

1-1-2007

# Neuroeconomics, Philosophy of Mind, and Identity

John B. Davis

Marquette University, [john.davis@marquette.edu](mailto:john.davis@marquette.edu)

---

Published version. "Neuroeconomics, Philosophy of Mind, and Identity," in *Economics and the Mind*. Eds. Barbara Montero and Mark D. White. New York: Taylor and Francis (Routledge), 2007: 58-74. [Publisher Link](#). © 2007 Taylor & Francis (Routledge). Used with permission.

## 4 Behavioral economics, neuroeconomics, and identity

*John B. Davis*<sup>1</sup>

### 1 Introduction

Behavioral economics originated largely as a critique of neoclassical thinking about rationality and individual preferences. Largely overlooked by the proponents of this critique is the fact that preferences constitute the basis on which the personal identity of the individual may be understood in neoclassical theory, and thus that the behavioral critique has implications for how we might think about personal identity in economics. Moreover, since the implicit account of personal identity in neoclassical theory in terms of preferences is circular and empty (Davis 2003), and since some account of personal identity is ultimately necessary to be able to talk about individuals in economics, also overlooked is the question: Does behavioral economics, in addition to what it offers for thinking about rationality and choice, offer a secure foundation for thinking about personal identity to fill the gap left by neoclassical theory? Alternatively, might it rather be the case that behavioral economics makes no improvement on the neoclassical account of individual identity, and thus follows neoclassical theory in begging the question of what makes individuals individual in economics? In light of these questions, it is interesting not only that cognitive psychology, from which behavioral economics is drawn, possesses a widely shared view of what constitutes personal identity, but that this view involves a reinterpretation of the very same philosophical foundations that underlie the unsuccessful neoclassical view of identity, namely John Locke's understanding of identity in terms of memory. Can Locke's memory view of identity, then, be taken up and reinterpreted in behavioral economics in such a way as to escape the problems that arise in connection with neoclassical theory?

Here I first argue that cognitive psychology's adaptation of Locke's view offers behavioral economics an escape from one difficulty inherent in the neoclassical account, but that as a purely cognitive theory it still leaves unresolved another difficulty as an account of identity. However, just as cognitive psychology has employed neuropsychology as one route of development, so an increasingly influential route of development for behavioral economics is neuroeconomics, understood as the combination of neuroscience and behav-

ioral theory (cf. Camerer *et al.* 2005). Thus, second, I argue that the relatively recent interest on the part of behavioralists and experimentalists in neuroeconomics represents a possible further step toward developing an alternative account of personal identity. This further step, moreover, goes to the heart of the neuroeconomics program, and thus evaluating it as a possible step in developing an account of personal identity provides insights both into the status of individuals in a revised preference approach to choice and rationality and also into the prospects for developing neuropsychological foundations for economics.

Section 2 briefly summarizes the emergence of behavioral economics as a set of critiques of neoclassical thinking about preferences and rationality, emphasizing what has been termed alternative or nonconventional approaches within non-expected utility theory, and then comments on how this development bears on thinking about individual identity in terms of preferences. Section 3 explains how personal identity is understood in cognitive psychology on the self-as-memory view, and distinguishes this approach from Locke's account of identity in terms of memory. Section 4 looks at how behavioral and experimental economics have inadvertently come upon the self-as-memory conception of the individual, reviews Vernon Smith's ideas on the subject as a particularly clear expression of this thinking, and argues that this conception still leaves one key problem unaddressed. Section 5 turns to neuroeconomics' conception of the individual as a potential further step in the self-as-memory view, considers what it might assume about the relationship between mind and brain, and then compares this view to a similar view advanced by Derek Parfit in his analysis of personal identity. Section 6 briefly introduces social psychology as a framework for additionally enlarging the neuroeconomics conception of the identity of the individual as applied to economics and law.

## 2 The behavioral economics critique of preferences

Though behavioral economics involves a wide and diverse set of critiques of the descriptive and normative adequacy of standard expected utility theory, over time these critiques have reflected less and less optimism as to whether standard expected utility theory can be retained as a general framework for understanding rationality and choice. In early contributions – for example Maurice Allais's (1953) discovery of common consequence and common ratio effects at odds with expected utility theory's independence axiom, or the evidence of preference reversals showing how orderings can depend on preference elicitation procedures (Lichtenstein and Slovic 1971) – the emphasis generally rested on seeing violations of the standard framework as anomalies, and then on revising the standard framework so as to be able to accommodate them in a more general theory. Thus, according to Chris Starmer, as a result of the accumulation of considerable evidence contradicting the independence axiom in particular:

a wave of theories designed to explain the evidence began to emerge at the end of the 1970s. Most of these theories have the following features in common: (i) preferences are represented by some function  $V(\cdot)$  defined over individual prospects; (ii) the function satisfies ordering and continuity; and (iii) while  $V(\cdot)$  is designed to permit observed violations of the independence axiom, the principle of monotonicity is retained.

(Starmer 2000: 337)

Starmer calls “theories with these properties *conventional theories*,” and states that the “general spirit of the approach is to seek ‘well behaved’ theories of preference preserving monotonicity that are nonetheless consistent with observed violations of independence” (Starmer 2000: 337–38). In order to deal with the disconfirming empirical evidence, then, in time-honored instrumentalist fashion the conventional approach “models choice as preference maximization and assumes that agents behave *as if* optimizing some underlying preference function . . . [though] there is no presupposition that the model corresponds with any of the mental activities actually involved in making choices” (Starmer 2000: 349).

Others, however, saw abandoning descriptive adequacy as inconsistent with the general goal of providing more realistic explanations of choice under risk, and by the end of the decade what Starmer calls a nonconventional approach had emerged that proposed alternatives to standard theory rather than attempting to revise it. The most influential of these alternative theories was prospect theory, developed by Daniel Kahneman and Amos Tversky in what became one of the most highly cited articles in *Econometrica* (Kahneman and Tversky 1979). In general, nonconventional theories depart from the traditional framework by adding a procedural element to choice behavior in supposing that individuals rely on decision heuristics or rules sensitive to context to frame their choices.<sup>2</sup> Thus in prospect theory choice is a two-phase process, with prospects “edited” in the first phase using different decision heuristics, and choices then made in the second phase from a restricted or reformulated class of prospects.

This two-phase process reflects the distinction made in the psychology subfield of behavioral decision research between investigation of judgment processes, which people use to estimate probabilities, and choice processes, which people use to select among actions given their judgments (Camerer and Loewenstein 2003). The two-phase analysis makes it possible to look at gains and losses relative to reference points and ultimately introduce well-observed phenomena at odds with standard framework predictions such as diminishing sensitivity and loss aversion (Tversky and Kahneman 1991). More generally, prospect theory opens the door to the investigation of a variety of descriptive and procedure invariance failures that cast doubt on the traditional idea that individuals possess stable and coherent preferences. Taken as a whole, then, the “one feature common” to all nonconventional theories that distinguishes them from those approaches meant to preserve a

revised standard expected utility framework is that “none of them can be reduced to, or expressed purely in terms of, a single preference function  $V(\cdot)$ ” (Starmer 2000: 339).

Thus, what appears to have occurred over the course of several decades in behavioral research is that the search for an alternative theory of choice under risk has gradually become an increasingly important if not the main objective of researchers in place of the original goal of repairing standard expected utility theory. One reflection of this is that researchers refer less to experimental results as violations and anomalies, and now consider the possibility that standard expected utility theory is more likely a special case within a more general alternative theory rather than a general theory itself. However, more may be involved than simply a change in focus if, as Starmer puts it, an increasing number of behavioral researchers are willing to assume that individual choice under risk cannot be explained in terms of a single preference function, since it is not clear that approaches which abandon this assumption fall into the same class as those that maintain it. Indeed, since having a single preference function is the way in which individuals are defined in neoclassical theory, supposing that choice cannot be explained in such terms raises the issue of whether behavioral economics (at least the nonconventional approaches) fundamentally departs from the neoclassical theory of the individual, and thus constitutes a significantly different approach to understanding individuals.

This deeper issue has been closer to the surface than it may initially seem in the form of a debate between behavioralists and experimentalists over the character of preferences consequent upon assuming the two-phase analysis of choice. Thus, on the one hand, those on the nonconventionalist side suppose that preferences are not the indifference curves of textbooks, but are highly malleable and dependent on the context in which they are elicited (Camerer and Loewenstein 2003), so that choice involves processes whereby individuals effectively “construct” their preferences (Payne *et al.* 1992; Slovic 1995). On the other hand, those likely to be conventionalists hold the “discovered preference hypothesis,” a term coined by Charles Plott (1996), which assumes that individuals have coherent and stable preferences, though they are not necessarily always revealed in their decisions, but can be “discovered” to underlie their apparent preferences after individuals engage in information gathering, deliberation, and trial-and-error learning. Proponents of the “discovered preference hypothesis,” we might thus say, depart from the instrumentalist “as if” view in saying that individuals have a single preference function  $V(\cdot)$  but that it is descriptively available only under the correct experimental procedures. While this shifts the debate away from the merits of instrumentalist reasoning regarding individuals to a consideration of the experimental controls needed to avoid confounding effects associated with laboratory subjects failing to act consistently with their underlying preferences, there are good reasons to suppose that it does not succeed in rescuing the conventionalist view (cf. Cubitt *et al.* 2001).

Where, then, does this leave thinking about the individual? If we are to assess behavioral economics as a whole, what we find is a general retreat from the neoclassical conception in the case of nonconventionalist approaches, together with a set of defenses of the traditional single preference function  $V(\cdot)$  basis for the individual in the case of conventionalist approaches that has gained an at best limited success. In neoclassical theory the traditional axiomatic treatment of preferences permits assigning the individual a single utility function, which allows for an account of the personal identity of the individual (albeit faulty) in terms of the notion of "own" preferences. However, this prospective foundation for identity ceases to be even an issue if it is not even possible to say that individuals possess a single preference function  $V(\cdot)$  in the first place. The important question, it thus seems, is what account of the individual and personal identity might be put forward by the nonconventionalists, who have not been reluctant to entirely cut the umbilical cord to the traditional neoclassical basis for understanding individuals in terms of preferences. As no such alternative account appears yet to have been fully worked out, the following section turns to the resources naturally available for one in the treatment of the individual and personal identity developed by cognitive psychologists.

### **3 Cognitive psychology and personal identity in terms of memory**

Cognitive psychologists investigate individuals apart from their interactions with other individuals, focus on their mental representations, and regard individuals' mental representations as being either perception-based or meaning-based. The self is understood in terms of two kinds of mental representations which individuals can have of themselves: perception-based self-images and meaning-based self-concepts (Kihlstrom and Klein 1994). Since individuals have memories of their self-images and self-concepts, over time the self may be identified with the individual's collection of remembered self-images and self-concepts, and accordingly cognitive psychologists have explained the self in terms of memory, and characterized the self as one's memory for oneself (Klein 2001).

The well-known antecedent for the view that identity is based on memory is Locke's definition of personal identity as continuity in one's own memories (Locke 1975 [1694]). Philosophers generally regard Locke's definition as an inadequate explanation of personal identity on the grounds of its being circular (cf. Perry 1975). The self cannot be explained in terms of its remembering its *own* memories since this presupposes that which is to be explained. Cognitive psychologists, however, are not interested in how individuals might be seen to construct their own personal identities through what they remember about what they believe to be their own past experiences, but are, rather, interested in how third-party social scientists might attribute identity to individuals in terms of observed memory reports expressed by those individuals, irrespective of whether those reports accurately recall individuals'



past experiences. In effect, cognitive psychologists make use of an alternative concept of memory essentially like the one developed by philosophers to overcome Locke's circularity problem, namely quasi-memory or q-memory (Parfit 1984), that only requires that an individual's memory seem to that individual to have been his or her memory, whether or not it in fact was.<sup>3</sup> Individuals are then identified in terms of their real memories and perhaps "false" memories, as observed by cognitive psychologists. This strategy escapes Locke's circularity problem, and provides a working framework in which to develop a more detailed analysis of the different forms of memory as they pertain to the individual's identity.

The broad distinction between two kinds of mental representations – perception-based and meaning-based – is first associated with two forms of memory that the individual exhibits: (1) episodic memory and (2) semantic memory (Tulving 1983). Within these two forms of memory there are memories that then relate to the self and memories that do not. Episodic memory is experiential memory, and may be either autobiographical, involving an individual's memories of experienced self-images, or non-autobiographical, involving an individual's memories of experiences that included no awareness of oneself. For example, an autobiographical episodic memory is remembering oneself seeing a friend, while a non-autobiographical episodic memory is just remembering seeing a friend. Semantic memory is context-free memory that in itself makes no reference to an experience.<sup>4</sup> As with episodic memories, there are semantic memories that relate to the self, and thus involve an individual's self-concepts, and semantic memories that do not relate to self. For example, a semantic memory that relates to the self is remembering that one was born in a certain year, while a semantic memory that does not relate to oneself is remembering that Africa is a continent. Generally, then, cognitive psychology's self-as-memory view encompasses both autobiographical episodic memories that involve mental representations in the form of self-images, and semantic memories that relate to the self involving mental representations in the form of self-concepts.<sup>5</sup>

An important assumption in cognitive psychology's general theory of memory is that both episodic and semantic memory take the form of declarative knowledge ("knowledge that"), and can always be represented by individuals in propositional form. Though this is not immediately obvious with respect to episodic memory, the standard argument is that any episodic memory must be representable in propositional form on the grounds that one cannot remember an experience unless one can remember *that* one has had that experience. At the same time, in contrast to memories that take the form of declarative knowledge, individuals also have memories that take the form of procedural knowledge ("knowledge how") that in itself is not represented propositionally (Anderson 1976). Further, whereas episodic and semantic memory can exhibit the self, procedural knowledge *per se* is thought not to include references to the self, as, for example, when one remembers that the way to operate a particular device is to follow certain instructions.

Nonetheless, for essentially the same reasons that episodic memory is representable in propositional form, so procedural knowledge is thought to be also. That is, one cannot remember how to do something unless, at least in principle, one could say *that* such and such was how to do something. And when individuals exercise memories regarding procedural knowledge they also remember *that* they have certain skills and capabilities. Cognitive psychologists have characterized these cases as involving a kind of "meta-knowledge," which itself is declarative or propositional in form, but whose object is procedural rather than declarative knowledge, as, for example, when one remembers one knows how to drive a car (Bandura 1977).

These latter points gain greater meaning when their general purpose is made clear. Thus, by including meta-knowledge memories about one's own skills and capabilities along with episodic autobiographical memories and semantic memories involving the self as kinds of propositional declarative knowledge, *all* forms of memory somehow involving the self can be treated as propositional declarative knowledge. Or, as it has been influentially put, when we think of the self as memory, we think of the self as a knowledge structure (Kihlstrom and Klein 1994). This conclusion is important by virtue of what it excludes. Specifically, whatever cannot assume the form of propositional declarative knowledge cannot be an object of memory. For example, should primitive single-cell animals be conditioned to respond to certain stimuli, this conditioning cannot properly be termed remembering, because it could never take the form of propositional declarative knowledge (Bennett and Hacker 2003: 156). Or, should human individuals exhibit patterns of conditioned functioning about which they can never become conscious, these also cannot be properly termed remembering. This basic conclusion underlies the main general theory of memory in cognitive psychology, the generic associative network model known as the Adaptive Control of Thought (ACT) model (Anderson 1983), which organizes memory as a system of declarative knowledge, and which makes it possible to explain the self as a set of such memories within this framework.

Within the general ACT framework, not surprisingly, there are competing models of the structure and organization of the overall memory system, and also, therefore, competing ACT models of the self as memory. The differences between these competing models are not important here.<sup>6</sup> Suffice it to say that empirical studies in cognitive psychology that employ the self-as-memory view attempt to support one or another model in experiments on well-identified phenomena such as self-reference effects (in which individuals engage in some sort of self-referent encoding task) or priming actions (in which memory paths are activated in connection with certain cues). Reference to these sorts of phenomena, of course, recalls research in behavioral economics in terms of framing heuristics, editing, and so on, that has played a role in nonconventionalist approaches to individual decision-making, and one might accordingly expect behavioral economists also to start from a self-as-memory view of the individual. But it appears that



behavioral (and experimental) economists have only begun to make use of the cognitive psychology approach to memory and individual identity inadvertently by way of other concerns. The following section thus turns to how elements of an account of personal identity in terms of memory show signs of emerging in behavioral and experimental economics in connection with efforts to understand recent experimental results inconsistent with standard economic assumptions.

#### 4 Personal identity in behavioral economics

A surprising though now widely accepted result of recent work in the economics laboratory is that, contrary to what economists have believed for over a century, individuals do not behave in an exclusively self-interested manner in their economic interaction with others, but in certain specific circumstances consistently deviate from self-interested behavior by emphasizing reciprocity in their interaction with others – for example in competitive markets with incomplete contracts, and also in situations involving public goods where there are strong material incentives to free ride but also opportunities to punish free riders (Fehr and Gächter 2000). One implication of this is that social context matters (or, more concretely, institutions), since it appears to determine the circumstances under which one observes self-interested behavior or reciprocity. A second, less frequently noticed implication, however, is that memory also matters, because behaving in a reciprocal way towards others requires that individuals remember their past interactions with them. Thus, economists' recent empirical discovery that reciprocity is fundamental to certain types of economic behavior has opened the door to making cognitive psychologists' research on memory a new concern of economics.

One good reflection of this new thinking can be found in the work of Vernon Smith. Though Smith is more an experimentalist than a behavioralist, he shares the latter's general view (one at odds with rational choice theory and conventionalist theories of individual decision-making) that individual behavior cannot be properly explained without reference to its human psychological foundations. Here I draw attention to his position in his "Experimental Methods in Economics," a contribution to the *Cognitive Science Encyclopedia* (2003), which as an encyclopedia contribution presumably represents views that Smith believes should be taken as authoritative in economics. Much of Smith's discussion focuses on two experimental games whose reciprocity results have been highly replicated in the literature, the ultimatum game and the trust game. In interpreting what these games show, Smith succinctly combines the two implications of reciprocity behavior noted earlier as follows: "context matters because all memory involves relationships and is associative" (V. Smith 2003: 1070). Here the term "associative" refers to the ACT model, a reference which is further supported by comments Smith makes, on priming effects and on the logic of how cues

trigger memories, that draw on Gazzaniga *et al.* (1998), a standard cognitive psychology source which adds elements of neuroscience to the investigation of memory via neuropsychological studies of brain-injured patients and brain-imaging studies of normal individuals. Or, more explicitly, Smith states – echoing his oft-stated principle that institutions matter – “context is important,” specifically because of “what is known about the autobiographical character of memory and the interaction between current and past experience in creating memory” (V. Smith 2003: 1072). Thus he argues that individuals’ behavior in ultimatum and trust games exhibits reciprocity because individuals operate in contexts that not only require they remember their past interactions with other players but also require that they autobiographically remember these past interactions as their own experiences. Individuals, that is, employ what is described in cognitive psychology as episodic autobiographical memory in their interactions with others, and accordingly operate with perception-based mental representations of themselves, or self-images.<sup>7</sup> Smith also asserts that the design of experiments provides players with “different instructional-procedural contexts” (V. Smith 2003: 1072). On the argument of Kihlstrom and Klein (1994), this would generate procedural knowledge for individuals, which, when it pertains to the self, could take the form of meta-knowledge which individuals could additionally use to state propositions to be remembered about themselves and games.

Thus Smith sees a cognitive psychology understanding of memory as central to understanding interactive individual behavior. But he also has a particular view of the origins and development of this behavior which goes beyond a purely behavioral framework, and which draws on neuropsychology and evolutionary psychology. Neuropsychologists treat the mind as a system of circuitry organized in interacting modules that are specialized for various functions, and evolutionary psychologists regard this mental circuitry as the product of millions of years of human evolution.<sup>8</sup> One such module, Smith believes, is the “cheater-defector module” for social exchange, which evolutionary psychologists see as adaptive to a skill in making judgments about who can and cannot be trusted among individuals with whom one interacts. In a laboratory ultimatum and trust games context, then, individuals employ this module when they exhibit reciprocity, so that the reliance on memory in social interaction has its basis in the evolution of human biology and psychology. Further, in a passage in an earlier draft version of his Cognitive Science Encyclopedia contribution – then titled “Experimental Methods in (Neuro) Economics” – which was omitted from the final published version, Smith speculatively suggested how one might connect this larger framework and its emphasis on memory to a concept of the self: “personal identity is defined by some combination of inherited mental characteristics and our developmental experience” (V. Smith n.d.: 10). We might interpret this as saying that having a personal identity is the result of an evolutionary process that has reinforced the operation of our

mental cheater-defector module by the development of enhanced memory processing. Thus while reconceptualizing the self was not Smith's original entry point – rather it was to explain experimental evidence regarding reciprocity at odds with the self-interest postulate – his effort to explain these results nonetheless leads to an understanding of the self which departs from the traditional neoclassical conception that treats individuals as uninfluenced by context, and which integrates core cognitive psychology concerns in this regard with respect to the role of memory.

Whether Smith's views will be widely adopted by other experimentalists and behavioralists remains to be seen, but his picture, together with the more fully developed account to be found in cognitive psychology, points us toward an alternative behavioral conception of the individual. But applying this understanding of identity to the individual in economics still leaves one essential aspect of individuals unexplained: namely their independence from one another. Cognitive psychology puts this issue aside, because it by assumption investigates the individual apart from other individuals to focus on individuals' mental representations *per se*. But mainstream economics concerns individuals who interact with one another in markets, and thus its conception of the individual needs to include not only an account of what individuals are, but also an explanation of how they are distinct from one another. In this regard, cognitive psychology may be of little help if its self-as-memory view relies on quasi-memories which may be both the individual's memories and those of others the individual nonetheless believes to be his or her own. Indeed, cognitive psychologists have also developed "connectionist" models of social memory in which non-experiential semantic memories are formed and shared across individuals, thus blurring the lines between one individual and the next, at least with respect to this form of memory (E. Smith *et al.* 1999). Thus cognitive psychology's view of the self does not seem to provide a sufficient basis for explaining individual identity in economics, and something additional is needed if behavioral economics is to offer an account of the individual to fill the gap left by neoclassicism. In the following section I argue that those who see neuroeconomics as a logical extension of behavioral economics may see this additional needed element in the existence of the human body.

## 5 Neuroeconomics and identity

Neuroeconomics, understood as the combination of neuroscience and behavioral economics (cf. Camerer *et al.* 2005), represents a strategy for grounding the behavioral regularities that cognitive science has established in brain structures. That is, the regularities in behavior we observe are caused by the ways in which the brain works. While this general idea has long been entertained, the recent development of brain-imaging technologies has now made it possible to investigate the idea more systematically. Thus neuroscientific research examines how different areas of the brain are

activated when an individual is active in different ways. Since we may observe individuals' memory reports just as we can observe other types of behavior, neuroscientists argue that in principle we should be able to link memory behavior to particular areas of the brain. This, then, would locate all an individual's memories in the individual whatever their origin, whether or not they are autobiographical, and irrespective of whether social memory models are correct. Then individuals would not be identified strictly in terms of their memories but rather in terms of their "embodied" memories.<sup>9</sup> What, then, does this further conception of the individual involve?

First, the relation of psychological states to brain states is fundamental to both the neuroeconomics conception of the individual and indeed to neuroeconomics itself. There are different possible views here, and I begin with the most radical in order to dismiss it as not reflective of behavioral and neuroeconomics thinking as it is currently constituted. This radical view is a form of reductionism termed eliminative materialism or eliminativism, particularly as developed by Steven Stich (1983) and Patricia Churchland (1986). Actually, eliminative materialism is not reductionist in the sense of reducing or translating all psychological predicates into brain predicates, which proponents accept cannot be done, but rather reductionist in the sense that it aims to remove all psychological predicates and explanations from science on the grounds that such "folk psychology" notions as belief, desire, feeling, pain, memory, and so on are fictitious entities that have no place in science, much as was previously learned about such now-abandoned notions as phlogiston, ether, and so on.

But whatever the merits – or demerits (cf. Hacker 2001) – of eliminative materialism, it does not appear to reflect the goals of behavioral economists, who see neuroeconomics as an opportunity to reform the vocabulary of psychology used in economics rather than an occasion for abandoning such vocabulary altogether. Thus, even on the professedly "radical" interpretation of the neuroeconomics research program (Camerer *et al.* 2005), what is recommended is a replacement of the traditional neoclassical constructs of preferences and beliefs by such neuroeconomics constructs as affective versus cognitive processing and automatic versus controlled processing. These latter constructs may not be traditional folk psychology concepts, but neither are they concepts that may be explained purely in terms of neural functioning. Rather they constitute a sort of hybrid kind of concept whose purpose is to reconfigure the descriptions of psychological activity economists employ in such a way as to more closely reflect neural processing, and in this respect simply reflect recurrent efforts in the history of the field of psychology to re-specify and refine the field's categories and concepts. Indeed, neuroeconomic explanations of behaviors of particular interest in economics, such as loss aversion or hyperbolic discounting, remain well within the orbit of these new behavioral explanations. Thus neuroeconomics is clearly not a reductionist program in the strong eliminativist sense.

Nor indeed, moreover, does neuroeconomics appear reductionist in any significant sense. While some proponents may be carelessly tempted to suggest that psychological states can be translated into brain states, the program as a whole seems more committed to showing that brain states underlie psychological states, so as to be able to provide a more secure grounding to evidence regarding psychological regularities. Thus I suggest that the standard view among proponents of neuroeconomics is that psychological states are supervenient on brain states, meaning that psychological states depend on brain states but are not reducible to them, just as is widely argued in the case of other science–science relations such as biology and chemistry. Alternatively, psychological states are emergent upon brain states in the sense that they exhibit characteristics that cannot be ascribed to their associated brain states. That is, as in many other fields that have developed at the point of intersection between existing sciences, neuroeconomics as the combination of economics and neuroscience seems destined to adopt the widely held view of the relative autonomy of related sciences.<sup>10</sup>

Second, then, at issue is whether the body as the location of memories secures independence and personal identity for individuals. Here the challenge is more significant, though I will argue that what is at stake in this case is how one chooses to proceed in further expanding this conception rather than whether that conception fails altogether. Indeed, the question of whether the body secures personal identity has been extensively debated by philosophers in connection with a series of provocative thought experiments concerning brain transplants (one person's brain in another person's body), fission or split-brain transplant cases (one brain divided and transplanted into two bodies), brain fusion cases (distinct brains combined in one body), and teletransporter cases (bodies destroyed and identically rematerialized elsewhere).<sup>11</sup> All of these cases challenge the basic idea that one individual is indeed one individual by imagining improbable but conceivable circumstances in which the person becomes two persons or the reverse. The effect of their discussion on philosophers, it seems fair to say, has been to lead many to conclude that no definitive view of personal identity is likely possible. Perhaps most representative of this consensus is Derek Parfit (1984), whose own view of the individual – a brain-based psychological continuity theory – fairly closely matches the behavioral economics–neuroeconomics conception of the individual. Thus here I briefly describe his view in order to comment on the status of that conception.

Parfit's teletransporter cases are designed to violate identity while leaving in its place a weaker form of connectedness for individuals across change.<sup>12</sup> Were it possible for an individual to be destroyed and a perfect replica of that individual created elsewhere, our usual intuition is that the replica is not us – identity does not hold – but that we nonetheless have more interest in that replica of ourselves than in an entirely different individual. Parfit terms our replicas our *descendent selves*, and argues that in the absence of personal survival or personal identity proper we still care about the survival



of our descendent selves. In effect, Parfit's approach is a second-best-type strategy in that he gives up personal identity to place a lesser substitute in its place. On the one hand, this reflects philosophers' general skepticism about whether an adequate account of personal identity is possible, while, on the other hand, it reflects increasingly realistic dilemmas modern medical science may soon encounter regarding human identity. These same issues, however, also arise for the neuroeconomics conception of the individual, since its addition of the brain (from neuroscience) to the self-as-memory account (from cognitive science) leaves it in the same situation Parfit and others describe, where changes in the bodily basis for being an independent individual inevitably raise questions regarding the continuing independence and identity of that individual. Thus the neuroeconomics conception, at least as currently developed, also arguably fails to account for personal identity.

This all may strike some as a rather unsatisfactory resolution of the identity issue on the grounds that the effect of Parfit's arguments is to essentially drop the issue of individuals having a personal identity. Yet from another perspective Parfit's arguments carry a more positive message, since they could also be interpreted to mean that the brain-based psychological continuity conception of the self is only incomplete – not necessarily mistaken – as an account of personal identity. That is, if we suppose, as seems reasonable, that some account of identity needs to be supplied for particular domains of social investigation such as economics or the law, but that we cannot guarantee personal identity in the abstract, we might go on to adopt assumptions specific to those domains that effectively “close” our account of identity, recognizing that any such account is relative to the purposes for which it is developed, and thus is more on the order of a contingent account of identity, not a solution to the “pure” problem of personal identity. In the brief final section, then, I sketch such a strategy for conceptions of the individual in economics, in order to indicate how the neuroeconomics conception of the individual might yet provide a partial basis for thinking about the individual in economics.

## **6 Social psychology's social interactionist perspective**

What Parfit's arguments challenge is the numerical identity assumption underlying personal identity, namely that an individual is one and the same person across change. Another way of capturing this idea of numerical identity is to say that individuals remain distinct and independent of one another across change – what may be understood as a matter of being able to individuate persons (Davis 2003). Consider, then, the different sorts of strategies that might be used to individuate persons. One could first reason in terms of properties of individuals *per se*, and attempt to show that certain properties successfully distinguish particular individuals from one another. This route is taken by neoclassicism that emphasizes individuals' subjectivity, and takes individuals to be distinct in virtue of each having their own individual



subjective states. The behavioral economics–neuroeconomics conception also takes this route when it makes individuals' memories necessarily their own in virtue of their location in independent brains. However, both accounts fail as means of individuating persons – the neoclassical conception in virtue of its circularity problems and the neuroeconomics conception in virtue of Parfit-type problems. Thus we need to consider an alternative strategy for establishing individuation.

In this case, rather than reason from the properties of individuals *per se*, we might ask whether the independence and distinctness of individuals is a function of individuals' interaction with one another. Arguably this is the general view in social psychology, which in contrast to cognitive psychology begins from the assumption that individuals' psychological states must be understood socially. On this view, roughly speaking, individuals occupy different positions, the occupation of which distinguishes them *vis-à-vis* one another. This is not the place to examine the adequacy of this strategy, and accordingly I restrict myself to linking this idea to the notion that identity conceptions may be “completed” relative to fields of investigation in which they operate. Taking economics and law, then, as social interactionist types of investigation, both may be said to invest agents with certain capacities – the capacity to choose and trade in economics and the capacity to act as a legal and moral agent in law. Such capacities can be said to distinguish individuals from one another as independent beings, because both economics and the law presuppose that these capacities are independently exercised in the ideal case. Supposing individuals have such capacities, then, does not solve the “pure” problem of personal identity. But were societies in fact to be organized according to economic and legal systems meant to ensure that individuals generally had such capacities, then individuals would effectively have personal identities understood in part in terms of those capacities and in part in terms of whatever substantive conceptions of individuals were employed (such as, for example, a brain-based psychological continuity theory). In this case, having a personal identity is a contingent matter reflecting how societies are organized, and not something that can be said to hold apart from any and all social-historical circumstances.

From this perspective, the emerging behavioral economics–neuroeconomics conception of the individual receives a mixed evaluation. On the one hand, since it draws on the first individuation strategy described earlier, it can ultimately no more provide a fully satisfactory account of personal identity than can Parfit. But, on the other hand, since the cognitive psychology foundations of this conception are not necessarily incompatible with the social interactionist perspective of social psychology, the door is still open to enlarging this conception in such a way as to provide an understanding of how individuals conceived in the first instance in terms of memories and brains might have personal identities in social settings. Further discussion of these strategies for a “social embedding” of the identity of the individual, however, must be postponed to another occasion.

## Notes

- 1 The author is grateful to Barbara Montero, Mark White, and the participants of the 2004 Amsterdam–Cachan History and Methodology of Economics Workshop for comments on a previous version of this chapter.
- 2 In this respect they recalled Herbert Simon's earlier, neglected approach. See Sent (2004) for a comparison of the origins and nature of the "old" and "new" behavioral economics.
- 3 Thus one might have been told about an experience one does not remember, and subsequently come to believe one remembers it.
- 4 Though one may remember the experience of learning a semantic memory, as a semantic memory it can be remembered independently of the experience of learning it.
- 5 For a substantially similar but nonetheless slightly different classification of forms of memory from the perspective of philosophy, see the pioneering discussion of Norman Malcolm (1963).
- 6 A key difference is whether a memory item is stored independently of other memory items, as in the independent storage model, or whether there are groupings and clusterings of memory items, as in hierarchical organization models (Kihlstrom and Klein 1994). Note that it is important to take the term "storage" metaphorically.
- 7 This is not to say that Smith thinks non-autobiographical episodic and semantic memory are not involved in games. Surely they are. He simply does not employ these additional categories in emphasizing players' reliance on experiential memory.
- 8 Smith refers to the often-cited Cosmides and Tooby (1992) for this argument.
- 9 A precursor view is that of philosopher Sydney Shoemaker, who thought Locke's memory argument for identity presupposed the body, so that the memory criterion for identity depends on bodily continuity (Shoemaker 1984).
- 10 See Dupr (2001) for a defense of this view.
- 11 Locke actually considered the mind transplant case in his prince–cobbler example. His memory-as-identity definition led him to conclude that the prince was still the prince with his mind in the body of the cobbler, and vice versa. The fission and fusion cases are due to Wiggins (1967); Parfit (1984) develops the teletransporter case.
- 12 Similar conclusions arguably follow from the fission and other cases.

## References

- Allais, M. (1953) "Le Comportement de l'homme rationnel devant le risque: Critique des postulats et axiomes de l'cole americaine," *Econometrica*, 21: 503–46.
- Anderson, J. (1976) *Language, Memory, and Thought*, Hillsdale, NJ: Erlbaum.
- (1983) *The Architecture of Cognition*, Hillsdale, NJ: Erlbaum.
- Bandura, A. (1977) "Self-Efficacy: Toward a Unifying Theory of Behavioral Change," *Psychological Review*, 84(2): 191–215.
- Bennett, M. and Hacker, P. (2003) *Philosophical Foundations of Neuroscience*, Oxford: Blackwell.
- Camerer, C. and Loewenstein, G. (2003) "Behavioral Economics: Past, Present, Future," in C. Camerer, G. Loewenstein, and M. Rabin (eds.) *Advances in Behavioral Economics*, Princeton, NJ: Princeton University Press, 3–51.
- Camerer, C., Loewenstein, G., and Prelec, D. (2005) "Neuroeconomics: How Neuroscience Can Inform Economics," *Journal of Economic Literature*, 43: 9–64.
- Churchland, P.S. (1986) *Neurophilosophy: Toward a Unified Science of the Mind/Brain*, Cambridge, MA: MIT Press.

- Cosmides, L. and Tooby, J. (1992) "Cognitive Adaptations for Social Exchange," in L. Cosmides and J. Tooby (eds.) *The Adapted Mind*, New York: Oxford University Press, 163–228.
- Cubitt, R., Starmer, C., and Sugden, R. (2001) "Discovered Preferences and the Experimental Evidence of Violations of Expected Utility Theory," *Journal of Economic Methodology*, 8: 385–414.
- Davis, J.B. (2003) *The Theory of the Individual in Economics*, London: Routledge.
- Dupr , J. (2001) *Human Nature and the Limits of Science*, Oxford: Oxford University Press.
- Fehr, E. and G chter, S. (2000). "Fairness and Retaliation: The Economics of Reciprocity," *Journal of Economic Perspectives*, 14(3): 159–81.
- Gazzaniga, M., Ivry, R., and Mangun, G. (1998) *Cognitive Neuroscience: The Biology of the Mind*, New York: Norton.
- Hacker, P. (2001) "Eliminative Materialism," in S. Schroeder (ed.) *Wittgenstein and Contemporary Philosophy of Mind*, London: Routledge, 60–84.
- Kahneman, D. and Tversky, A. (1979) "Prospect Theory: An Analysis of Decision under Risk," *Econometrica*, 47: 263–91.
- Kihlstrom, J. and Klein, S. (1994) "The Self as a Knowledge Structure," in R. Wyer and T. Srull (eds.) *Handbook of Social Cognition*, Vol. 1: *Basic Processes*, 2nd edn., Hillsdale, NJ: Erlbaum, 153–208.
- Klein, S. (2001) "A Self to Remember: A Cognitive Neuropsychological Perspective on How Self Creates Memory and Memory Creates Self," in C. Sedikides and M. Brewer (eds.) *Individual Self, Relational Self, and Collective Self*, Philadelphia: Psychology Press/Taylor and Francis, 25–46.
- Lichtenstein, S. and Slovic, P. (1971) "Reversals of Preference between Bids and Choices in Gambling Decisions," *Journal of Experimental Psychology*, 89: 46–55.
- Locke, J. (1975 [1694]) *An Essay Concerning Human Understanding*, ed. P.H. Nidditch, Oxford: Clarendon Press.
- Malcolm, N. (1963) "Three Forms of Memory," in *Knowledge and Certainty*, Englewood Cliffs, NJ: Prentice-Hall, 203–21.
- Parfit, D. (1984) *Reasons and Persons*, Oxford: Clarendon Press.
- Payne, J., Bettman, J. and Johnson, E. (1992) "Behavioral Decision Research: A Constructive Processing Perspective," *Annual Review of Psychology*, 43: 87–131.
- Perry, J. (ed.) (1975) *Personal Identity*, Berkeley: University of California Press.
- Plott, C. (1996) "Rational Individual Behaviour in Markets and Social Choice Processes: The Discovered Preference Hypothesis," in K. Arrow, E. Colombatto, M. Perlman, and C. Schmidt (eds.) *The Rational Foundations of Economic Behaviour*, Basingstoke: Macmillan, 225–50.
- Sent, E.-M. (2004) "Behavioral Economics: How Psychology Made Its (Limited) Way Back into Economics," *History of Political Economy*, 36: 737–62.
- Shoemaker, S. (1984) "Personal Identity: A Materialist Account," in S. Shoemaker and R. Swinburne (eds.) *Personal Identity*, Oxford: Blackwell, 67–132.
- Slovic, P. (1995) "The Construction of Preferences," *American Psychologist*, 50: 364–71.
- Smith, E., Coats, S., and Walling, D. (1999) "Overlapping Mental Representations of Self, In-Group, and Partner: Further Response Time Evidence in a Connectionist Model," *Personality and Social Psychology Bulletin*, 25: 873–82.
- Smith, V. (2003) "Experimental Methods in Economics," in L. Nadel (ed.) *Cognitive Science Encyclopedia*, London: Nature Publishing.
- (n.d.) "Experimental Methods in (Neuro) Economics," unpublished.
- Starmer, C. (2000) "Developments in Non-Expected Utility Theory: The Hunt for a Descriptive Theory of Choice under Risk," *Journal of Economic Literature*, 38: 332–82.

- Stich, S. (1983) *From Folk Psychology to Cognitive Science*, Cambridge, MA: MIT Press.
- Tulving, E. (1983) *Elements of Episodic Memory*, Oxford: Oxford University Press.
- Tversky, A. and Kahneman, D. (1991) "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *Quarterly Journal of Economics*, 106: 1039-61.
- Wiggins, D. (1967) *Identity and Spatio-Temporal Continuity*, Oxford: Blackwell.