THEORIA ET HISTORIA SCIENTIARUM, VOL. VIII, N° 2
Ed. Nicolaus Copernicus University 2008

Uriah Kriegel

The Reduction of Conscious Emotion

Abstract

The main purpose of this paper is to outline a possible reductive explanation of emotion in neurophysiological terms. But it will also be argued that such a reductive explanation is more difficult to achieve than is commonly thought, in that it has to address conscious emotional experience. It will be argued that when an emotion is conscious, what makes it the emotion it is, and an emotion at all, is its phenomenal character, and when an emotion is unconscious, what makes it the emotion it is, and an emotion at all, is the phenomenal character it *would* have if it *were* conscious. This has the consequence that the theory of emotion cannot be insulated from the theory of consciousness, and a reductive explanation of emotion must target the phenomenal character of conscious emotional experiences. A possible reductive explanation of this sort will be outlined.

Key Words: Emotion; Consciousness; Inter-theoretic reduction; Phenomenal character; Representational content; Higher-order monitoring

Introduction: Emotion, Consciousness, and Inter-Theoretic Reduction

The overarching purpose of this paper is to outline a way in which the realm of emotions *might* be reductively explained in neurophysiological terms. But I also want to argue that such a reductive explanation should be more difficult to achieve than is commonly thought, because it would require us to square off with the phenomenon of conscious emotional experience.

Since Freud, we have been accustomed to thinking of the realm of emotion as completely independent of the realm of consciousness. Freud showed us that each emotional state may occur either consciously or unconsciously. It seems to

follow that consciousness is inessential to the emotions. Thus, anxiety or anger can occur either consciously or unconsciously, and therefore the theory of emotion *as such* can abstract from the consciousness involved in some states of anxiety and anger. This means that the peculiar difficulties attending the project of reducing consciousness do not infest the project of reducing emotion. These difficulties can be dealt within the context of the theory of consciousness, while the theory of emotion, being insulated from the theory of consciousness, is prepared for inter-theoretic reduction into neurophysiology and ultimately chemistry and physics.

In the second part of the paper (§§7-10), I will argue that such insulation of the theory of emotion from the theory of consciousness is in fact impossible. Although Freud's demonstration that every emotional state can occur both consciously and unconsciously establishes a certain degree of independence of emotion from consciousness, there are forms of dependence that survive it. Indeed, the dependence of the realm of emotion on the realm of consciousness is so thorough and fundamental that the reduction of emotion cannot proceed without a reduction of conscious emotion.

In the third part of the paper (§§11-14), however, I will outline a way a reductive explanation of conscious emotion in neurophysiological terms might proceed. Drawing on previous work on consciousness (Kriegel, 2002a, 2003a, 2003b, 2003c) and conscious emotion (Kriegel, 2002b), I will attempt to sketch a specific way the reduction of conscious emotion might proceed.

Before doing so, let me discuss the concept of inter-theoretic reduction in the sciences and tackle some of the foundational, a priori difficulties that it presents. This is the topic of the first part of the paper (§§2-6). Recent work in philosophy of science and especially philosophy of psychology suggests that a reduction of psychology into neurophysiology may be impossible. I will argue that this prospect forces on us a heterodox model of inter-theoretic reduction in the sciences.

I. Reduction and Reductionism

1. From Non-Reductive Physicalism to Eliminativism

Inter-theoretic reduction is a relation between scientific theories. But a successful reductive enterprise involves not only the reduction of a *theory,* but also of what the theory is *about.* Thus, the reduction of zoology to organic chemistry would entail not only that the theory of animals was reduced to another theory, but also that animals themselves were reduced to another kind of entity. Reduction of entities, as opposed to theories, is what is often called *ontological reduction.*

Inter-theoretical reduction entails ontological reduction, in that the successful reduction of a theory requires the successful reduction of the entities it is a theory of. So inter-theoretic reduction of the theory of emotion to neurophysiology requires ontological reduction of emotional phenomena to neurophysiological phenomena.

Ontological reduction must be distinguished from ontological *elimination.* Zebras will be chemically reduced when all zebra phenomena are explained in chemical terms. But reductive explanation of zebra phenomena must be such as to allow us to conceive of the zebra as essentially a complex chemical entity. This is different from the fate that befalls such alleged entities as witches and ghosts. We expect a scientific explanation of all witchcraft phenomena, but not one that will allow us to see witches as essentially another kind of scientifically posited entity, but rather one that will allow us to see witches as non-existent. That is, the scientific explanation of witchcraft phenomena *eliminates* witches, whereas the scientific explanation of zebra phenomena *reduces* zebras. Only the latter explanation is reductive.

Likewise, in seeking a reductive explanation of emotional phenomena in neurophysiological terms, we expect to reduce emotions to neurophysiological entities, not to eliminate emotions. However, recent work in Philosophy of Psychology suggests that central mental phenomena may eventually face elimination rather than reduction. This would be in line with an eliminativist tradition according to which the mind does not really exist - not any more than witches and ghosts do (Churchland, 1984). In this and the next section, I discuss this recent literature. In later sections, I will attempt to "save" mental phenomena from their looming elimination.

Reductive physicalism is the thesis that mental properties, and in fact all the properties invoked in the special sciences (zoology, geography, sociology, etc.), are reducible to physical properties, that is, properties invoked in physics. Around the middle of the twentieth century, reductive physicalism was deemed so obvious as to require no positive argumentation. Thus, when Smart (1959) set out to defend reductive physicalism, his strategy was to claim that no argument *against* it could be made to work. The only *positive* consideration in favor of reductivism he saw a need to adduce was the following (Smart, 1959: 169):

> It seems to me that science is increasingly giving us a viewpoint whereby organisms are able to be seen as physicochemical mechanisms: it seems that even the behavior of man himself will one day be explicable in mechanistic terms.

This observation is still today the chief motivation behind reductivism. But today most philosophers are of the opinion that reductive physicalism is false and many properties invoked in the special sciences are irreducible to physical properties.

The cause of this shift is a single argument, namely, Putnam's (1967/1991) *argument from multiple realizability*. Putnam noted that mental properties are multiply realizable. In humans, pain is realized in one sort of physical event, say, C-fiber firing, but in other creatures it may be very different. Suppose there are extra-terrestrial creatures, Venusians perhaps, who feel pain just as we do, but who have no neural system. Instead, the physical event underlying pain in Venusians is some event taking place in their silicon brains. The mere possibility of such a scenario, in which pain is realized in silicon, precludes the identification of pain with C-fiber firing. For there is no more reason to identify pain with C- -fiber firing than with the relevant silicon event in the Venusian brain. And yet pain cannot be identified with *both* C-fiber firing and the silicon event, since this would entail (by the transitivity of identity) that C-fiber firing is identical with the silicon event, which it evidently is not[1]. Therefore, the property of being in pain is irreducible to any physical property. The same point was extended by Fodor (1974) to all properties invoked in the special sciences, e.g., the property of being a mountain. Fodor noted that two chunks of land can be a mountain in virtue of such different microphysical properties that the property of being a mountain cannot be identified with any one microphysical property.

The argument from multiple realizability led to a new anti-reductivist orthodoxy in Philosophy of Psychology and more generally Philosophy of Science. However, recent work by Jaegwon Kim (1989a, 1989b, 1998) has mounted a serious challenge to non-reductive physicalism. Kim argues that if mental (and other) properties are irreducible to physical properties, they are likely to be causally inert, and therefore explanatorily useless and theoretically dispensable. If so, non-reductive physicalism may ultimately lead to an *eliminative* physicalism according to which mental properties (and other properties invoked in the special sciences) do not really exist[2].

Kim's "Master Argument" can be set out with two basic premises. The first is the principle of the Causal Closure of the Physical. This is the thesis that all physical events have complete and sufficient physical causes. Formulated in terms of properties rather than events, the thesis is that every instantiation of a physical property has as its (complete and sufficient) cause the instantiation of another

[1] The assumption here is that identities are absolute. If one embraces relative identity (Geach, 1967), one can claim that pain is identical with C-fiber firing *relative to hitmans* but identical with the relevant silicon event *relative to Venusians*. This reductive strategy consists in effectively denying the transitivity of identity. Most philosophers, however, would be disinclined to embrace relative identity.

[2] There is a question as to what it means for a property to exist or fail to exist. Perhaps a more cautious formulation of eliminative physicalism is as the claim that mental (and other) properties are never instantiated in the actual world.

physical property[3]. The second premise is the principle of Causal Exclusion. This is the thesis that events cannot in general have more than one sufficient cause. It may happen occasionally that an event has two separate causes, each of which would be sufficient by itself to bring about the event in question. But this is bound to be a most infrequent occurrence, and cannot possibly be a pervasive feature of our world. Normally, an event has a single sufficient cause. Formulated in terms of properties, this premise states that the instantiation of a property normally has only one property instantiation as its cause. Property instantiations cannot generally have two separate and sufficient causes.

With these two premises, we can lay out Kim's Master Argument. Suppose a subject *S* experiences pain and winces in response. We want to say that S's pain experience was the cause of her wincing - that the pain caused the wincing. However, non-reductive physicalism excludes such a causal connection. In accordance with the Principle of Causal Closure, the wincing must have a (sufficient) physical cause. There must be a neurophysiological event in S's brain that (by itself) causes S's wincing. If S's pain also caused the wincing, then the wincing had *two* sufficient causes: S's pain and the neurophysiological event in S's brain. But according to the Principle of Causal Exclusion, the wincing cannot have two *separate* sufficient causes[4]. This means either (i) that the pain is not really a *separate* cause of the wincing, or (ii) that the pain is not a cause of the wincing at all. According to the first option, the two causes are not really separate but are actually one and the same, meaning that the pain just is the neurophysiological event in S's brain. This is in fact reductive physicalism. According to the second option, S's pain is causally inert and has no role in the production of S's wincing. If the non-reductivist holds on to the irreducibility of pain to the neurophysiological events, she must reject the first option and embrace the second.

Thus the non-reductive physicalist is forced into the position that mental events and properties are normally epiphenomenal, that is, causally inert. Once mental properties are construed as causally inert, it would seem that any explanatory role

[3] The two formulations would come to the same under the conception of events as property instantiations - which, incidentally, Kim (1976) defends. If we reject this conception of events, the two formulations would probably constitute two different theses. But it would still be unlikely that one of the two were true while the other were false, though see Raymont (2003) for a claim that they would.

[4] It is possible to claim here that this case may be one of the abnormal cases in which an event does have two separate sufficient causes. But this move would just postpone the non-reductivist's problem. For the same reasoning applied above to S's wincing can be applied to any motor response to which we would like to assign a mental cause. And according to the Principle of Causal Exclusion, it is impossible that all such motor events are abnormal in this way.

they may have in psychology must be strictly illusory. Since they have no causal role in bringing about the behavior of agents, their occurrence would not in the least explain why these behaviors occurred[5]. Therefore, inasmuch as psychology attempts to explain events whose explanation appears to be mental or psychological, non-reductive physicalism would take psychology out of business.

According to Kim, it is but a short step from epiphenomenalism to eliminativism, and this step is what he calls Alexander's Dictum: to be is to be causally efficacious. If causal efficacy is a necessary condition for a property's existence, then if mental properties are causally inert, they do not exist. Thus non-reductive physicalism leads directly to eliminative physicalism.

2. From Reductive Physicalism to Eliminativism

Kim's own inclination is to reject the irreducibility of mental properties to physical properties. If mental properties just are physical properties, as the reductivist maintains, then there is no competition between the mental causes and physical causes of agents' behavior. Thus, S's wincing can be caused both by S's pain and by the relevant neurophysiological event if the two are not *separate* causes but *one and the same* cause.

The problem is how to reconcile such reductivism with the multiple realizability of mental properties. Kim explores two possible avenues: *disjunctive reduction* and *local reduction.*

The former is the notion that, although being in pain cannot be reduced to the property of undergoing C-fiber firing, it may be reducible to the *disjunctive* property of undergoing C-fiber-firing-or-the-relevant-silicon-event. A full disjunction of all the *possible* realizations of pain will be a suitable reducer of pain. Kim (1992, 1998) argues against such reduction on the grounds that disjunctive properties are scientifically useless because they are not projectible. For us to formulate scientific generalizations, such as "All gold is yellowish," we must invoke projectible properties, such as gold. Gold is projectible in that the discovery of a million instances of yellowish gold does suggest that the next instance of gold will be yellowish. This feature, projectibility, is missing from disjunctive properties, such as being a dog or a fish. The fact that a million dogs-or-fish have turned out to be warm-blooded does *not* suggest that the next instance of a dog-or-fish will be warm-blooded. For if it so happens that the million instances are all dogs, whereas the next instance is a fish, the warm-

[5] This would be especially so on a causal account of explanation (Lewis 1993), which is very plausible in general but particularly plausible in the context of psychological explanation.

bloodedness of the examined instances will nowise bear on the question of whether the next instance will also be warm-blooded. Therefore, there are no scientific generalizations to be sought about dogs-or-fish. Likewise, there should be no scientific generalizations to be sought in terms of C-fiber-firing-or-the- relevant-silicon-event.

Kim prefers the second sort of reduction, local reduction. Local reduction means that (say) pain is reduced to physical structures *relative to* kinds of organism or system. So, pain is reduced to C-fiber firing relative to humans and to the relevant silicon event relative to Venusians. The reduction is local to humans, Venusians, or other creatures, rather than being global to all of them. Another way to put this is to claim that, although the property of being in pain cannot be physically reduced, the properties of being in *human* pain and of being in *Venusian* pain can. For local reductionism, what is up for reduction in the sciences is not pain *simpliciter*, but such things as human pain. And while pain *simpliciter* may be multiply realizable, there is no reason to think that *human pain* is multiply realizable as well[6].

The problem with local reductionism is the price it asks us to pay, which is to give up on pain *simpliciter*. According to Kim, the multiple realizability of pain *simpliciter* means that it cannot be reduced to a physical property. But its irreducibility entails that it is causally inert and hence theoretically dispensable. That is, local reductionism also leads to eliminativism, although on a smaller scale than non-reductive physicalism. For although it preserves and reduces such properties as being in human pain, it eliminates the property of being in pain. The same reasoning applies to other mental properties. Thus, the property of believing that it is raining would be eliminated from Kim's ontology, although the property of human-believing that it is raining is reduced to a physical property.

Kim's elimination of such "simpliciter" mental properties is implausible. After all, it is not for nothing that human pain and Venusian pain are both pain. They share many features, phenomenological as well as functional. *Ex hypothesi*, what it feels like for a Venusian to be in her pain is the same as what it feels like for us to be in our pain; and the functional role of pain in the Venusian's mental life is the same as the functional role of our pain in our mental life. On the most natural conception of properties, this sameness is already ground enough to admit a property shared by the human and the Venusians (see Armstrong, 1978). This property is surely the property of being in pain. Yet Kim claims that there is no such property.

The upshot is that both reductive and non-reductive physicalism ultimately transform into eliminative physicalism about straightforward mental properties

[6] In fact, many philosophers have argued that even human pain is multiply realizable (including Putnam, 1967/1991). This issue will be discussed in the next section..

("simpliciter" properties). More accurately, eliminative physicalism is entailed by the theses of multiple realizability, causal exclusion, and the causal closure of the physical. All three theses are hard to deny, but their acceptance leads straight to eliminativism. If we are to avoid eliminativism, one of these three theses must go. There have been several attempts in the literature to undermine the force of the theses of causal exclusion and causal closure, attempts which - for reasons I will not discuss here - appear to fall short of their task. Here I would like to explore the possibility of denying the multiple realizability of mental properties.

3. Multiple Realization and Multiple Realizability

It may initially appear completely far-fetched to deny that mental properties are multiply realizable. After all, it is plain that Venusians could feel pain, even if they did not have a nervous system at all. However, it appears just as far-fetched to deny the causal exclusion principle, the causal closure principle, and the existence of mental properties. The problem we face is that these four propositions cannot be held conjointly, even though all seem undeniable. We must therefore deny what initially appears undeniable, and reexamining multiple realizability is as good a place as any to start.

As a first order of business, let us distinguish between multiple realization and multiple *realizability.* The thesis of multiple realization is stronger than the thesis of multiple realizability, in that it claims that the existence of pain which is not realized in C-fiber firing is not only possible, but actual. According to the thesis of multiple realizability, there *could* be creatures which would feel pain even when their C-fiber did not fire. According to the thesis of multiple realization, there *are* such creatures. Putnam himself cited the octopus as a creature which likely experiences pain but in whom the physical realization of pain is unlikely to be similar to its realization in humans. Indeed, it has even been suggested that different human individuals may be in different physical states when they undergo a pain experience (Horgan, 1997).

Many philosophers have accepted Putnam's claim regarding the multiple realization of pain, but it is surely more likely to turn out false than any of the theses we have discussed thus far[7]. I suspect that, in acquiescing with the notion of multiple realization, philosophers have had in mind too restrictive a view of what a legitimate reducer might be. It is quite unlikely that when the octopus is in pain, the very same cell assemblies in its brain are excited as when a human is in pain; even pain in two human individuals is likely to be realized in different

[7] A similar point is made by Kim (1998: 94-5).

neural cell assemblies. But a reduction of pain does not require such a coarse, neuro-anatomical reducer. As a dynamic, complex, self-organizing system, the brain has many neurophysiological properties that are much more complex and more subtle than its neuroanatomical properties, and among those there is likely to be found a property shared by the octopus brain and the human brain when both are in the same state of pain, and certainly by two human individuals who feel a similar pain[8]. In any event, this is what scientists assume when they seek the "neural correlates" of pain; they assume, that is, that pain *has* a neural correlate. If so, pain is *uniquely realized,* whether or not it is uniquely realizable.

In what follows, I will assume that mental properties are indeed uniquely realized. In the next two sections, I will explore lines of argument that take us from the thesis of unique realization to the reducibility of mental properties. Neither line of argument will be without problems or grounds for dissatisfaction. But as I noted at the beginning of this section, this seems to be a feature of all possible positions on the issue at hand. Our aim is not so much to find a fully satisfactory position as to find the least unsatisfactory position.

4. Rejecting Multiple Realizability

One line of argument would be that the unique *realization* of pain entails its unique *realizability,* because it provides sufficient grounds for *identifying* pain with its unique realizer. That is, because in the actual world pain is uniquely realized by C-fiber firing, we are justified in identifying pain and C-fiber firing. And once they are identified, anything that is not C-fiber firing must be admitted to be *different* from pain. Thus, the silicon event in the Venusian brain, although correlative with an experience that feels the same as pain, is not pain.

This is, after all, what we do with water. Because water is uniquely realized by H,O, we *identify'* water with H_2O. Any substance we may find on Venus that is not H_2O is therefore *not* water. Even if the substance is clear, liquid, and quenches thirst, we would claim that it is not water, but only superficially similar to water. If the substance is XYZ rather than H_2O, we might call it twin-water, but we would not call it water. Water is H,O, and that substance is not H,O. So that substance is not water. Similarly, pain is C-fiber firing, and the silicon event in the Venusian brain is not C-fiber firing. So that silicon event is not pain.

[8] It is of course likely that the octopus experiences its pain as much more dull than we do our pain. It is then likely that there would be a neurophysiological property shared by a normal octopus pain and a dull human pain. More generally, pains that are phenomenologically alike probably share a neurophysiological realizer - within the actual world, that is.

By contrast, anything we might find on Venus that looks like jade we would happily consider to be jade. This is because jade is multiply realized in the actual world[9]. Because jade is realized by both $NaAlSi_2O_6$ and $Ca_2(Mg, Fe)_5Si_8O_{22}(OH)_2$, it cannot be identified with either of them. It is for this reason - because it cannot be identified with any one substance in the actual world - that jade is multiply realizable. If it was identical with some actual substance, jade would not be realizable in anything other than that substance. But if jade was uniquely realized in the actual world, we would have no reason not to identify it with its unique realizer.

On the view here being explored, multiple realizability *depends upon* multiple realization, and unique realization *entails* unique realizability. What motivates this view is the fact that theoretical identifications, as performed in the sciences, have nothing to go on except what happens in the actual world. Science is not in the business of investigating possible worlds. When a property turns out to be uniquely realized in the actual world, there is nothing else science need (or can) establish before it identifies the property in question with its realizer. This is how the reduction of water proceeded, and why the reduction of jade was undercut. And this is the way a reduction of pain will proceed or be undercut. Under the assumption, hesitantly put forth in the previous section, that pain is indeed uniquely realized in the actual world, the reduction of pain should proceed without a priori difficulties. And similarly for other mental properties.

An objection to this line of argument is based on Kripke's (1980) claim that when it comes to pain and other conscious phenomena, there is no gap between appearance and reality. Thus, when we encounter the XYZ substance on Venus, we have no difficulty declaring that, even though the substance *looks* (appears) like water, it *is not* water. This is because we have no difficulty distinguishing the way water *is* from the way it *looks* (appears). However, when we encounter the Venutian silicon event, we would be uncomfortable declaring that, even though the event *feels like* pain, it *is not* pain. This is because we are disinclined to distinguish the way pain *is* from the way it *feels.* On the contrary, we are inclined to say that to be in pain is to feel pain. There is nothing more to being in pain than feeling pain.

Some philosophers have denied Kripke's claim, insisting that some mental events feel like pain but are not pain[10]. I do not wish to pursue this rejoinder

---

[9] Throughout this discussion, I am assuming that Venus does *not* in fact include creatures with silicon brains and experiences phenomenologically similar to us and substances that are superficially similar to water and jade. So these are all counterfactual states of affairs that reflect only on the *possible* realizers of pain, water, and jade - not on their actual realizers.

[10] Suppose a blindfolded person is told that she is about to be cut in the back of her neck. When an ice cube is then placed on the back of her neck, this person might experience the cold sensation, in the first split second, as pain. It has been argued that in this case a mental state which feels like

here, however, as it strikes me as clearly misguided. Kripke is right that there is a disanalogy between the case of water and the case of pain, inasmuch as the phenomenological properties of pain have just as good a claim to be the *essential* properties of pain - the properties that make pain what it is - as the physical- realization properties of pain, whereas the superficial properties of water do not have nearly as good a claim to be the essential properties of water as do the physical-realization properties of water. To avoid this objection, let us explore a different line of argument from unique realization to reduction.

5. Reduction, Identification, and Constitution

If pain necessarily is the way it feels, then the fact that Venutians feel pain means that what they experience *is* pain. Yet their experience is not realized by C-fiber firing. Therefore, pain is not identical with C-fiber firing. But this entails that pain is not *reducible to* C-fiber firing only if identity is a necessary condition for reduction, that is, only if it is necessary that the reduced property be identical with the reducing property. This assumption was presupposed throughout the above discussion, but in light of the fact that there seems to be no way to escape eliminativism as long as we hold on to it, we may wish to reconsider it.

To be sure, there must be an intimate relation between reduced and reducer. If the relation in question falls short of identity, it must still be a relation close enough to identity. In any case, it must be a relation that would warrant a "nothing but" judgement. That is, the relation between the reduced property and the reducer property must allow us to see the former as *nothing but* the latter. This seems to be the key to the reduction relation. And although it is clear that the identity relation does warrant nothing-but judgements, it is not clear that other relations do not.

My suggestion is that we require a *constitution relation* between reduced and reducer[*] [11]. On this suggestion, whenever a property *F* is constituted by a property *G*, *F* may be reduced to *G*. The idea is that even if unique realization of *F* by *G* does not entail the identity of *F* and *G*, it may yet entail that *F* is constituted by, and hence reducible to, *G*. Crucially, the constitution relation seems to warrant nothing-but judgements, as we will see momentarily.

The constitution relation is familiar mostly from the realm of individual objects. A statue and the clay from which it is made are not identical to each

---

pain is not pain. To my mind, this is mistaken, and the sensation is correctly regarded, in the first split second, as real pain - precisely because it feels like pain.

[11] It is part of this suggestion that the constitution relation is weaker than the identity relation. For argumentation to that effect, see Johnston, 1992.

other, since they have different modal properties (e.g., the statue could have been made of bronze, whereas the clay could not have been made of bronze). But even though they are not identical with each other, they *coincide* with each other, and this means that there is an intimate relation between them. This is the constitution relation: the statue is *constituted by* the clay. Moreover, because the statue is constituted by the clay, there is a sense in which it is "nothing but" the clay. There is nothing to the statue that goes beyond the clay. This makes it plausible to claim that the statue is *reducible* to the clay.

The constitution relation holds also among events. Suppose Smith makes a trip to Warsaw. Clearly, Smith's trip to Warsaw is not identical with Smith's trip to Poland, if only because their modal properties are different (e.g., Smith's trip to Poland could have involved a visit to Krakow, but her trip to Warsaw could not have involved a visit to Krakow (although it could have been *conjoined with* a visit to Krakow)). At the same time, it may well be that Smith's trip to Poland is *constituted by* her trip to Warsaw, namely, in case her trip to Poland involves a trip to nowhere else but Warsaw. In such a case, we would be well justified to say that Smith's trip to Poland was *nothing but* her trip to Warsaw, and hence that the former event is reducible to the latter.

If the constitution relation can hold among individual objects and among events, then we should expect it to hold among properties. Indeed, Johnston (1997) has argued that water is not identical to H_2O, but is rather constituted by it. Johnston's argument is that if water is identical to H_2O, then vapor (as well as ice) is also identical to H_2O, but this is impossible, since water is not identical with vapor (nor ice). Johnston concludes that the relation between water and H_2O (and between vapor and H_2O and ice and H_2O) is that of constitution rather than identity. Now, the discovery that water is constituted by H_2O is a paradigmatic scientific reduction. This suggests that reduction indeed requires no more (nor less) than a constitution relation between reduced and reducer[12].

Now, it seems to me that the reason H,O constitutes water, even though it is not identical therewith, is that it is water's unique realizer. In any case, it is clear that water really is *nothing but* H_2O, and is therefore reducible to it. A similar model of reduction could apply to mental properties: pain is nothing but C-fiber firing, and is therefore reducible to it, because pain is constituted by C-fiber firing. And pain is constituted by C-fiber firing because the latter is its unique realizer[13].

Observe that the multiple realizability of mental properties is not denied here, but is rather neutralized. On the view I am recommending, the multiple

[12] This does not exclude identity as a source of reduction. To say that reduction does not *require* anything more than constitution is not to say that it does not *allow* anything more than constitution.

[13] By contrast, because jade is multiply realized, it is false that jade is nothing but $NaAlSi_2O_{fi}$ or nothing but $Ca_2(Mg, Fe)_5\ Si_{\#}O_{22}(OH)_2$.

realizability of mental properties does not undercut their reduction to physical properties, because multiple realizability is fully compatible with the constitution of a mental property by a physical property. At the same time, Kim's Master Argument does not threaten this sort of view, because the Causal Exclusion Principle only claims that there cannot normally be two *separate* causes of the same event, whereas mental properties, being constituted by their uniquely realizing physical properties, are not separate from them, since they are "nothing but" these physical properties.

In summary, the only way to avoid an eliminativist account of pain *simpliciter* appears to be to employ a constitution model of reduction that is based on unique realization as the relation underlying reduction[14]. This requires that we deny the multiple realization of mental properties, though not their multiple *realizability.* To be sure, the constitution model of reduction is unorthodox, but the elimination of mental properties, which appears inevitable without it, would be even more heterodox.

In the remainder of this paper, I will assume a constitution model of reduction. This means that in order to effect a successful neurophysiological reduction of the emotions, we must find neurophysiological properties that uniquely realize each emotional property. We will have occasion to revert to this issue in the third part of the paper.

II. The Dependence of Emotions upon Consciousness

6. Emotion and Consciousness After Freud

As is well known, Freud completely revolutionized our way of thinking about the emotions. Before Freud, emotions were commonly thought of as essentially conscious. On this conception of emotion - which, for lack of a better term, we may call *Cartesian* - to be jealous, angry, or anxious is to undergo a certain conscious experience characterized by a special inner feeling associated with it. This "inner feeling" is the *phenomenal character* of the emotional experience. What makes a particular mental state a state of anxiety, rather than jealousy, is

[14] One advantage of the constitution relation is that, tike the reducibility reduction, it is asymmetric. If what makes water reducible to H,O is that water is identical to it, this would presumably mean that H,0 is also reducible to water, which is strange to say the least. By contrast, there is a sense in which water does not constitute H_2O even though H,0 constitutes water. This helps explain why water reduces to 11,0 but not the other way around. (It may well be that some other explanation is available, but the point for now is that the constitution model of reduction has its own resources to explain, rather straightforwardly, the asymmetry of reduction.)

that it feels to the subject the way anxiety feels, not the way jealousy feels. On this conception of emotion, then, emotions are conceived of as not much more than raw feelings.

Through his meticulous study of unconscious emotions involved in such psychological phenomena as repression and denial, and his parallel theoretical work on the concept of the unconscious, Freud taught us to think of emotion in dissociation from consciousness (Freud, 1915). On the Freudian conception of emotion, a great deal of our emotional life goes on outside the sphere of consciousness. This means that many genuine emotions have no phenomenal character, and therefore involve no inner feeling. This conception of emotions inevitably led to a broader notion of emotion as involving essentially certain cognitive and motivational components. Rather than being little more than raw feeling, emotion is here construed as continuous with full-fledged cognition.

This broader notion of emotion, premised on the Freudian dissociation of emotion from consciousness, has proved immensely useful, both experimentally and clinically. But my contention is that it has also led to an overreaction to the Cartesian conception of emotion[15]. In particular, I want to argue that the realm of emotion is conceptually dependent upon the realm of consciousness in a very fundamental way. In this part of the paper, I will argue that the existence of unconscious emotions does not establish the complete independence of emotion from consciousness, and that indeed there are good reasons to think that emotion is essentially connected to consciousness, in that what makes an unconscious emotion the emotion it is, and an emotion at all, is the way it *would* feel if it *were* conscious. This entails that there can be no complete understanding of emotion in dissociation from an understanding of consciousness. That is, the theory of emotion would be incomplete without an account of conscious emotion and hence emotional consciousness. Accordingly, an inter-theoretical reduction of emotion would require a reduction of emotional consciousness.

7. Is Emotion Completely Independent from Consciousness?

At a most basic level, a theoretical framework for emotion must answer the following two questions:

1. What makes a given emotion the emotion it is (rather than another emotion)?
2. What makes a given emotion an emotion at all (rather than a different kind of mental state)?

[15] Consider a straightforward example of such overreaction. It is customary to hear psychologists today claiming that emotion and feeling are "two different things." But the fact that conscious emotion involves more than just a feeling does not in any way entail that it does not *also* involve feeling (Perkins, 1966).

Consider a particular unconscious emotional state, such as anger at one's mother. What makes this unconscious state a state of anger, rather than, say, a state of shame or disappointment? And what makes it an emotional state at all, rather than a state of memory or perception?

These questions receive a straightforward answer within the Cartesian framework.

On the Cartesian conception of emotion, what makes a conscious emotional experience of anger a state of anger, and not a state of disappointment, is that it *feels* the way anger feels, not the way disappointment feels. The feeling of anger and the feeling of disappointment are very different - anger involves expansion of the soul, if you will, whereas disappointment involves shrinking of the soul - and the conscious experience under discussion involves the former rather than the latter[16]. Moreover, what makes this conscious experience an emotional state at all, rather than a perceptual state, is that it feels the way emotions feel, not the way perceptual experiences feel. The phenomenal character of being angry is very different from the phenomenal character of seeing yellow, and the conscious experience under discussion has the former phenomenal character and not the latter.

The Cartesian framework answers the above questions by explicit appeal to the phenomenal character, or inner feeling, of conscious emotional experiences. But this sort of answer cannot be given from our Freudian perspective, which countenances the existence of *unconscious* emotions involving no feeling. Within the Freudian framework, the question arises of what makes an unconscious emotion the emotion it is, and an emotion at all. Given that unconscious emotions have no phenomenal character or inner feeling, it cannot be that what makes an unconscious emotion the emotion it is, and an emotion at all, is how the emotion feels.

This may lead us to think that consciousness has no role in answering those basic questions. However, the existence of unconscious emotions does not rule out a subtler role for consciousness in answering those questions. In particular, we can still hold the following thesis: what makes an unconscious anger a state of anger, rather than disappointment, is that *if* it were brought up to consciousness it *would* feel the way conscious anger feels, and not the way conscious disappointment feels; and what makes it an emotional state at all, rather than a perceptual state, is that *if* it were conscious it *would* feel the way a conscious emotional experience feels, not the way a conscious perceptual experience feels.

These answers to the above questions are not ruled out by the existence of unconscious emotional states, yet they portray emotions as intimately dependent

[16] Obviously, I mean the descriptions "expansion of the soul" and "shrinking of the soul" in a figurative way only. In particular, I do not wish to commit to the existence of a soul in the Cartesian sense - an immutable, simple, and indestructible entity.

upon consciousness. On this view, the essential property of an unconscious emotion is that it *would* feel a certain way if it *were* conscious. We may call this sort of property a *counterfactual phenomenal property*. The view under consideration is that the counterfactual phenomenal properties of unconscious emotions play the same role in making them the emotions they are (and emotions at all) as the *actual* phenomenal properties of conscious emotions play in making *them* the emotions they are (and emotions at all).

This view of emotions theorizes them as importantly dependent on consciousness. According to it, there is nothing that makes a particular unconscious state an emotion, let alone the particular kind of emotion it is, other than the phenomenal character it would have if it were conscious. This means that the existence of the phenomenon of consciousness is a necessary condition for the existence of emotions. In other words, there would be no emotions in a world without consciousness. For in a world without consciousness, there would be nothing to make a given unconscious state an emotional state rather than a mental state of some other kind.

The view of emotion I have portrayed is quite intuitive. When a therapist assesses that a patient harbors unconscious anger at her mother, what makes the therapist think of the patient's unconscious state as a state of anger is that, if the therapy goes well and the patient "gets in touch" with her emotion in such a way that the latter becomes consciously felt, the way it will feel to the patient is the way anger feels. The objective of the therapy is precisely to make the patient *feel* the anger she now merely *harbors.*

The view of emotion under consideration carries intuitive conviction, then. The question is whether it can withstand critical scrutiny. The main tenet of this view is the notion that there is nothing about an unconscious emotion that makes it the emotion it is (and an emotion at all) other than its counterfactual phenomenal properties. In the next two sections, I will consider two other candidates for making unconscious emotions the emotions they are (and emotions at all): their *functional* properties and their *representational* properties (respectively). Both candidates will be shown to be insufficient to undercut the dependence of emotion on consciousness. This will consolidate the case for this fundamental sort of conceptual dependence of emotion upon consciousness.

8. Emotion, Consciousness, and Functional Role

One way to answer our two theoretical questions about emotions is by invoking the functional role of emotion in the mental life of the subject. An emotional state typically has a specific set of causes and effects, which set constitutes its functional role. It could thus be claimed that what makes an unconscious emotion the emotion

it is (and an emotion at all) is that it has the functional role of the relevant kind of emotion. That is, what makes an unconscious anger a state of anger rather than disappointment is that its functional role is like that of a conscious anger and unlike that of a conscious disappointment. On this view, the essential property of an emotion is its functional role (DeLancey, 2001; see also Rey, 1980).

The problem, however, is that the causes and effects of conscious and unconscious anger are very different. Indeed, there seem to be principled functional differences at play here, both on the side of causes and on the side of effects.

On the side of causes, it is clear that the causal circumstances behind the formation of a conscious anger must be different from those behind the formation of unconscious anger. For something must cause the anger to be conscious. What causes the anger to be conscious is an element in the functional role of a conscious anger that is necessarily missing from the functional role of an unconscious anger (if it was part of the functional role of the latter it would cause it to be conscious, contrary to assumption). Conversely, the mechanisms that drive the suppression of certain emotions, in a way that keeps them unconscious, are evidently moot when an emotion does become conscious (if these mechanisms were active, the emotion would not become conscious, contrary to assumption)[17].

On the side of effects, there are good reasons to think that conscious and unconscious anger do not have the same effects. To suppose that they do is to suppose that consciousness contributes nothing to the fund of causal powers of a conscious emotion, that is, that consciousness is epiphenomenal[18]. Moreover, the fact that conscious anger has different causal powers than unconscious anger is the *raison d'etre* of psychotherapy. It is widely assumed that emotions which remain repressed are bound to manifest themselves in neurotic or psychotic symptoms. This is to assume that unconscious emotions have the power to produce behavioral effects that conscious emotions would not produce. Thus, if anxiety remains unconscious ("unprocessed") overlong, it might erupt in panic attacks, but conscious anxiety will not.

In conclusion, the functional roles of conscious and unconscious anger appear to be very different, on both sides of the functional characterization. Therefore, it cannot be that an unconscious anger is the emotion it is (anger, rather than disappointment) in virtue of having the sort of functional role a conscious anger (and not a conscious disappointment) has.

[17] Or if they are not moot they are at least not *as* active (i.e., not active to the same degree) as they are when their activity does result in the suppression of an emotion.

[18] It is not incoherent, of course, to maintain that consciousness does not contribute *anything* to a mental state's fund of causal powers - that consciousness is causally inert, or epiphenomenal (Velmans, 1992). But as we saw earlier, this leads to an unappealing eliminative account of consciousness.

9. Emotion, Consciousness, and Intentionality

A more plausible alternative is the suggestion that the essential properties of emotions are intentional, or representational. Emotions are intentional, in that they are typically *about* something. It is impossible to be angry without there being something one is angry about. Anger thus has an object, and a state of anger must employ a representation of that object. It could therefore be suggested that what makes an unconscious anger at one's mother a state of anger, rather than disappointment, is that the representational properties of the unconscious anger are similar to those of conscious anger at one's mother and dissimilar to the representational properties of a conscious disappointment with one's mother.

One immediate objection to this suggestion would be that the representational properties of anger at one's mother and disappointment with one's mother are the same: both states represent one's mother. If so, it cannot be that the representational properties of an unconscious anger are like those of conscious anger and unlike those of conscious disappointment, because the representational properties of conscious anger and disappointment are the same.

To respond to this objection, the proponent of the representationalist suggestion would have to argue that the representational properties of anger and disappointments are different after all. This may not be all that implausible. Thus, I have argued elsewhere that whereas anger at one's mother represents one's mother *as angering,* disappointment with one's mother represents one's mother *as disappointing* (Kriegel, 2002b)[19]. These are two very different representations: one represents an angering object, the other a disappointing object.

Another objection may be that some emotions do not have *any* representational properties. Thus, it is commonly said that anxiety is an objectless emotion. Other examples may be elation, depression, and various moods (Searle, 1983: 1). If so, it cannot be that the representational properties of unconscious anxiety are similar to those of conscious anxiety, since neither has representational properties at all.

One problem with this objection is that much of the common wisdom on the lack of intentionality in some emotions may simply be false. Seager (1999: 183) has argued quite convincingly that depression does have a representational import. According to Seager, depression does not represent any particular object, but "colors" the representation of any objects one represents while depressed. When one is depressed, everything looks depressing, that is to say, everything one is

[19] More on this in the third part of the paper.

aware of one represents to oneself as uninteresting, dull, and worthless. Similar remarks apply to elation and anxiety: when one is elated, everything looks exciting; when one is anxious, everything looks worrisome.

A third objection to the representationalist suggestion would be that the representational properties of conscious and unconscious anger are simply not the same. Thus, according to Tye (2000), the phenomenal character of an emotional experience consists in the proprioceptive representation of certain bodily events. Unconscious emotions do not have a phenomenal character, and therefore do not represent those bodily events. So the representational properties of conscious and unconscious emotions are different.

The proponent of the representationalist suggestion could respond by denying that conscious emotions do in fact involve a proprioceptive representation of bodily events. But also, she could modify her position, claiming that what makes an unconscious anger relevantly similar to a conscious anger and dissimilar to a conscious disappointment is the similarity in their *non-proprioceptive representational content.* Clearly, emotions represent more than just bodily events. A state of anger at one's mother may represent some bodily events (e.g., irritation), but it certainly also represents one's mother, and moreover, represents her as angering. It could be argued that what makes an unconscious anger at one's mother a state of anger is the fact that, despite not representing the relevant bodily events, it does represent one's mother as angering.

The various objections to the representational suggestion I have reviewed thus far appear unsuccessful, then. But there is one rejoinder to the representational suggestion which does not even attempt to refute it. This is to claim that the intentionality of mental states is itself fundamentally dependent upon consciousness. Thus, McGinn (1988: 299-300) writes[20]:

> One view, by no means absurd, is that all [representational] content is originally of conscious states. There is no (underivative) intentionality without consciousness... Our attributions of content to machines and cerebral processes is, on this view, dependent or metaphorical or instrumental; there would be no content in a world without consciousness.

Variations on this view have been recently defended by several authors (Horgan and Tienson, 2002; Kriegel, 2003d; Searle, 1991, 1992 Chap. 7; Williford, 2005).

There are two main arguments for this view. Both arguments proceed by claiming that there is a certain asymmetry between conscious and unconscious representations, such that unconscious representations must derive their aboutness, or intentionality, from conscious representations, whereas the latter do not derive

[20] The page numbers refer to the reprint of McGinn's article in Block et al. (1997).

theirs from anything else, hut rather have it in and of themselves. The difference between the two arguments is in the specific asymmetry they identify.

According to Searle (1991), an unconscious belief that there are rabbits in England is nothing more than a neurological event in the brain, and as such cannot be inherently about something. Rather, it derives its aboutness from conscious belief about rabbits: the only thing that makes an unconscious belief a belief about rabbits, rather than about (say) dolphins, is that *if* it were conscious, it would be inherently about rabbits and not about dolphins. A similar, if somewhat subtler argument, is developed by McGinn, according to whom conscious representation, unlike unconscious representation, is Janus-faced: it has an outward face (which it shares with unconscious representation) directed at whatever external object is being represented; but it also has an inward face (which unconscious representation lacks) directed at the subject of experience. This second face of conscious representation is responsible for making the external object *present* to the subject. Without such presence to the subject, McGinn claims, a mental state is not inherently about the object.

We cannot here undertake a full examination of the merits and demerits of these two arguments[21]. But it is quite plausible that there is some sort of asymmetry between conscious and unconscious representation which makes the former dependent upon the latter. If so, then as McGinn puts it in the above quoted passage, there would be no intentionality in a world without consciousness. This means that even if a representationalist answer to the two questions posed in §8 could be given, it would not undermine the thesis that there would be no emotion in a world without consciousness.

In conclusion, there are good reasons to think that the realm of emotion is fundamentally dependent upon the existence of consciousness, in that the essential properties of emotional states are their actual or counterfactual phenomenal properties. My argument for this view has been basically an argument by elimination: nothing else seems to account for what makes an emotion the emotion it is and an emotion at all. To be sure, the argument I have developed in this part of the paper has been canvassed in broad strokes, and can certainly not be taken to definitively establish the dependence of emotion on the existence of consciousness. At the same time, it suggests that the conception of emotion we have been accustomed to working with since Freud, in which emotion and consciousness are completely dissociated, may have less going for it than we ordinarily tend to think. In any event, the above discussion suggests that a full theoretical understanding of emotion is impossible as long as our understanding of consciousness is incomplete. In particular, a reductive account of emotions

[21] For such detailed examination, see Kriegel, 2003d.

must include a reductive explanation of what makes them emotions (and moreover, the particular emotions they are). If the argument of this part is on the right track, this would involve a reductive explanation of the phenomenal properties of emotional experiences. In the next part of the paper, I outline such a reductive explanation.

III. The Reduction of Emotional Consciousness

10. The Phenomenological Structure of Conscious Emotional Experience

In the first part of this paper, I argued that the reduction of emotion is not ruled out by a priori considerations pertaining to the multiple realizability of emotional properties such as the property of being angry. What a reduction of emotion would require, on the view I have defended, is the empirical discovery of a unique realizer for each emotional property. Such a discovery would entitle us to claim that the unique realizer constitutes the emotional property in question, making the latter "nothing but" its realizer. In the second part I argued that some a priori considerations suggest that a reduction of emotion cannot proceed without a reduction of emotional consciousness. On the view I have defended, it is impossible to isolate the theory of emotion from the theory of consciousness in such a way that the reductive explanations of these two phenomena unfold independently of each other. Superimposing the results of the first two parts of the paper, it is clear that a necessary condition for the reduction of emotion is the empirical discovery of a unique realizer of emotional consciousness. In this third part, I want to bring further a priori considerations to bear in speculating about the nature of any such unique realizer[22]. The question I want to address is, What sort of brain structure might possibly constitute emotional consciousness?

Before we tackle the issue of a reductive explanation of emotional consciousness, it is imperative that we get clear on the nature of the explanandum. When a mental state is conscious, there is *something it is like* for the subject to have it (Nagel, 1974). Thus, when I am consciously angry, there is something it is like for me to be angry. In particular, there is an anger-ish way it is like for me to be in the mental state I am in. This "anger-ish way it is like for me" is the

[22] Let me stress that I am not using the term "a priori" in a strict sense, let alone assume a dichotomist conception according to which every proposition is purely a priori or purely a posteriori. Rather, following Quine (1969), I conceive of the a priori and the a posteriori as two opposing poles of a continuous spectrum. So when I say that I will discuss certain a priori considerations, my claim is that the consideration I will discuss are *relatively* a priori, that is, relatively observation-independent.

phenomenal character of my emotional experience. And this phenomenal character has a certain internal structure that must be brought out before a theoretical account of it can be provided or even envisaged.

In the first instance, two dimensions of a phenomenal character such as the "anger-ish way it is like for me" should be distinguished: (i) the anger-ish aspect, and (ii) the for-me aspect (Levine, 2001). Let us call the former the *qualitative character* of the experience and the latter the *subjective character* of the experience. On the one hand, the experience has a certain quality, which is its anger-ish character. On the other hand, I am *aware* of this anger-ish quality. The experience not only *has* this quality, I am also aware of the quality. In this sense, the experience is not just *in me*, it is *for me*. The phenomenal character of a conscious experience is a matter of these two elements - the quality and the awareness of it - coming together in a single mental state.

The subjective character of an emotional experience is normally a matter of what I have called elsewhere *peripheral self-awareness* (Kriegel, 2004). It is a phenomenon of awareness, to be sure, since there is no sense in which a mental state could be *for me* if I am unaware of its occurrence. It is, moreover, a form of *self*-awareness, since it involves not only my awareness of my anger, but an awareness of it precisely as *my* anger. However, only very rarely do we explicitly dwell on our emotions in a way that makes us focus *on* them. Ordinarily, in having an emotional experience, we are focused mainly on the object of the emotion. Thus, if I am angry that the phone bill is wrong again, the focus of my experience is not on my anger, but on the wrong phone bill. Yet as we have just seen, I am necessarily aware of my anger. So this form of awareness must be a non-focal awareness, or as I prefer putting it, *peripheral* awareness[23]. Our awareness of our concurrent emotional experiences, which constitutes these experiences' subjective character, is therefore normally a form of peripheral self-awareness.

As for the qualitative character of our emotional experiences, it seems to exhibit an internal structure as well. To see this, consider the seldom appreciated fact that life without conscious emotional experiences is not worth living. Take away a person's capacity to undergo conscious emotional experiences and there would be all but nothing to keep this person going. This raises the question, What is it about emotional experience that bestows on it this all-important worth? What is it about emotional experience that makes us care about it so much, indeed, that *defines* our caring about anything?

[23] The distinction between focal and peripheral awareness is most easily drawn in the case of visual awareness, where there is clearly a distinction between what we are aware of through foveal vision and what we are aware of through peripheral vision. But the same distinction applies to self-awareness, with a distinction between what we are focally self-aware of and what we are peripherally self-aware of. For more on this, see Kriegel, 2004.

A plausible answer to this question is that we care about emotional experiences because they involve essentially sensations of *pleasure and pain.* Thus, experiences of sadness, shame, and jealousy are unpleasant and often painful, while experiences of joy, pride, and love are pleasant. Plausibly, every conscious emotion involves, if ever so dimly, a *good feeling* or a *bad feeling*[24]. And it may well be that we care so much about our conscious emotions precisely because - and to the extent that - they involve good or bad feelings. We want to avoid the bad feelings and indulge in the good feelings that come with these emotions[25]. There is nothing to the motivational force of conscious emotion over and above that.

At the same time, everything we know about the neuroscience of pleasure and pain suggests that the pleasure (or pain) involved in various emotions is in itself the selfsame sensation. Jealousy and shame feel different, and they both hurt (i.e., they feel bad). On the face of it, it may seem that they do not hurt in the same way: jealousy hurts in one way and shame in another. However, upon reflection it appears that the pain involved in jealousy and shame is - inasmuch as it is pain - one and the same. There is only one kind of pain (although it may come in different intensities), and it shows up as a component of different emotions and feelings. The pain itself always feels the same.

Consider, by analogy, painful tactual experiences. A cut in one's finger and a burn in one's finger both hurt, and they do not hurt in the same way. However, these tactual experiences may be decomposed into two separable components: the purely tactual quality of a cut or a burn, and the pain quality simultaneous with it. The overall feeling is different, but the pain component is the same. The difference between the way a cut hurts and the way a burn hurts is not in the pain *per se,* but in the purely tactual sensation that goes with it. Likewise, I would maintain that the qualitative character of conscious emotional experiences is always a combination of two separate factors, which we may call the *somatic factor* (pleasure/pain) and the *purely emotional factor* (the non-somatic residue)

In conclusion, the phenomenal character of emotional experiences, which is often construed as simple and incomposite, exhibits in reality a quite intricate internal structure, in that it is a whole with identifiable component parts. The

[24] According to some, every conscious experience whatsoever (emotional or other) involves an element of good feeling or bad feeling (see Searle, 1992, Ch. 6). This strikes me as correct, though I will not argue for it here.

[25] As against this, someone might argue that we care about our conscious emotions beyond the good feeling and bad feeling they involve. But this in itself does not answer our question as to why we should care as much as we do about conscious emotions. The virtue of the view discussed in the text is that it is *explanatory,* it illuminates the fact that we care so much about our emotional experiences.

main component parts are subjective character and qualitative character. The latter involves a somatic component and a purely emotional component. An account of the phenomenal character of emotional experiences (hence of emotional consciousness) would have to account for all three dimensions of conscious emotion. In the next section, I sketch a reductive account of qualitative character, which is developed in greater detail elsewhere (Kriegel, 2002b, 2002c). In the section after that, I will sketch a reductive account of subjective character, also developed in greater detail elsewhere (Kriegel, 2002a, 2003a, 2003c). Together, they amount to a proposal regarding the unique realizer of phenomenal consciousness, including (as a special case) emotional consciousness.

11. A Reductive Account of Qualitative Character

Several philosophers today defend a *representational theory of consciousness* (Dretske, 1995; Harman, 1990; Tye, 1995, 2000). According to this theory, the phenomenal character of a conscious experience is nothing but its representational content. When I look up at the blue sky, there is a bluish way it is like for me to have my visual experience. According to representationalists, this bluish way it is like for me is a matter of the experience's representation of the sky's being blue.

One obvious problem with the representational theory of consciousness is that it is unclear how the representation of the sky's being blue can account for the for-me-ness of the visual experience. That is, the subjective character of conscious experience does not seem to be accounted for by the fact that something blue is represented. For something blue can be represented by unconscious mental states as well, that is, mental states that do not have a subjective character. Thus, when I have a subliminal perception of the blue sky, the perceptual representation of the sky occurs *in me*, but it is not/or *me*. It has no subjective character[26].

This objection suggests that the representational theory may be *at most* an account of the qualitative character of experience, not of its phenomenal character. However, even as an account of qualitative character the theory faces serious difficulties. A full discussion of these difficulties will take us too far afield, but several other philosophers have offered a modification of the representational theory that seems to handle these difficulties. This is to suggest that my visual experience does not represent the sky's being blue, but rather the sky's appearing blue (or the appearance of the sky's being blue). On this suggestion, which I endorse, the bluish character of a visual experience of the sky is a matter of the experience's representation of the sky's appearing blue (Kriegel, 2002c;

[26] For an overview discussion of the phenomenon of subliminal perception, see Dixon 1971.

Shoemaker, 1994, 2002; Thau, 2002). What makes a given property an *appearance property* is, according to Shoemaker, that it can only be instantiated relative to a sentient subject. That is, the sky's property of appearing blue is in fact a *relation* between the sky and the perceiver[27].

This modified representational account of qualitative character can be applied to emotional experiences as well (Kriegel, 2002b). What gives a conscious experience of anger its anger-ish qualitative character is the fact that the experience represents certain external objects as angering. Thus, when I am angry that the phone bill is wrong again, my emotional experience represents the bill's wrongness, but it does not represent it neutrally; rather, it represents the bill's wrongness *as angering.* Now, the property of being angering is subject-relative, since two objects or states of affairs that are intrinsically indistinguishable can anger one person and fail to anger another (Kriegel, 2002b). (Another person, who is more patient than I am with the phone company, might not be angered by the very same wrong phone bill.) So the property of being angering is an appearance property (even though we need not use the noun "appearance," or its cognates, in *designating* it).

It might be objected, justifiably, that none of this seems to account for the somatic factor in emotional experience. Being angry feels unpleasant, but the unpleasant feeling is not accounted for by the representation of an external object as angering. However, several philosophers have offered a representational account of pain. Thus, Tye (1990) claims that the hurt-ish character of a pain experience is a matter of the experience's representation of tissue damage or some such bodily event. Shoemaker (1994) offers a representational account that focuses on the appearance properties of such bodily events: a feeling of pain in the toe is a mental state that represents the appearance of tissue damage in the toe. This account can evidently be used to complement the modified representational account of the purely emotional factor in the qualitative character of emotional experience.

The great advantage of a representational account of qualitative character is that we have today a clear notion as to how the phenomenon of mental representation is to be reductively explained in information-theoretic terms, hence in physical terms. This is mainly due to the work of Fred Dretske (1981, 1988). 1 will not go into Dretske's so-called "informational semantics" here. Suffice it that we accept that some fully satisfactory information-theoretic account of

[27] This may or may not be the case with the property of being blue. If it is, then *being blue* simply is *appearing blue,* and the modified representational account of qualitative character is the representationalist's only option (Kriegel, 2002c). Let me note here that according to Thau the relevant appearance properties are not relational but intrinsic. However, this leaves Thau without an account of what makes these properties appearance properties in the first place.

representation will eventually emerge. If this is the case, then a representational account, even in its modified form, holds the key to a three-step reduction of qualitative character: first, qualitative character is reduced in terms of mental representation; second, mental representation is reduced in information-theoretic terms; and finally, information theory is reduced to physics (augmented, presumably, with probability theory).

12. A Reductive Account of Subjective Character

As for the subjective character of experience, it too may be representationally reducible. Thus, one theory of consciousness that is quite popular among philosophers and cognitive scientists alike is the *higher-order monitoring theory* (Armstrong 1968; Lycan 1996; Rosenthal, 1990, 2004). According to the higher- order monitoring theory, a conscious experience has phenomenal character because it is represented by another mental state of mine. The argument for this is straightforward: for-me-ness requires that I be aware of the experience, and to be aware of something is to represent it; therefore, an experience exhibits for- me-ness only if it is represented (Lycan, 2001). This requires, according to the higher-order monitoring theorist, that I enter *another* mental state, which would represent my experience.

Again, the higher-order monitoring theory is more safely construed as an account of subjective character than phenomenal character, for the presence higher-order representation does not seem to account for the qualitative character of an experience. Thus, it is not because my experience of the sky is represented by me that the experience is *bluish*, though it may well be that this is why it is bluish *for me.*

However, even construed more narrowly as an account of subjective character, the higher-order monitoring theory faces a number of difficulties, which again I will not dwell on. These difficulties have recently given rise to the rebirth of an account of consciousness with a venerable tradition behind it. According to this account, which we may call the *same-order monitoring theory*, a mental state has subjective character when it is represented, not by a *separate* mental state, but *by itself.* On this view, conscious experiences, whatever else they may represent, always, and by necessity, also represent their own occurrence (Kriegel 2002a, 2003a; Natsoulas, 1996; Smith, 1986)[28].

[28] As I said, this view has a venerable tradition. According to Caston (2002), it was in fact Aristotle's view. This is no doubt where Brentano (1874) inherited the view from. Through Brentano's influence, it has become the default position within the phenomenological tradition (see Zahavi, 1999 for a recent overview). Incidentally, it seems to have been Freud's view as well (see Natsoulas, 1984).

The same-order monitoring account applies readily to emotional experiences. The idea would be that conscious emotional experiences, whatever else they represent, represent also themselves. On the account of emotional consciousness here being outlined, conscious emotions represent both themselves and certain appearance features in the external environment - and it is in virtue of representing these two sorts of things that they have their phenomenal character. That is, it is in virtue of this complex representational (and self-representational) profile that they are the conscious emotional experiences they are (and conscious emotions at all). Thus, when I am consciously angry about the phone bill being wrong, my emotional state represents both the fact that the phone bill's being wrong is angering and the fact that I myself am thereby angered by the phone bill's being wrong. And it is precisely in virtue of representing all this that my emotional state is the mental state it is.

As noted above, a representational account has the advantage of paving the way for an information-theoretic reduction. However, the capacity of conscious experiences to represent themselves introduces a complication, since it is not obvious how a mental state could carry information about itself in a non-trivial manner (that is, not by just *being* itself). Elsewhere, I have suggested a three- phase mechanism that may mediate the formation of self-representing states (Kriegel, 2002a, 2003b). Applied to the case of emotional experience of anger, the mechanism in question proceeds as follows. In the first phase, the subject forms a representation of the fact that some object or state of affairs is angering. In the second phase, the subject forms a higher-order representation of that representation (that is, a representation of the representation of the angering object). In the third, and crucial, phase, the two representations are integrated, or unified, through one of the cognitive system's processes of information integration.

Cognitive processes of information integration are not unfamiliar. At the personal level, there is the conscious inference in accordance with "conjunction introduction," as when one consciously infers that the wall is white and rectangular from one's thoughts that the wall is white and that the wall is rectangular. At the sub-personal level, there is the widely discussed process of *binding*, as when the brain binds information from the visual cortex and from the auditory cortex to form a single, unified visuo-auditory representation of a single object (see Engel et ah, 1999 for a comprehensive review). On the suggestion I am making, a conscious emotion arises when an emotional representation and the representation of that representation are integrated into a single mental state through a cognitive process of this sort[29]. The mental state produced through such a process would

[29] The process in question is probably different from either feature binding or conscious inference in accordance with conjunction introduction. But there is no reason to think that these are the only

fold within it, as it were, a representation and a representation of that representation, in such a way as to make it, in effect, self-representing. Thus, the integration, or unification, of a representation of the phone bill as angering with the representation of that representation would result in the formation of a conscious anger about the phone bill.

In summary, according to the reductive explanation of subjective character propounded here, the reduction of the subjective character is to proceed in similar fashion to the reduction of qualitative character, with the complication that a cognitive mechanism of information integration is posited whereby first-order and second-order representations are integrated or unified. My suggestion is that each product of the operation of the mechanism in question is the unique realizer of a conscious emotional experience, and hence an appropriate reducer of it.

13. Conclusion: How to Reduce the Realm of Emotion

The outlook I have developed in the present paper features the following tenets. A reductive account of emotion must target first and foremost the properties of emotional states that make them the emotional states they are, and emotional states at all. Emotional states can be either conscious or unconscious. When they are conscious, what makes them the emotions they are, and emotions at all, is the phenomenal character they have. When they are unconscious, what makes them the emotions they are, and emotions at all, is the phenomenal character they would have if they were conscious. Therefore, a reductive account of emotion must target the phenomenal character of conscious emotions. What such a reductive account requires is the identification of the unique realizer of the phenomenal character of conscious emotions. This unique realizer, by virtue of constituting the phenomenal character of conscious emotions, is a reducer of it.

The phenomenal character of conscious emotions has two main components, the subjective character and the qualitative character of the conscious emotional experience. Both can be accounted for representationally. The qualitative character can be accounted for in terms of the emotional experience's representation of certain appearance properties of external objects, the subjective character in terms of the emotional experience's representation of itself. Thus what makes a conscious emotional state the emotional state it is, and an emotional state at all, is its particular representational character; and what makes an unconscious

processes of integration employed by our cognitive system. Any process in which two separate mental states or contents are unified in such a way that they are superseded by a single mental state or content that encompasses both will qualify as a process of cognitive integration.

emotional state the emotional state it is, and an emotional state at all, is the particular representational character it would have if it were conscious. This representational character is reductively explicable in information-theoretic terms, hence ultimately in physical (and probability-theoretical) terms.

References

Armstrong, D. M. 1968. *A Materialist Theory of the Mind.* New York: Humanities Press.

Armstrong, D. M. 1978. *A Theory of Universals,* vol. 2. Cambridge: Cambridge UP.

Brentano, F. 1874. *Psychology from Empirical Standpoint.* Ed. O. Kraus. Ed. of English edition L. L. McAlister, 1973. Translation A. C. Rancurello, D. B. Terrell, and L. L. McAlister. London: Routledge and Kegan Paul.

Caston, V. 2002. Aristotle on consciousness. *Mind,* 111: 751-815.

Churchland, P. M. 1984. *Matter and Consciousness.* Cambridge MA: MIT Press. DeLancey,

C. 2001. *Passionate Engines.* New York: Oxford UP.

Dixon, N. F. 1971. *Subliminal Perception: The Nature of a Controversy.* London: McGraw- - Hill.

Dretske, F. I. 1981. *Knowledge and the Flow of Information.* Oxford: Clarendon Press. Dretske,

F. I. 1988. *Explaining Behavior.* Cambridge MA: MIT Press.

Dretske, F. I. 1995. *Naturalizing the Mind.* Cambridge MA: MIT Press.

Engel, A. K,, Fries, P, Konig, P, Brecht, M., and Singer, W. 1999. Temporal binding, binocular rivalry, and consciousness. *Consciousness and Cognition,* 8: 128-151.

Fodor, J. A. 1974. Special sciences. *Synthese,* 28: 97-115.

Freud, S. 1915. The unconscious. In his *Metapsychological Essays* (pp. 159-215). Trans. J. Strachey. New York: Collier/Macmillan, 1963.

Geach, P. T. 1967. Identity. *Review of Metaphysics,* 21: 3-12.

Harman, G. 1990. The intrinsic quality of experience. *Philosophical Perspectives,* 4: 31-52.

Horgan, T. 1997. Kim on mental causation and causal exclusion. *Philosophical Perspec tives,* 11: 165-184.

Horgan, T., Tienson, J. 2002. The intentionality of phenomenology and the phenomenology of intentionality. In D. J. Chalmers (ed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 520-533). Oxford and New York: Oxford UP.

Johnston, M. 1992. Constitution is not identity. *Mind,* 101: 89-105.

Johnston, M. 1997. Manifest kinds. *Journal of Philosophy,* 94: 564-583.

Kim, J. 1976. Events as property exemplifications. In Brand, M., Walton, D. (eds.), *Action Theory* (pp. 159-177). Dordrecht: Kluwer Academic Publishers.

Kim, J. 1989a. The myth of nonreductive materialism. *Proceedings and Addresses of the American Philosophical Association,* 63: 31-47.

Kim, J. 1989b. Mechanism, purpose, and explanatory exclusion. *Philosophical Perspectives,* 3: 77-108.

Kim, J. 1992. Multiple realization and the metaphysics of reduction. *Philosophy and Phenomenological Research,* 52: 1 -26.

Kim, J. 1998. *Mind in a Physical World.* Cambridge MA: MIT Press.

Kriegel, U. 2002a. Consciousness, permanent self-awareness, and higher-order monitoring. *Dialogue,* 41: 517-540.

Kriegel, U. 2002b. Emotional content. *Consciousness and Emotion,* 3: 213-230.

Kriegel, U. 2002c. Phenomenal content. *Erkenntnis,* 57: 175-198.

Kriegel U. 2003a. Consciousness, higher-order content, and the individuation of vehicles. *Synthese,* 134: 477-504.

Kriegel, U. 2003b. Consciousness as sensory quality and as implicit self-awareness. *Phenomenology and the Cognitive Sciences,* 2: 1-26.

Kriegel, U. 2003c. Intrinsic theory and the content of inner awareness. *Journal of Mind and Behavior,* 24: 171-198.

Kriegel, U. 2003d. Is intentionality dependent upon consciousness? Forthcoming in *Philosophical Studies.*

Kriegel, U. 2004. The functional role of consciousness. Forthcoming.

Kripke, S. 1980. *Naming and Necessity.* Cambridge MA: Harvard UP.

Lewis, D. C. 1993. Causal explanation. In Ruben, D.-H. (ed.), *Explanation.* Oxford: Oxford UP.

Lycan, W. G. 1996. *Consciousness and Experience.* Cambridge MA: MIT Press.

Lycan, W. G. 2001. A simple argument for a higher-order representation theory of consciousness. *Analysis,* 61: 3-4.

McGinn, C. 1988. Consciousness and content. *Proceedings of the British Academy,* 76: 219-239. Reprinted in Block, N. J., Flanagan, 0., and Guzeldere, G. (eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 295-308). Cambridge MA: MIT Press, 1997.

Nagel, T. 1974. What is it like to be a bat? *Philosophical Review,* 83: 435-450. Natsoulas, T.

1984. Freud and consciousness: I. Intrinsic consciousness. *Psychoanalysis and Contemporary Thought,* 7: 195-232.

Natsoulas, T. 1996. The case for intrinsic theory: I. An introduction. *Journal of Mind and Behavior,* 17: 267-286.

Perkins, M. 1966. Emotion and feeling. *Philosophical Review,* 75: 139-160.

Putnam, H. 1967/1991. The nature of mental states. In Rosenthal, D. M. (ed.), *The Nature of Mind* (pp. 197-203). Oxford and New York: Oxford UP, 1991.

Quine, W. V. O. 1969. Epistemology naturalized. In his *Ontological Relativity* (pp. 68- -90). New York: Columbia UP.

Raymont, P. 2003. Kim on overdetermination, exclusion, and nonreductive physicalism. In Walter, S., Heckman, H.-D. (eds.), *Physicalism and Mental Causation* (pp. 225- 242). London: Imprint Academic.

Rey, G. 1980. Functionalism and the emotions. In Rorty, A. (ed.), *Explaining Emotion* (pp. 152-163). Berkeley CA: Berkeley UP.

Rosenthal, D. M. 1990. A theory of consciousness. ZiF Technical Report 40, Bielfield, Germany. Reprinted in Block, N. J., Flanagan, O., and Guzeldere, G. (eds.), *The Nature of Consciousness: Philosophical Debates* (pp. 729-754). Cambridge MA: MIT Press, 1997.

Rosenthal, D. M. 2004. *Consciousness and Mind.* Oxford: Oxford UP.

Seager, W. 1999. *Theories of Consciousness.* London and New York: Routledge.

Searle, J. R. 1983. *Intentionality.* Cambridge: Cambridge UP.
Searle, J. R. 1991. Consciousness, unconsciousness, and intentionality. *Philosophical Issues,* 1: 45-66.
Searle, J. R. 1992. *The Rediscovery of Mind.* Cambridge MA: MIT Press.
Shoemaker, S. 1994. Phenomenal character. *Nous,* 28: 21-38.
Shoemaker, S. 2002. Introspection and phenomenal character. In Chalmers, D. J. fed.), *Philosophy of Mind: Classical and Contemporary Readings* (pp. 457-472). Oxford and New York: Oxford UP.
Smart, J. J. C. 1959. Sensations and brain processes. *Philosophical Review,* 68: 141-156. Reprinted in Rosenthal, D. M. fed.), *The Nature of Mind* (pp. 169-176). Oxford and New York: Oxford UP, 1991.
Smith, D. W. 1986. The structure of (self-)consciousness. *Topoi*, 5: 149-156.
Thau, M. 2002. *Consciousness and Cognition.* Oxford and New York: Oxford UP.
Tye, M. 1990. A representational theory of pains and their phenomenal character. *Philosophical Perspective,* 9: 223-239.
Tye, M. 1995. *Ten Problems of Consciousness.* Cambridge MA: MIT Press.
Tye, M. 2000. *Consciousness, Color, and Content.* Cambridge MA: MIT Press. Velmans, M.
1992. Is human information processing conscious? *Behavioral and Brain Sciences,* 14: 651-669.
Williford, K. 2005. The intentionality of consciousness and the consciousness of intentionality. In Forrai, G., Kampis, G. (eds.), *Intentionality: Past and Future.* Dordrecht: Kluwer Academic Publishers.
Zahavi, D. 1999. *Self-awareness and Alterity.* Evanston IL: Northwestern UP.