 'Water is H<sub>2</sub>O' would not represent water's composition and so would not represent the scientific discovery philosophers intend with the expression 'Water is H<sub>2</sub>O'.

Many other statements said to represent *a posteriori* identities – about gold, clouds, lightning, heat – resemble ‘Water is H<sub>2</sub>O’ in being irreversible, composition statements. They represent analyses by reference to component physical parts.

I mentioned four kinds of alleged *a posteriori* identities. In each case the belief in the *a posteriori* identity is tied to *some* fact that is knowable only *a posteriori*. Each object referred to has *a posteriori* properties – properties it can be known to have only *a posteriori*. These include being named ‘Tully’, being visible in both the east in the morning and the west in the evening, having lost at Waterloo, having a particular molecular composition.

#### IV. Might Pain’s Badness Have A *Posteriori* Properties?

Might pain’s badness resemble water or lightning in having a particular atomic constitution that is knowable only *a posteriori*? Pain’s badness differs from water and lightning in this respect. Pain’s badness is not the sort of thing that has an atomic makeup. It is not a three-dimensional substance or a three-dimensional event that occupies some sector of space. Pain’s being bad is similar to a pain’s being rare, an argument’s being valid, and a soup can’s being to the left of the cat. None of these properties is a three-dimensional substance. None has a molecular composition scientists might discover.

People might think even if pain’s badness is not itself an orthodox material property, it at least has an *a posteriori* property of *correlating* with such a property, and so of being intimately related in some way with one. Pain’s badness might have the *a posteriori* property of *supervening* on an orthodox material property. These suggestions do not conflict with my thesis. In contending pain’s badness is not an orthodox material property I do not claim, what is quite different, that it does not supervene on, or otherwise correlate with, such a property.

#### V. Conclusion

Were orthodox neural materialism true, the exhaust and dispense principle would be true. In principle, we could specify properties of mental states in physical science language. Suppose Sheena hoped to specify all intrinsic properties of her feelings. Would it be possible in

principle for her to do so while dispensing with psychological words and solely by using physical science words to specify orthodox material properties of neural events in her brain? I think not. Many properties of mental states seem not to be specifiable in physical science language. In this paper, I have appealed only to one kind of property: the axiological. We cite this property in the analysis of pain and pleasure. Particular axiological properties of pain and pleasure, which I regard as defining properties, are not specifiable in physical science language. The exhaust and dispense principle is false. Orthodox neural materialism is false.

Might physical scientists expand their ontology and accept intrinsic goodness as a defining property of the physical? The development would not threaten my argument's conclusion. I conclude only that *orthodox* neural materialism is false, and hence that some properties of some mental states are not elements in the materialist's *current* ontology.<sup>8</sup>

## References

- Armstrong, D. 1999. *The Mind-Body Problem: An Opinionated Introduction*, Boulder: Westview Press.
- Goldstein, I. 1980. "Why People Prefer Pleasure to Pain." *Philosophy* 55: 349–62.
- . 1989. "Pleasure and Pain: Unconditional, Intrinsic Values." *Philosophy and Phenomenological Research* 50: 255–276.
- . 1994. "Identifying Mental States: A Celebrated Hypothesis Refuted." *Australasian Journal of Philosophy* 72 (1): 46–62.
- . 2000. "Intersubjective Properties by Which We Specify Pleasure, Pain, and Other Kinds of Mental States." *Philosophy* 75: 89–104.
- . 2002. "The Good's Magnetism and Ethical Realism." *Philosophical Studies* 108 (1–2): 1–14.
- Kripke, S. 1971. "Identity and Necessity." In *Identity and Individuation*, ed. Milton Munitz, 135–164. New York: New York University Press.

---

<sup>8</sup> Thanks to David Barnett, Errol Bedford, Sheena Goldstein, John Heil, Dorothea Lotter, Al Mele, David Ohreen, Ulrich Meyer, David Papineau, J. J. C. Smart, Brendan O'Sullivan, Stuart Rachels, and Joan Weiner for valued feedback. This paper is dedicated to a cherished friend: Errol Bedford.