Dissociable Realization and Kind Splitting

Carl F. Craver†‡

It is a common assumption in contemporary cognitive neuroscience that discovering a putative realized kind to be dissociably realized (i.e., to be realized in each instance by two or more distinct realizers) mandates splitting that kind. Here I explore some limits on this inference using two deceptively similar examples: the dissociation of declarative and procedural memory and Ramachandran’s argument that the self is an illusion.

1. Introduction. The main point of clarifying the concept of realization is to better understand how our taxonomy of kinds ought to track or accommodate what we know about their realizers. Discussion of this accommodating relation has focused almost exclusively on cases of multiple realization—cases in which the same kind could be realized by different realizers in different instances. Yet there has been almost no discussion of dissociable realization—cases in which instances of what was previously thought to be a single kind are discovered to be realized by more than one distinct and independent realizer. It is a common (but rarely explicit) methodological assumption in neuroscience and elsewhere that discovering a kind to be dissociably realized mandates splitting the kind into as many as there are dissociable realizers.

Here I explore some limits on this assumption using two cases: one that is accepted as an exemplar of successful application of this principle, and one that is a bit more suspect. The first is the dissociation of procedural and declarative memory and the corresponding inference that memory is not a single kind. The second is V. S. Ramachandran’s (1998) argument that “the self is an illusion”: because the self is dissociably

†To contact the author, please write to: Department of Philosophy, Philosophy Neuroscience and Psychology, Washington University, One Brookings Dr., Busch Hall, Rm. 225, St. Louis, MO 63130; e-mail: ccraver@artsci.wustl.edu.
‡Thanks to Lindley Darden, Stuart Glennan, Jeff Poland, Rob Wilson, Peter Machamer, and Barbara Von Eckardt for comments on earlier drafts.
realized, what we previously thought of as “the self” should be split into as many selves as there are realizers—which is a bit like having no self at all. The arguments are deceptively similar and the empirical evidence in each case is equally compelling. Yet there is a way to avoid the conclusion in the second without jeopardizing the inference in the first. The difference between the two turns on the fact that each exemplifies a different variety of realization, one of which sustains the argument for splitting and one of which does not.

2. Splitting Memory. Memory, as understood from the perspective of contemporary cognitive neuroscience, is not a single kind of thing. This conviction is typically justified by noting that different kinds of memory can be “dissociated” from one another through brain damage or experimental manipulation. One example is the splitting of declarative and procedural memory.

The argument for splitting typically opens with the story of H. M. (Scoville and Milner 1957). In an effort to relieve life-threatening epileptic seizures, H. M. consented to experimental surgery removing bilaterally a structure known as the hippocampus. H. M. can still read and write, he can learn new skills, he can remember much of his childhood, and his IQ if anything increased after the surgery. The tragedy of H. M.’s life is that he permanently lost the ability to retain new memories for facts and events. He has lost declarative memory but has maintained procedural memory (see Corkin 2002). From the case of H. M. and from studies in nonhuman animals, researchers have concluded that the mechanisms for procedural and declarative memory are distinct. Since the two are realized by distinct mechanisms, the kind “memory” is confused, lumping together what we now know to be two distinct kinds of memory.

Arguments from dissociable realization are central to the contemporary neuroscience of memory. Daniel Schacter writes:

we have now come to believe that memory is not a single or unitary faculty of the mind, as was long assumed. Instead, it is composed of a variety of distinct and dissociable processes and systems. Each system depends on a particular constellation of networks in the brain that involve different neural structures, each of which plays a highly specialized role within the system. (1996, 5)

Similar arguments can be found in Weiskrantz 1990, Squire 1992, Schacter and Tulving 1994, and Squire and Knowlton 1994. Philosophers of neuroscience have used this case as an illustration of the ability of neuroscientific research to drive conceptual revision in psychology (e.g., Bickle 1998, Churchland 1986).

In each case, the authors argue for kind splitting on the basis of a
premise about dissociable realization. The argument might be represented as follows:

M1. Declarative memory and procedural memory can be independently disrupted.

M2. If phenomena $X$ and $Y$ can be independently disrupted, then they have distinct realizers.

M3. Procedural and declarative memory have distinct realizers.

M4. No Dissociable Realization (NDR). Instances of a natural kind have one and only one realizer. If there are two distinct realizers for a putative instance of a kind, there are really two kinds, one for each realizer.

M5. So procedural and declarative memory are distinct kinds of memory.

Here I will be concerned primarily with NDR. NDR is stated as a sufficient condition for kind splitting; of course there are other kinds of evidence as well. Researchers studying memory (and cognition generally) have a variety of accepted techniques for distinguishing cognitive phenomena, thereby leaping from M1 to M5 directly (an example opens the next section). The distinction between procedural and declarative memory predates H. M. But the case of H. M. (and subsequent research) seemingly removed any doubt: the declarative/procedural distinction is reflected in the architecture of the brain. NDR is phrased to reflect the force of such evidence for kind splitting.

NDR stops short of requiring one-to-one correspondence between realized kinds and their realizers. It asserts only that dissociating realizers mandates splitting kinds. It is consistent with NDR that there might be two realized kinds and one realizer. (Plaut 1995 argues that one can independently disrupt memory categories that share a common neural network realizer.) NDR also leaves it an open question whether the unification of two or more realizers that were previously presumed to be disparate would mandate unifying the realized kinds (as in, perhaps, the case of rusting, burning, and breathing). And finally, NDR takes no stance as to whether discoveries about realized kinds should influence views about lower-level realizers. These are all important (and underdiscussed) questions; but here I confine my attention primarily to NDR.

3. Splitting the Self. In Phantoms in the Brain, V. S. Ramachandran (1998) provides a structurally similar argument that our sense of self is an illusion. He develops his argument with the aid of “Titchener’s circles” (Figure 1). A circle looks larger when surrounded by small circles than when it is
Figure 1. Titchener’s circles.

surrounded by large circles. The effect, which can be measured by allowing subjects to change the size of one center circle until it visually matches the other center circle, can increase the apparent size of the center circle by as much as 30%. Recent experiments by Salvatore Aglioti et al. (1995) show that if the experimental subject reaches for a poker chip the size of the center circle, the fingers are right every time: the forefinger and thumb open the right amount to pick up the chip, and they open as much when the circle is surrounded as when it is alone. It appears that the system reporting perceptual experience and the system doing the reaching are distinct: one is fooled; the other is not.

This behavioral argument for dissociation is reinforced by evidence from people who have brain damage. Consider David Milner’s case-study, DF, who suffered anoxic brain damage when carbon monoxide from her water heater overwhelmed her (Milner and Goodale 1995). When she came to, she was blind. Although she gradually regained the ability to recognize colors and textures, she could not visually recognize the shapes of objects or faces. She says that she can’t see, and she fails tests of object recognition and naming. Nonetheless, she can perform complex reaching tasks—such as putting a letter through a mail slot or reaching for a pencil—without error. And if you ask DF to reach for the center chip, her hands aren’t fooled. DF’s lesion uniquely damaged her visual system’s
ventral pathway (featuring inferotemporal cortex), which houses crucial mechanisms involved in consciously recognizing “what” objects are. Yet the event spared her dorsal pathway (featuring the posterior parietal cortex), which houses crucial mechanisms involved in guiding visual interactions with objects (and so is sometimes called the “how” pathway). Ramachandran concludes:

The most obvious fact about existence is your sense of being a single, unified self “in charge” of your destiny; so obvious, in fact, that you rarely pause to think about it. And yet Dr. Aglioti’s experiment and observations of patients like DF suggest that there is in fact another being inside you that goes about his or her business without your knowledge and awareness. And as it turns out, there is not just one such zombie but a multitude of them inhabiting your brain. If so, your concept of a single “I” or ‘Self” inhabiting your brain may be simply an illusion—albeit one that allows you to organize your life more efficiently, gives you a sense of purpose and helps you interact with others. (83–84)

Although the self may be a “practical kind” (helping to organize our lives and our commerce with others), it is not a natural kind. Ramachandran clarifies his conclusion: “Here and elsewhere, when I say that the self is an ‘illusion’ I simply mean that there is no single entity corresponding to it in the brain” (272). Ramachandran does not notice an inferential gap between the claim that there is no single entity in the brain corresponding to the self and the claim that there is no such thing as the self (i.e., that our sense of self is illusory). He does not notice it because the inference is so imbedded in the methodological medium of contemporary biology and cognitive neuroscience.

Ramachandran’s argument is superficially similar to the argument for splitting memory:

S1. Conscious visual object recognition and visually guided action can be independently disrupted.

S2. If phenomena X and Y can be independently disrupted, then X and Y have distinct realizers.

S3. So conscious visual object recognition and visually guided action have distinct realizers.

S4. (NDR). Instances of a natural kind have one and only one realizer. If there are two distinct realizers for a putative instance of a kind, there are really two kinds, one for each realizer.

S5. There is no single thing, the self, that is responsible for both
conscious visual object recognition and visually guided action; the notion of a single unified self is an illusion.

If anything, a stronger case can be made for premise S3 in the case of the self than can be made in the case of memory. After all, the central dogma of contemporary neuroscience holds that functions localize in the brain, and if that is right, then the brain is composed of distinct and independent mechanisms realizing more or less distinct and independent kinds of phenomena. If NDR is right, the conclusion follows.

But the conclusion is suspect. There are many areas of neuroscience and psychology where the concept of the “self” is not merely respectable but fashionable: research on alien thoughts and actions, dissociative identity disorder, development of the self-concept, egocentric maps of space, episodic or autonoetic memory, kinesthetic and proprioceptive self-awareness, mirror self-recognition, and the recent panoply of work on consciousness and the biological basis of self-reference—these are all areas of cognitive science in which the notion of the self plays a central role in the causal and inferential structures described by their theories. Perhaps Ramachandran would take this diversity of topics as reinforcing his point: each of these fields focuses on a different capacity with different realizers and labels it, exclusively, the self. But one resisting Ramachandran’s eliminativist conclusion (as I am inclined) has available the option of rejecting NDR. Why can’t the self simply be the bundle of these dissociable capacities?

There are other clear violations of NDR. We did not give up on mice or frogs as kinds (or split them into numerous distinct and independent kinds of things) when we learned that instances of them were made up of innumerable distinct mechanisms. Nor did we give up on hearts, or kidneys, or solar systems. In fact, in cases such as these, the robustness underwriting their status as kinds depends precisely on the fact that they are associated with a cluster of regularly co-occurring properties and activities. This association between clusters of properties and kinds has been persuasively urged by, for example, Boyd (1999), Griffiths (1997), and Wilson (2001). Further, some of our best evidence for belief in different kinds of entities comes from the fact that they can be detected and manipulated with multiple causally and theoretically independent techniques. These considerations, as I will argue in Section 5, suggest a principled way to restrict NDR so as to block the conclusion for mice, frogs, and selves that follows so naturally for memories. But first, we must consider a necessary condition for splitting realized kinds.

4. Constitutive, Contextual, and Etiological Realizers. Realization is typically understood as a constitutive relationship. The realized phenomenon
is some activity $\Psi$ of an object or system $S$ in context $C$, and $\Psi$'s constitutive realizer is located within $S$'s spatial boundaries. Wilson (2001) rightly points out that a “constituency” thesis is part of nearly every explicit analysis of realization in contemporary philosophy. But, as Wilson notes, many of the activities in which objects engage cannot be exhaustively (sufficiently) realized by the parts of those objects. Sometimes the activities are realized (at least in large measure) by causal (or other) relations between $S$ and its context. And sometimes activities are realized (at least in part) by the historical processes leading up to them. To accommodate these possibilities, we may wish to recognize cases of contextual realization and cases of etiological realization (cf. Craver 2001, section 4) as distinct from the more familiar constitutive variety. Each of these approaches has figured prominently in discussions of multiple realizability. Fully articulating either would require considerable fine tuning. Yet the effort is unlikely to pay off in the present cases. The contextual and etiological realizers in each of our cases are as dissociable as the constitutive realizers.

Consider memory first. Declarative memories are triggered by the presentation of facts or the occurrence of events in the life of the person, and they play important roles in, for example, conversation, autobiography, or the simple act of reminiscing. Nondeclarative forms of memory (like procedural memory, iconic memory, priming, etc.) have their own unique triggering conditions (procedural memories are acquired by doing things, iconic memories by visual impressions, etc.) and play different roles in the life of the organism. These differences are reflected in the different kinds of stimuli used to produce and evoke memories of the different types. Similarly, the etiological realizers of declarative and procedural memory are likely to be distinct and independent as well. Phylogenetically, procedural forms of memory appear earlier than declarative memory (they can be found, for example, in sea slugs). Ontogenetically, skills, habits, and priming predate the ability to lay down lasting memories for facts and events. Appealing to contextual or etiological realizers in the memory case reinforces the splitter inference.

Similar arguments can be made for the two systems dissociated in Ramachandran’s argument about the self. Contextually, the inferotemporal “what” system damaged in DF is privy to different information than the dorsal “how” system. And it is clear from Aglioti’s experiments that the two systems are able to support different sorts of behavior in the world. The “how” system is contextually linked to reaching and the “what” system is contextually linked to object recognition. Etiologically, the “how” system is phylogenetically much older than the “what” system, and the two develop ontogenetically at different time scales. So, as with
the memory case, appeal to contextual or etiological realizers only reinforces the dissociation of realized kinds.

Differences in contextual or etiological realization do seem necessary for kind splitting. Maximally redundant dissociable realizers illustrate this point. The mechanisms of each kidney can be disrupted independently of those in the other, they are the same kinds of mechanism, and they both do the same thing. It seems reasonable to say that there are two mechanisms of the same kind jointly filtering the blood rather than that there are two kinds of filtering. After all, discussions of multiple realizability center on the fact that even different kinds of realizers might realize the same realized kind. A more complicated case involves two realizers, each of which is necessary for a realized kind, neither of which is sufficient for the kind, and neither of which produces any intelligible behavior operating in isolation (each lurches and hiccups incomprehensibly in the absence of its partner). In this case, failure to find an intelligible way to distinguish two realized phenomena by attention to their contextual or etiological realizers may trump the discovery that the constitutive realizers are dissociable. Each example illustrates that our willingness to split kinds extends only to those cases in which it is possible to distinguish them contextually and etiologically as well.

5. Varieties of Constitutive Realization. Focusing attention back on constitutive realizers, one could perhaps restrict NDR by recognizing different varieties of constitutive realization, depending upon what is getting realized and what is realizing (cf. Poland 1994). In particular, kinds of entities stand in a different relationship to their realizers than do kinds of properties or activities. The distinction between the two rests on the importance of explanatory relevance in articulating the latter, but not the former, sort of realization. It is open to those inclined against Ramachandran’s argument to hold that the self, conceived as an entity like mice or frogs, need not track realizing material constituents, properties, or mechanisms.

Some realized kinds are entities or objects, e.g., gems, corkscrews, and computers. The realizers of entities are typically described as material realizers. Material realizers are constituents: the material parts (however those are to be individuated) within the entity’s spatial boundaries. Nephrite is said to be realized by its molecular constituents, corkscrews by steel, and computers by silicon. Such simplified examples typically abstract from the messy details scientists confront; they turn heterogeneous material realizers (in which \( S \) is realized by many different kinds of materials) into homogeneous material realizers (in which the stuff realizing \( S \) is the same throughout). Either of the jades is a good example of the latter. But steel corkscrews can have silver engravings, and computers contain a great deal
besides silicon. Most entity kinds in our world have heterogeneous material realizers, and it is rare that the discovery of heterogeneous material realization leads us to split realized entity kinds. Precious metals and gems are items for which homogeneity frequently counts, but there is little temptation to split *Aplysia*, pyramidal cells, or even water into more than one distinct kind simply because we learn that they are composed of heterogeneous substances. The idea that these kinds would split does not arise naturally from consideration of theory building in science.

Learning how entities are materially realized is often an important opening stage of science. Cell fractionation and centrifugation were essential to the development of biochemistry and molecular biology, for example, precisely because they allowed investigators to discover cellular constituents. But when attention shifts from describing to explaining, entities and their material realizers are not the primary focus. Instead, attention shifts to the properties or activities as realized kinds and to component parts, their properties, and their activities as realizers.

Realizers for properties and activities lie along a continuum between aggregative and more mechanistic cases. *Aggregate realization* is realization by constituents where the property of the whole is a simple sum of the properties of the parts (see Wimsatt, 1997). The mass of a pile of sand is realized aggregatively by the masses of the individual grains (other units are equally useful). In aggregate cases, we do not typically split kinds when we split lower-level realizers. If we found that our pile of sand were made up of sands of ten different grains (coarse to fine), we would not be tempted to say that there are ten masses (any more than we would be tempted to say that there are right and left masses or bottom and top masses); and given that there are arbitrarily many ways to divide up the total mass into constituent units, there would be arbitrarily many ways of dividing kinds of mass. Note, though, that even in cases of aggregate realization, realizers include only some of the properties of the constituents: those that are explanatorily relevant to the realized property or activity. It is the mass of the individual grains (and not their color, melting point, or texture) that explains the mass of the whole pile.

In cases of *mechanistic realization*, the burden of realization is borne by some constituents (working parts or components) more than others, and organization among the components figures increasingly in our description of the realizer. The realized kind might be some activity, such as sensitization in *Aplysia*, a pyramidal cell’s firing, or a computer program running. Realizing mechanisms are composed of the working parts of the mechanism (such as amino acids, neurons, and filters) and their activities, the things that these entities do (their binding, firing, filtering). These components are organized such that they produce the behavior of the mechanism as a whole (Machamer et al. 2000). This organization may be
spatial (locations, orientations, shapes) or temporal (order, rate, duration) or active (who acts and interacts with whom when) (cf. Bechtel and Richardson 1993; Glennan 1996; Thagard 2000; Cummins 1975). There are no mechanisms simpliciter; all mechanisms are mechanisms of something, and it is by reference to that something that the relevance of components is established. Relevance thus has a central place for mechanistic realizers as well as for property realizers.

In general, realized entity kinds track neither kinds of realizing properties nor kinds of realizing mechanisms. As noted above, entity kinds are typically characterized in terms of clusters of co-occurring properties and our best evidence for the existence of an entity comes from the multiple independent lines of evidence for detecting and manipulating it via different of those properties. And entities don’t track realizing mechanisms for exactly the same reason: Entities (especially complex entities) tend to be associated with myriad lower-level mechanisms. There is no mechanism of the *Aplysia*, nor a mechanism of the cell; rather, they are composed of innumerable mechanisms for each of the many things that *Aplysia* and cells do. Explanatory relevance is central to our notion of realization for properties and activities (Lepore and Loewer 1989 and Kim 1993 include explanatory relevance in their account of realization), but not to realization for entities. We do not seek explanations for *Aplysia*, the cell, or computers. We do not ask “How the *Aplysia*?” or “Why the pyramidal cell?” but rather, “How do *Aplysia* sensitize to repeated tail shocks?” and “Why do pyramidal cells have their characteristic dendritic arborization?”

This distinction provides some basis for distinguishing cases where NDR holds from those where it does not, one that we might apply to avoid Ramachandran’s conclusion without jeopardizing the memory case. The different kinds of memory (declarative, procedural, sensitization, priming, etc.) are different kinds of activities. The self (as typically conceived) is an entity. If we take complex composite objects as our kinds, we are almost always going to find them to be realized by several independent kinds of constituents, their properties are almost always going to be realized by several independent sets of properties, and their activities

---

1. In some cases, a constitutive realizer explains why the realized is the kind of thing that it is. This thing’s being water is explained by its being composed of H2O molecules, this rock’s being jade is explained by its being composed of a silicate of magnesium and calcium. Being water and being jade are properties and should be handled as such.
are going to be explained by different and in many cases nonoverlapping mechanisms. Entity kinds do not typically track kinds of constituent substances, kinds of constituent properties, or kinds of constituent mechanisms.

In clear cases of dissociation, we are able to identify two or more distinct sets of relevant entities, properties, and/or activities, one set for each of two or more independently detectable realized kinds. The parts, properties, and activities that we appeal to in explaining procedural memory are not all of them relevant to explaining declarative memory and vice versa. But the realizers of mice and *Aplysia* are localized sums of materials, properties, and activities, and such sums can be formulated without special reference to the relevance of those constituents to the realized kind.

The self is no exception. Like most complex entities, selves are complexes of diverse capacities that are only loosely associated with one another. Selves are conceived as centers of conscious awareness, as agents of actions, as authors of narratives, as seats of autobiographical memory, as things that can refer to themselves, and as things that can recognize themselves in mirrors. They could not be all of these things without a variety of different capacities, many of which can be explicated by mechanisms that are independent. Think of the impact on the self suffered by people with brain damage. When H. M. lost his declarative memories, he lost one of these capacities, but by no means lost his self. (This is so even on a relatively simple formulation of Locke’s memory criterion, given that H. M. maintains memory ties to himself prior to the surgery.) Memory is but one of the psychological capacities contributing to a self. Ramachandran’s patient DF has lost conscious object recognition and maintained visually guided behaviors. The corpus of contemporary neuroscience is replete with tragic cases of people who, by virtue of having suffered damage to parts of their brains, have lost the ability to engage in activities that lie very close to the heart of what we call “the self.” But in truth, normal selves, prototypically, are (or, better, *do*) all of these things.

6. Conclusion. The problem of dissociable realization is to come up with a principled way to restrict NDR to only those cases in which it is truly appropriate. Here I have identified a range of cases in which NDR arguably fails. These different cases correspond to different varieties of realization at work in contemporary cognitive neuroscience. The metaphysics of realization ought to help us sort out the problems that arise in the process of refining our taxonomies of kinds. To play such a role, the metaphysics ought to be sensitive to different varieties of realization encountered in the process of building theories and ought to recognize the possibility that those different varieties of realization may carry vastly different implications for our taxonomy of kinds.
REFERENCES


