Neuroeconomics: Constructing Identity

John B. Davis

Department of Economics
Marquette University
PO Box 1881
Milwaukee, WI 53201-1881 USA
(voice) 414-288-5438; (fax) 414-288-5757
john.davis@mu.edu

Amsterdam School of Economics
University of Amsterdam
Roetersstraat 11
1018 WB Amsterdam, The Netherlands
(voice) +31-020-525-4187; (fax) +31-020-525-4254
j.b.davis@uva.nl

September 14, 2010

Forthcoming: Journal of Economic Behavior and Organization.

Abstract: This paper asks whether neuroeconomics will make instrumental use of neuroscience to adjudicate existing disputes in economics or be more seriously informed by neuroscience in ways that might transform economics. The paper pursues the question by asking how neuroscience constructs an understanding of individuals as whole persons. The body of the paper is devoted to examining two approaches: Don Ross's neurocellular approach to neuroeconomics and Joseph Dumit's cultural anthropological science organization approach. The accounts are used to identify boundaries on single individual explanations. With that space Andy Clark's external scaffolding view and Nathaniel Wilcox's socially distributed cognition view are employed.

Keywords: neuroeconomics, behavioral economics, neurocellular economics, collective brainset, external scaffolding, socially distributed cognition

JEL codes: A12, B41

Neuroeconomics is unique as a research program in economics for a number of reasons. First, it relies on a technical scientific apparatus whose sophistication and complexity has no equivalent in any other research program in economics. Second, the science behind that technical apparatus, as well as its goals, is less understood by economists than the science drawn upon by any other research program in economics. Third, the neuroscience scientific community itself is more diverse, multi-sided, and interconnected in its network of fields and subfields, is more organized in its protocols and scientific practices, and has more points of contact with the non-science world than any other science community that other economics research programs draw upon. In short, in stark contrast to all the other recent new research programs in economics, neuroeconomics operates in a highly unfamiliar, highly advanced science environment of considerable magnitude and social influence. However, contrary to what this suggests, this does not imply that neuroeconomics is destined to fail as a research program. Neuroeconomics could still develop as a relatively independent and active research program in economics making essentially non-expert use of neuroscience's technical scientific apparatus for its own program of investigation. At the same time, for neuroeconomics to become a truly significant new research program in economics, it seems it needs to go deeper into neuroscience to investigate what it means to operate in terms of *its* rationales, and then somehow use these to reframe the questions economics asks. That is, if it is not to only pursue corroboration of economics' pre-given concerns, and constitute a genuinely 'new' research program, it needs to genuinely ask new questions, and one way in which this can be done is by looking through the lens of other sciences. And, in the case of neuroeconomics, clearly, the status and significance of neuroscience as a twenty-first century pioneer science offers a remarkable opportunity to do this.

In this paper, then, I ask how neuroscience might inform economics in a fundamental way with respect to an issue which seems central to deep neuroscience concerns and which offers a natural point of contact with economics, namely, what the individual is, or in language less familiar in economics, how we understand the identity of individuals. The individual, of course, has been a primary theoretical construct in economics for over a century, and the preference conception of the individual has been standard in economics for the last half century. But all of the new research programs in economics of the past several decades question the standard view, indeed in different and in some cases in far-reaching ways (Davis, 2009). Thus it is fair to say that the

question, 'what is the individual,' is on the agenda at the research frontier in economics. In neuroscience, the matter is even more salient. Neuroscientists put single individuals into the scanner, where they treat them there as collections of relatively independent neural structures whose coordination is theorized in different ways according to the goals of the research team involved and the kinds of phenomena being investigated. Neuroscience begins with the disaggregated rather than whole individual, and its multiplicity of its research concerns inevitably produces a multiplicity of theories regarding how individuals function as single persons. At the same time, given the organization of the neuroscience community, no one has any special interest in asking how this might be true in general. Indeed, that is likely to simply seem a poorly formulated question for most neuroscientists.

This paper, then, investigates this possible point of contact between neuroeconomics and neuroscience by examining and comparing two distinct points of entry regarding thinking about individuals: Don Ross's (2005) game theoretic neurocellular approach to neuroeconomics and Joseph Dumit's (2004) science system of organization approach to neuroscience. The two approaches employ significantly different underlying assumptions, yet can be shown to operate by means of strategies of analysis that mirror one another, and jointly map out a social-institutional space in which we seek to explain the behavior of single individuals. The first section of the paper begins with an assessment of the behavioral economics-based neuroeconomics view of the individual in neuroeconomic research, and then the second and third sections turn to the Ross and Dumit views about individuals in neuroeconomics and neuroscience respectively. The fourth section discusses the relationship between their different types of explanations, and argues for a combined explanatory strategy appropriate to the subject of individual economic behavior. The concluding section comments briefly on neuroeconomics as a new research program in economics.

i. Individuals in neuroeconomics

An influential view of neuroeconomics, promoted by Colin Camerer and others, is that neuroeconomics is an extension and development of behavioral economics which aims to secure additional new evidence from neuroscientific research for many of the conclusions reached by

psychologists and behavioral economists about the behavior of individual economic agents (e.g., Camerer, Loewenstein, and Prelec, 2005; Camerer, 2006, 2008). As Ross aptly puts it, on this view neuroeconomics is 'behavioural economics in the scanner' (Ross, 2008, p. 374). Ross opposes to this conception an understanding of neuroeconomics he terms 'neurocellular economics' which uses the modeling techniques and mathematics of economics – specifically optimization and equilibrium analysis – to represent the functioning of different parts of the brain without making any assumptions about how neurocellular processes are related to the individual as a whole. For Ross, whether in neuroeconomics or other areas of investigation, an agent ought to be defined in minimalist fashion as an optimizer in a narrowly defined reward space (Ross, 200., pp. 378-9). More fully, an agent is "any system that observes certain consistency conditions in behavior, such that it can be interpreted *as if* it is maximizing the value of a function that maps a system of preferences over commodity bundles onto the real numbers" (Ross, 2005, p. 245). Note that there is nothing in this definition that requires economic agents be identified as single individuals – or as Ross puts it, as "whole *people*" (Ross, 2008, p. 378). Indeed the objective functions that agents optimize, whether they be utility functions as in the passage above or production functions, can in principle be ascribed to all types of agents, including *sub*-personal agents – "*that* neuron, or *that* neurotransmitter system, or *that* quasi-modular circuit" (*Ibid*., p. 379). In a word, Ross rejects the idea that any sort of whole person 'anthropomorphism' is the necessary starting point neuroeconomics or indeed any other sort of investigation that concerns economic agency.

In contrast, neuroeconomics as 'behavioural economics in the scanner' unhesitatingly assumes that economic agents are single individuals. If we take Daniel Kahneman and Amos Tversky's prospect theory (Kahneman and Tversky, 1979) as the basic model from which most developments in the behavioral economics literature have been generated (cf. Heukelom, 2008), then the rationale for assuming this is that their individual value function V is meant to function as an analogue to the traditional individual economic agent utility function.[1] A neuroeconomics derived from behavioral economics thus works from the assumption that single individuals (suitably re-represented) are agents for an analysis of how the different regions of functioning in

[1] "As in utility theory, V is defined on prospects, while v is defined on outcomes" (Kahneman and Tversky, 1979 p. 276).

the brain work together to produce the behavior of the single individual economic agent. This of course begs the question whether people function as single individuals. It is entirely conceivable that they do not, and indeed much social science as well literature from the humanities paints a more complicated picture.[2] Thus we could equally reason that different structures in the brain effectively produce different selves for people, and that individuals' interaction with others and their wider environments is segmented according to which of these selves and which of these structures is involved. People as single individuals do not interact; the different selves of people interact. Thus that you place one body in the scanner at a time does not imply that the behavior investigated is that of a single agent. Indeed it is a species of the mereological fallacy (Bennett and Hacker, 2003, 68-107) – the view that the brain and the mind can be treated as equivalent – to suppose that having one body ensures one functions as a single individual.[3]

I suggest, then, that a behavioral economics-based neuroeconomics is prone to this error because it is motivated by the general strategy of introducing greater 'realism' into standard economics, and sees implementing this strategy to be a matter of operating with a revised atomistic individual conception. That is, its aim is to empirically upgrade the standard atomistic individual conception in economics by expanding its empirical foundations via psychology and neuroscience. But it can be argued that this underlying commitment to the single individual assumption is fundamentally at odds with the general nature of neuroscientific research. Neuroscientific research, it seems fair to say, is novel in that the experimental access to the brain's different structures and multiple sources of neural processing that scanning technologies allow makes it possible to go beyond indirect inferences about the brain's functioning based on observations of overall behavior. However this makes the integration and unity of the brain's many activities a central question. That is, once we disaggregate the brain's many activities, the question of their aggregation in relation to people's behavior is clearly on the agenda. Neuroscience, then, begins with the individuation problem, or the problem of determining how or under what conditions individuals might function as single beings (if they do), while a

[2] For example, in social psychology social identity theory is quite ambivalent on the subject (cf. Brewer, 2001; Brown, 2000).

[3] If one considers the four (Cartesian) possibilities that the two sets of self/body alternatives allow, [one self/multiple selves] and [one body/multiple bodies], 'behavioural economics in the scanner' operates in terms of [one self/one body] and neurocellular economics operates for Ross in terms of [multiple selves/one body]. A theory of 'remote prosthetics' might constitute a basis for the two other cases, but the discussion in this paper only concerns the first two cases.

behaviorally-based neuroeconomics based on a revised atomistic individualism eliminates this central neuroscience issue before it can even be raised. Like the standard atomistic individual conception, a revised one settles the individuation problem by simply assuming it away.

Let me explain this claim by further characterizing the atomistic individual conception (revised or not) as an 'internalist' type of conception of the individual. On my view, 'internalist' conceptions of the individual define single individuals strictly in terms of their own characteristics apart from any characteristics they may have in virtue of their relationships with other individuals (cf. Davis, 2008). But defining being a single individual strictly in terms of an individual's 'own' characteristics is circular, and thus provides no explanation of why we should individuate persons as single individuals. Compare, then, Ross's neurocellular economics approach to neuroeconomics. As we will see more fully in the next section, he asks how collections of sub-personal agents might coordinate with one another to function as single individuals while interacting with other collections of sub-personal agents attempting to coordinate with one another as other single individuals. That the individuation of the single individual consequently depends on interaction between different individuals (or between different sets of coordination games) makes Ross's conception of the individual what I term an 'externalist' rather than 'internalist' conception, since it defines single individuals in terms of their relationships to other individuals. This avoids the circularity problem, and also avoids assuming away neuroscience's task of explaining how the brain's many activities coordinate to produce single individual behavior.

But Ross's alternative strategy of explaining what goes on internally within the individual only in combination with what goes on externally for the individual vis-à-vis other individuals raises an interesting problem concerning how far this latter side of the story should reach. Though his basic model only requires that two collections of sub-personal agents interact in order to show how they can each be single individuals, he recognizes that human interaction is many-sided and influenced by the larger environment in which it occurs, and that this environment includes historically accumulated social structures and institutions. The question this raises, then, is how much of this wider social-institutional content needs to be incorporated into our explanations of interaction to say that collections of sub-personal agents can be individuated as single people in

terms of their interaction. Notice, however, that this rather open-ended 'how much' question immediately raises another question, or a 'how much' is 'too much' question. Thus we know that at some level of description of social structure our explanations cease to be fine-grained enough to refer to single individuals, and that individuals drop out of explanations that are chiefly concerned with social aggregates. Consequently we need to ask not only how much additional social-institutional content beyond interaction is needed to explain individuating interaction, but also how much of this content is *too* much to explain interaction because individuals drop out of the analysis.

Nathaniel Wilcox offers a framework which I believe helps us bring these questions into focus. The problem of individuating people as single individuals first arises out of the fact that people may be dissolved into collections of different sub-personal selves; this is what Wilcox calls the "*fission* of agency." The flip-side of this – the problem I characterize as the individuation problem (how collections of different sub-personal selves can be single individuals) – is for Wilcox the problem of how we explain "*fusions* of agency" (Wilcox, 2008, p. 527). Let us say, then, that this particular fission-fusion combination identifies the 'lower bound' for the analysis of single individuals, since they 'emerge' as single individuals – a fusion of the agency of their sub-personal selves – when we explain the conditions under which they can be individuated as single individuals. In parallel, however, there is also an 'upper bound' issue regarding how much of our social-institutional space should and can be incorporated in individuating single people. Some of this content does presumably need to be incorporated, and we rely on it to reinforce the fusion of single individual agency as explained by means of a social-institutional account of interaction. But too much social-institutional content moves our explanations beyond reference to individuals to produce a further fusion of agency in the form of social groups and others social aggregates into which individuals effectively 'dissolve.' At this upper bound, that is, concerns move to the emergence of social groups through a fusion of the agency of individuals. Both lower and upper bounds of single individual analysis, then, are two-sided, balancing fission and fusion issues. Between the two of them, there exists the social space in which we explain the behavior of single individuals as agents.

To begin to fill in this space, in what follows I look at the Ross and Dumit types of explanations of neuroscience, one addressing the relationship between individuals seen as multiple selves and single individuals – the lower bound of the analysis – and the second addressing the relationship between single individuals and social explanations that eliminate them – the upper bound of the analysis. In the next section I look more closely at Ross's dual-sided account of the fission of single individuals into their sub-personal agents and their fusion as interacting sets of sub-personal agents. Ross's view, culminating in an account of individuals' self-narratives, introduces a set of concepts that direct us to look upward into the social-institutional space that arguably frames these narratives. Then in the section following, to get a sense of the 'upper bound' of the analysis, I consider the vantage point of cultural anthropologist Joseph Dumit's examination of the social organization of science involved in neuroscience research. Here, I rather look downward from the neighborhood of the 'upper bound' where single individuals fuse in explanations that focus on the behavior of entire scientific communities. In the closing section of the paper I discuss the relation of the two strategies to one another in order to comment on the nature of the social-institutional space in which single individual explanations operate.

ii. Don Ross: Sculpting Individuals

Ross's neurocellular economics understanding of neuroeconomics follows in the line of thinking initiated by George Ainslie and Paul Glimcher that starts with individuals' multiple selves (Ainslie, 2001; Glimcher, 2003).[4] His conception of the individual (Ross, 2005, 2006, 2007; cf. Davis 2007), we saw, characterizes individuals as collections (akin to communities or countries) of optimizing sub-personal agents who interact in coordination games internal to the individual, and who also interact in coordination games external to the individual with other collections of sub-personal agents. Why begin with sub-personal agents? Ross takes the idea of an agent to be primitive and basic – an agent is "anything whose behavior is well modeled within the constraints of a small set of consistency axioms" (Ross, 2006, p. 247), and then argues that neuroscience gives us good reason to see individual neurons and other neural structures as

[4] The multiple selves view is sometimes associated with Paul Glimcher's neuroeconomics research, but Ainslie's view is criticized in Glimcher, Kable, and Louie (2007). Ross *et al*. (2008) defends Ainslie.

functioning as in this way as agents (Ross, 2006, p. 251).[5] The issue for him, then, is rather whether single individuals play a role in a neurocellular economics. Indeed, it not immediately clear from his starting point that we even need to individuate whole persons, since if an individual's sub-personal neural agents play coordination games with one another, it stands to reason that they could also play coordination games with other individuals' sub-personal neural agents (as suggested above), thus eliminating any need to talk about the person *per se*. Then the analysis would be no different than a generalized game theoretic view of the world in which agents play different games with different agents, and there are no agents made up of collections or communities of agents. But this is not Ross' view. Essentially this representation of the world is too 'flat' to capture the various structures he believes we see in it. Thus his game theoretic world is one in which certain collections of agents find it in their interest to successfully play coordination games with certain other collections of agents. Individuals' sub-personal neural agents, in particular, find it in their interest to prioritize coordination games with sub-personal agents internal to the same individual. The basis for this is largely evolutionary. These sub-personal neural agents share the environment of the individual human body, and their survival obviously depends upon the body's survival.

More fully, then, what I have referred to as additional social-institutional content is understood by Ross in terms of ways in which evolutionary pressures operate upon (single individual) collections of sub-personal neural agents. Individual sub-personal neural agents would not survive were the single individuals in which they operate not to survive. This however does not imply that the former are subsumed within the latter, or that fission gives way to fusion. Certain neural structures can have dominant influences on individual behavior at one point in time and others at other points in time. That is, the equilibrium solutions to individuals' internal coordination games not only have different possible values for different sub-personal neural agents, but also have different impacts on the functioning of the individual as a whole. The latter is then a factor influencing how the external coordination games that single individuals (as collections of sub-personal neural agents) play with one another work out. Accordingly we need

[5] Other elements in his characterization of what it means to be an agent are that agents cannot be 'money pumps' (or indeed other kinds of self-eliminating 'pumps' – such as, I suggest, a 'social identity pump'), and that agents exhibit servosystematicity or the control of entropy through negative feedback mechanisms (Ross, 2005, pp. 246-8).

to maintain an account of both the individual's disaggregation into multiple sub-personal selves and aggregation as a single person simultaneously. Indeed the interdependence between coordination games on these two levels is what generates variations in behavior upon which selectionist evolutionary forces operate.

Ross's evolutionary view, then, is set out more explicitly in terms of social dynamics, or evolving forms of social interaction, whose pressures 'sculpt' individuals into whole persons. Central to this argument is a special role that Ross assigns human language. We know that humans are different from other animals in terms of their particular capacities for language. An especially important dimension of this is individuals' capacity to produce self-narratives or discursive representations of themselves. But as language is a capacity exhibited at the level of the single individual, individuals' self-narratives are representations of the whole individual that on Ross's argument emerge simultaneously as a result of the games played between an individual's sub-personal selves and the games played between individuals as communities of sub-personal selves. The questions above about individuating the whole person, then, can be reformulated as follows: why should individuals produce single whole person-individuating self-narratives rather than just confused speech that is not individuating? Why not a collection of self-narratives each representing different competing sub-personal selves. Ross' answer is that environments for social interaction are complex and highly interdependent, and so to play games with others whole individuals need

> to *make* themselves predictable. They do this *so* they can play and resolve coordination games with others. (To be predictable to others, they must be predictable to themselves, and vice versa.) Then all of this is compounded by the fact that nature doesn't exactly partition games the way analysts do in game theory texts. A person can't keep the various games she simultaneously plays with different people in encapsulated silos, so a move in a game G_i with the stranger will *also* represent moves in other games $G_{k, \ldots, n}$ with more familiar partners – because these partners are watching, and will draw information relevant to $G_{k, \ldots, n}$ from what she does in G_i (*Ibid*., p. 250).

The interdependent nature of games between individuals, then, constrains the games internal to individuals between their sub-personal selves to produce self-narratives that are recognizable as those of whole persons. For Ross, that is, "social dynamics *are* logically and ontogenetically prior to individual selves, because selves are sculpted into being by social processes" (*Ibid*., p. 257).

Note that this argument does not reduce to the simple (and potentially empty) claim that evolutionary processes individuate single persons. Evolutionary pressures indeed operate, but that individuals respond to them by producing self-narratives assumes a particular type of reflexive activity that goes beyond the idea of a dynamic feedback principle operating from social interaction to the individual. The idea of dynamic feedback itself is sufficient to make Ross's analysis of the individual relational, since individuals do not exist as single beings except in interaction with others. But the reflexivity idea adds to this the claim that (game-theoretic) interaction within the individual and between individuals produces a representation of the whole individual, specifically a *self*-report, rather than simply a set of representations of aspects of the whole. Nothing in Ross' analysis of game-theoretic interaction between an individual's sub-personal selves, however, tells us why there should be this orientation. Games internal to an individual should just be about those games and their sub-personal players, and as there are many of these, they fall well short of a view of the whole. One is consequently suddenly be struck by the question, where does this self-concept in the form of a self-narrative come from? In my view Ross does not have an entirely clear answer to this question, but he nonetheless offers a framework for one that can be set out in outline: individuals develop individuating self-concepts, because of their interaction with others (the relational aspect of his view), because of evolutionary pressures on the interaction within themselves (the neuroeconomic aspect of his view), and because of the special properties of human language (what I term the social-institutional aspect of his view). The last is worked out in terms of Daniel Dennett's idea of an intentional stance (Dennett, 1987).

Dennett's intentional stance idea, or intentional stance functionalism, involves explaining the behavior of non-human entities using the concept of intentionality. An intention in the most basic sense is an orientation toward something, and its representation is always in the form of

language as in such expressions as 'believes,' 'thinks,' 'wants,' etc. Thus when we represent a non-human entity intentionally, we say, for example, that a thermostat 'perceives' a change in temperature, or a computer 'remembers' certain files. Dennett's goal in authorizing such expressions is not to undermine the language of intentionality as associated with human intentions, but rather to free it of its traditional association with so-called internal states of the human mind, and in effect 'externalize' its meaning by applying it to relations between things in the world. This arguably increases the power of attributions of intentions to all sorts of entities in the world, because rather than looking 'deeper and deeper' into some sort of mysterious mind to understand those intentions, we are compelled to draw on features of the environment and the world to explain the nature of those intentions – a process understood as 'triangulation' by Ross (Ross, 2005, pp. 49-50, 61). Thus the intentional stance, as simply a principle of orientation requiring expression in a distinct kind of language, can be taken by all sorts of entities towards all sorts of things.

A collection of sub-personal selves, accordingly, could conceivably take an intentional stance toward the individual as a whole just as they could than toward any other objects in their environment, and this 'stance' could be characterized as a self-concept. In effect, the equilibrium values of a successful coordination game could include an orientation on the individual as a whole just as they could include an orientation on some other object. The language of intentions, it should be emphasized, has syntactic and semantic properties. Their semantic character supports their breadth of reference, while their syntactic character structures this reference. We thus use language syntactically in subject-object relationships, but subjects, or selves, can semantically speaking also be objects in language. Within this framework, Ross can thus say that the interaction of sub-personal selves within individuals combined with the interaction between individuals could tend to sculpt out whole persons. The argument is only an outline, because it is not clear just how the individual-individual interaction would drive within-individual interaction to produce equilibrium self-concept type intentional stances towards the whole person on the part of the person's sub-personal selves. Ross's further thinking in this regard concerns the impact of evolutionary pressures on the individual self-narrative (cf. Dennett, 1991; Bruner, 1992, 2002).

A narrative, including self-narratives, is a discursive structure that takes the form of a story from a particular perspective, and as such must have consistency and integrity as a single account of how events transpire for a subject or number of subjects. That is, a story must hold together, and in the case of self-narratives must do so for the subject at the center of that narrative (cf. Eakin, 2007). Consequently, if a person's self-concept is seen in extended form as self-narrative, the whole person is individuated as a single being as long as that person's story retains its integrity. Of course saying that a story holds together is very subjective, but Ross has a way of avoiding this problem. His view is that social dynamics increasingly sculpt the person by narrowing the possibility space the individual's life occupies as the individual ages, so that the continually reduced space in which the individual's self-narrative takes place always defines the self, even if the 'story' never comes down to one simple final message (Ross, 2007). This convergence argument arguably allows for a kind of dynamic re-identification of the individual as a single being, because the narrative of the person at earlier points in time is always contained in the narrative of the person at later points in time.

The problem the view faces, however, is that of life reversals and discontinuities in a person's story. Ross' convergence view relies on treating the person's identity through time as a kind of continuity, but continuity itself cannot be assumed in his framework. Since social interaction influences individuals' formation of their self-concepts/self-narratives, large changes in it can contribute to significant changes and possible discontinuities in a person's life story. Ross seems required to say that a new person emerges in such circumstances, but might we still somehow argue that the same person is there but with a new life and story? Ross in fact has additional resources at his command that might assist in this regard that he arguably does not fully exploit. Thus he accepts Andy Clark's idea that individual cognition is supported by external scaffolds whereby people offload to constructed and natural structures in the world those cognitive functions they are not especially good at while retaining and specializing in those functions at which human beings excel (Clark, 1998).[6] From this perspective, we might argue that self-

[6] Clark's (anti-Cartesian) idea of external scaffolds was developed following the thinking of psychologist Lev Vygotsky, and sees human mental processing, whether involving cognitive or motor tasks, as relying on found and created features of the external world (Clark, 1998, pp. 45-7). Essentially we retain those activities at which we excel, offload those at which our external prosthetics excel, and use continually reconfigured combinations of the two to achieve our goals. Similarly Wilcox argues that we should see human cognition as being distributed across individuals and external cognitive artifacts in social systems of information processing "which implement *external*

narratives are not merely the product of individuals' interaction, but also reflect accumulated social cultural practices and inherited expectations regarding the form and nature of story-telling in the form of self-representation, thus functioning as external scaffolds to individuals' self-narratives. Then, though individuals' self-narratives may exhibit reversals and discontinuities, from this perspective they might nonetheless be regarded as the self-narratives of single individuals in virtue of having been produced by means of this external scaffolding.

This argument obviously only outlines a way in which Ross's argument might be extended to secure his relational single individual conception. It does so, moreover, by moving his argument upward from the lower bound of the range in which we speak about single individuals further into the social-institutional space than he takes it with his emphasis on the functional stance and self-narratives. From Clark's external scaffolding perspective, the question this raises is, what are the actual structures that operate in this social-institutional space that might play this scaffolding role? In the following section, then, to get some further sense of how we might begin to investigate such structures, I move to Dumit's point of entry via explanation of the multi-disciplinary organization of neuroscience. When we speak about the organization of science and the groupings it involves, of course, we emphasize social aggregates. Thus we will use Dumit's discussion to look downward from the upper bound of the range in which we speak about single individuals to see what this additionally tells about individuals in neuroscience as seen thus far from Ross's upward-looking entry point.

iii. Joseph Dumit: Picturing Personhood

Picturing Personhood is the title of Dumit's book on brain scans and individuals' biomedical identities (Dumit, 2004). Dumit is a cultural anthropologist, and though he writes from the perspective of science and technology studies about neuroscience in general rather than about neuroeconomics *per se*, the issues he addresses are no less relevant to neuroeconomics and its representation of the individual. Indeed much of his discussion focuses on laboratory experimentation, a central focus of neuroeconomics. A key difference from neuroeconomics,

---

information representations and algorithms," and reduce "*the importance of individual brains* to economically important information-processing tasks" (Wilcox, 2008, p. 524).

however, is that Dumit focuses on how the laboratory is socially organized by disparate groups of scientists (physicists, chemists, nuclear chemists, biologists, computer scientists, electrical engineers, statisticians, psychologists etc.) who play different roles in making experiments possible, most of whom are only marginally interested in the experimental outcomes themselves, and who are primarily concerned with the technical-scientific development and employment of their own areas of expertise. This scientific production process, moreover, is also explained in terms of the organization of neuroscience as a whole around research institutes, universities, government agencies, and proprietary entities, all of which are represented as being in competition for resources in a climate in which scientific and wider often popularly understood social goals for the use of science are intermixed and contend.

Ross, we saw, argues that individuals are communities of sub-personal agents who are constructed as single individuals through a sculpting process understood in terms of interaction between single individuals each made up of communities of sub-personal agents. In contrast, Dumit brings whole communities of individual scientists in complex social settings into this sculpting process and the business of determining when and how neuroscience counts people as single individuals. The biomedical identity of single individuals, of course, is hardly the only concern of neuroscience, but it has been an important one on account of the development of neuroscience as a means of addressing mental illness, including schizophrenia. If behaviorally speaking people are 'not themselves,' then an ambition of neuroscience is to determine why and how they might be treated so as to be able to 'be themselves.' As we will see, whether particular individuals are able to 'be themselves' is determined in neuroscience by comparing them to control groups of individuals who are specifically represented as able to 'be themselves.' Neuroscience for Dumit, then, employs a relational conception of individuals, though one that explains individuals in relation to individuals in groups in which they share membership rather than in relation to individuals with whom they play games. Individuals still acquire their status as individuals in relation to other individuals, but only relative to social groups in which they are included which moreover exist only in virtue of way in which research is organized in the neuroscience community. Thus Dumit is clearly concerned with what I treat as the upper bound of the space in which we discuss individuals.

Dumit's framework can also be understood in terms of Clark's idea of external scaffolds, since neuroscience across its many constituent disciplines makes heavy use of a wide array of advanced technologies (all the apparatus of brain scans with all the supporting science, techniques, procedures, equipment, systems of organizing labs, etc. that scans require) which allows it to accomplish computational and information-processing tasks that far exceed and yet also simultaneously enhance scientists' natural abilities. In addition, Dumit's emphasis on the organization of science opens the door to the main subject of Wilcox's thinking: the social distribution of human cognition. Wilcox's main view is that much of neuroscience and neuroeconomics is misguided in operating from a perspective of 'cognitive individualism,' meaning an undue focus on individual cognition, whereas his own view is that human cognition is socially distributed across individuals and social structures, so that different people perform different cognitive tasks in a general division of cognitive labor (Wilcox, 2008, p. 524). Clark's and Wilcox's views, then, are complementary. Clark emphasizes natural and constructed artifacts as scaffolding for cognition; add Wilcox, and we conclude that this scaffolding specifically supports a social division of cognitive labor that Dumit explains in terms of the multi-disciplinary, socially-located organization of a technology-based neuroscience. Thus on this combined view, the concept of 'social-institutional' begins to be less the very general notion with which we began as we give it this more specific structural character as materially and socially distributed, a character, it should be emphasized, that can be described in dense detail and made quite concrete through Dumit's type of anthropological research.

What does this then tell us about single individuals from the perspective of neuroscience and neuroeconomics? Recall that neuroscience as we saw it from Ross's perspective begins with the problem of individuation and the task of negotiating the lower bound of the space where we talk about tipping the balance between the fission and fusion of single individuals toward the latter. Indeed historically speaking neuroscience finds as one of the key rationales for its wider social existence the medical problems of mental illness, for example, schizophrenia, and the fragmentation of otherwise normal healthy single individuals into multiple or divided selves. Then, how in fact does neuroscience actually approach the unhealthy and divided individual when we re-approach it now from the upper bound of the space in which we talk about single individuals, but rather in terms of tipping the balance from fusion to fission? Essentially, this is

a matter of how the science community – working first with popular conceptions of what 'normal human' involves – sets the neuroscience boundaries on what being a 'normal human' means in the laboratory, and then takes this constructed idea of a single integrated individual as the basis on which particular individuals are evaluated as succeeding or failing to behave as such (Dumit, 2004, p. 59). That is, we get whole single individuals not by assembly out of multiple selves, as in Ross's framework, but by discerning who counts or makes it as a single individual relative to the class of 'normal human' single individuals. Consider the steps involved in producing such a class, as in keeping with standard scientific experimental procedure, according to Dumit.

First, one must isolate a sample of the 'relevant' human population for scanning.[7] Second, a set of controls is imposed on this base group of individuals (for such things as smoking, a history of past depression, etc.), as well as for the state of these subjects prior to scanning (what each subject eats or drinks beforehand, their state of rest, instructions regarding how to behave while being scanned, etc.). Third, the individual brainset data produced for each subject scanned is normalized for differences in brain size, physical functioning, etc., in order that each individual's "brain locations can be correlated with those of others," so as to produce a kind of average individual.[8] These normalized or average individual brainsets are then combined to produce a "collective group brainset" which establishes what being a 'normal human' involves, as represented by this particular population sample. It is by this standard that mental illness is evaluated. That is, whether a particular individual functions as a healthy single individual is determined by that individual's functioning (subsequently scanned) relative to that we would see in a collective individual group brainset constructed out of a collection of particular individuals subject to controls and normalized to represent a single 'normal' individual – all of course according to accepted scientific experimental procedures and making use of the technical apparatus of neuroscience and division of tasks across types of experts.

[7] There are many problems associated with the representativeness of these samples, but they go beyond the subject of this paper.

[8] There are different types of normalization procedures – the original standard being the 'Talairach brain' – but they all essentially adjust non-average individual characteristics toward the average. I also put aside here that various researchers intentionally 'inflate' or 'flatten' the scans they produce to enhance their usefulness in exhibiting phenomena of special interest (cf. Alač, 2004).

Thus whether particular individuals are single individuals – understood as 'healthy' individuals – is a matter of whether they fall into a scientifically constructed class of like objects all characterized as 'normal single individuals.' Of course, as there are many avenues of investigation into the nature of mental illness, there are obviously multiple classes of like objects which neuroscience constructs of different kinds of 'normal single individuals.' Further, to compound the factors involved in determining 'normal single individual,' these multiple classes of like objects are each produced by large heterogeneous team of diverse types of researchers, most of whose work is motivated by objectives that are often only tangentially related to the construction of this class of 'normal human,' and who accordingly are working with only vaguely related conceptions of 'normal human' as pertains to their roles in the community's general research project. But the general logic is nonetheless the same whatever the sub-science community and type of neuroscience problem involved with respect to how single individuals are understood. The fusion-fission balance is tipped toward the latter with a 'normal human' construction that stabilizes what a separate single instance is in a class of like objects. Absent the research community succeeding in this individuating stabilization, single individuals do not emerge from mass of characteristics neuroscience ascribes to the general human population. Alternatively, the community fails to break-up the mass of neuroscience observations into classes of phenomena like representative individuals exhibit.

iv. The space of single individual explanations

Let us compare the types of single individual explanations that Ross and Dumit develop. The general outcome for Dumit's organization of science explanation is the same as for Ross's neuroeconomics explanation – single individuals are individuated, and the conception of the individual is a relational one, albeit in a social-institutional framework. But we can clearly see that there are important differences between how individuals emerge in Dumit's neuroscience community framework at upper bound of the individual space compared to how they emerge in Ross's framework at the lower bound of this space. At the upper bound single individuals emerge as instances of a class in terms of which they must be described, whereas at the lower bound individuals emerge as a result of coordination games between their sub-personal selves, which continue to be played while the single individuals they inhabit play games with each other.

Notice the completely different types of explanation these two bounds consequently involve. At the upper bound, single individuals are understood to behave as they do as instances of a class, such that the form of the explanation is, 'members of this class generally behave in manner *x*, so this individual as a member of this class behaves in manner *x*.' I label this a descriptive-prescriptive type explanation, because it prescribes what behavior should be found in the individual according to how the class is described to which the individual belongs. In contrast, at the lower bound, how single individuals are understood to behave is contingent upon the outcomes of interaction between their sub-personal selves and the general environment in which these communities of sub-personal selves operate, so that the form of the explanation is, 'agents interact in conditions described as *y*, producing a range of behaviors associated with contingent outcomes *z*.' I label this an agent-descriptive type explanation, because it describes the behavior of the individual in terms of the concept of agency and the contingencies its exercise generates.

The symmetry, then, of there being two bounds to the space in which we talk about single individuals does not produce forms of explanation that are symmetric. Though the bounds themselves can both be understood as fission-fusion or fusion/fission bounds, the ways in which single individuals emerge as an explanatory concern in each case are quite different, and rely on large incommensurate logics of explanation. There is no obvious basis in prescriptive-descriptive explanations for saying that a single individual as an instance of a class of like individuals is an agent, and indeed attributing agency status to a single individual as a member of a class of like individuals risks undermining the individual's class membership by creating the possibility that the individual's behavior may become atypical. Also, agent-descriptive explanations are really not about the likeness of individuals but rather about agent-specific behavior. Agents by definition are relatively autonomous, and while we can certainly talk about classes of agents (as we do for example in economics), this comes at a fairly high level of generality, and does not pertain to the idea of agency itself.

Thus we are left with a complex explanatory space for characterizing single individuals in the sense that it exhibits largely incommensurate explanatory logics, and this raises the issue of what meta-strategies might be adopted for adopting coherent explanations of individuals in general for neuroscience and neuroeconomics. I distinguish two broad strategies, and then argue in favor of

a version of the second. The first, easily dismissed, is reductionism, an old time-tested and popular form of general explanation in philosophy and the social sciences, in which the goal is to reduce the concepts of an unpreferred framework to those of a preferred framework, and where this is patently not possible, eliminate the offending, unpreferred concepts. The chief proponents of this strategy in the current connection are methodological individualists in economics who operate with naïve single individual economic agent views. They fail to recognize both bounds of single individual explanations, ignoring equally the issue of multiple selves and the issue of individuals' relation to the social groups or aggregates of which they are a part. Second, then, is the meta-strategy of pluralism, or following Helen Longino, 'nonmonism,' meaning "a denial that there is exactly one (correct, complete, consistent) account" of the world. Longino distinguishes three versions of this: an antirealist or constructivist view – multiple accounts are compatible with whatever empirical basis is at issue, and none correspond to it; an eliminativist view – no correct accounts are possible; and a realist view – no single theory or model is complete in itself but each satisfactory account captures some aspect of the world (Longino, 2002, pp. 91-2). It is this last version of nonmonism/pluralism that I recommend in view of its realist commitment and the incommensurabilities that obtain between the two explanatory frameworks discussed here.[9]

Indeed, Longino's pluralism is the epistemological view appropriate to the Wilcox-Clark ontological characterization I have used to give structure to the social-institutional space in which single individual explanations operate. Longino describes how multiple scientific explanations operate at the same time when none are complete by themselves but each satisfactorily captures some aspect of the world. Wilcox-Clark describe the distribution of cognitive agency when we reject cognitive individualism and recognize the role external scaffolding plays in human cognitive activity. That is, as these structures provide an account of cognition as socially distributed, they argue for an account of scientific explanation that is also socially distributed. Further, we may conjecture that Longino's nonmonist/pluralist view that our different science explanations can only be partially successful explains why human cognitive activity has become socially distributed – something about which Wilcox-Clark are relatively agnostic. If our explanations are by nature incomplete, it not unreasonable to suppose that

[9] Thus I reject the alternative unity of science view. Cf. Ross (2005) for an opposing understanding.

human cognition diversifies its activities to fill niches they leave unaddressed. On this understanding, then, we should expect 'general' explanations of subjects, such as has been discussed here in connection with the question of what single individuals are, to be made up of dissimilar and relatively non-communicating sub-explanations whose relation to one another is always contested. The strategy for dealing with this problem of explanation here has been to set boundaries on the overall space in which such explanations can be said to lie. Understanding that overall space, I claim, depends on methodologically employing an understanding of it as distributed and complex. This seems well illustrated in terms of how the relatively incommensurate explanatory logics of Ross's neurocellular neuroeconomics and Dumit's social organization of neuroscience achieve different but complementary tasks in a distributed representation of single individuals. Such a framework also has the obvious advantage that it widens single individual explanation to allow consideration of such issues as multiple selves and membership in groups.

v. Neuroeconomics as a research program in economics

Recall the issue with which this paper began was whether neuroscience might sufficiently inform neuroeconomics so as to lead it to ask genuinely new questions in economics. Neuroeconomics of course is a new and novel research program in economics simply in virtue of its adoption of neuroscience as a basis for investigation of economic questions. But the extensive development of neuroscience and the independence of its scientific concerns from those of economics creates the very real possibility that economics will make only incidental use of neuroscience, largely confining neuroeconomics to corroborating existing theories and conceptions of economics. In this case it would be difficult to judge whether neuroeconomics has become a genuinely new research program in economics in the sense of being one that significantly impacts the development of economics.

My view of the matter is that the current status and direction of neuroeconomics is unclear. On the one hand, much neuroeconomics research seems primarily instrumental in nature in that it sees neuroscience as a means of securing further evidence for propositions which economists already have some confidence in or alternatively as an opportunity for settling existing disputes

between economists regarding theory selection. This instrumental orientation seems unlikely to raise questions motivated by neuroscience's own concerns that might inform neuroeconomics in ways that would have substantial impact on the nature of economics.[10] If this becomes the dominant form of neuroeconomic research, it seems that neuroeconomics will neither be a long-lasting research program nor have a significant impact on economics. On the other hand, there is also neuroeconomic research that raises issues that have potentially far-reaching significance for economics, as I have argued in this paper to be the case with respect to research that addresses the boundaries of the conception of the individual. Other more far-reaching questions are likely being raised by other neuroeconomics researchers, but by its nature this more exploratory type of research is often difficult to recognize in its early stages. Thus it is likely too early to judge the significance of neuroeconomics as a new research program in economics. But I suggest we may begin to chart its progress according to how deeply we believe it is becoming truly engaged with the ambitions and horizons of neuroscience.

References

Ainslie, G., 2001. Breakdown of the Will. Cambridge University Press, Cambridge.

Alač, M., 2004. Negotiating Pictures of Numbers. Social Epistemology 18 (2-3), 199-214.

Bennett, M., Hacker, P., 2003. Philosophical Foundations of Neuroscience. Blackwell, Oxford.

Brewer, M., 2001. The Many Faces of Social Identity: Implications for Political Psychology. Political Psychology 22, 115-125.

[10] There is also the risk that neuroeconomics carried out on this basis will spend much time heralding 'discoveries' already well established outside of economics.

Brown, R., 2000. Social Identity Theory: Past Achievements, Current Problems and Future Challenges. European Journal of Social Psychology 30, 745-78.

Bruner, J., 1992. Acts of Meaning. Harvard University Press, Cambridge, MA.

Bruner, J., 2002. Making Stories: Law, Literature, Life. Farrar, Straus and Giroux, New York.

Camerer, C., 2006. Behavioral Economics. In: Blundell, R., Newey, W., Persson, T., (Eds.), Advances in Economics and Econometrics. Cambridge University Press, Cambridge, pp. 181-214.

Camerer, C., 2008. The Potential of Neuroeconomics. Economics and Philosophy 24, 369-379.

Camerer, C., Loewenstein, G., Prelec, D., 2005. Neuroeconomics: How Neuroscience Can Inform Economics. Journal of Economic Literature 43, 9-64.

Clark, A., 1998. Being There. MIT Press, Cambridge, MA.

Davis, J., 2007. Review: Economic Theory and Cognitive Science by Don Ross. Economics and Philosophy 23 (July 2007), 245-252.

Davis, J., 2008. The Individual in Recent Economics: Internalist and Externalist Conceptions. Storia del pensiero economico, Nuova serie, Anno V, no. 1 (gennaio/giugno 2008), 5-24.

Davis, J., 2009. Competing Conceptions of the Individual in Recent Economics. In: Kincaid, H., Ross, D. (Eds.), The Oxford Handbook of the Philosophy of Economic Science. University Press, Oxford, pp. 223-244.

Dennett, D., 1987. The Intentional Stance. MIT Press, Cambridge, MA.

Dennett, D., 1991. Consciousness Explained. Little Brown. Boston.

Dumit, J., 2004. Picturing Personhood: Brain Scans and Biomedical Identity. Princeton University Press, Princeton.

Eakin, J., 2007. The Economy of Narrative Identity. In: Weintraub. E., Forget, E. (Eds)., Economists' Lives and Autobiography in the History of Economics. Duke University Press, Durham, 117-133.

Glimcher, P., 2004. Decisions, Uncertainty and the Brain. MIT Press, Cambridge, MA.

Heukelom, F., 2008. Building and Defining Behavioral Economics. Nijmegen University working paper.

Kahneman, D., Tversky, A., 1979. Prospect theory: An Analysis of Decision under Risk. Econometrica 47, 263-291.

Longino, H., 2002. The Fate of Knowledge. Princeton University Press, Princeton.

Ross, D., 2005. Economic Theory and Cognitive Science: Microexplanation. MIT Press, Cambridge, MA.

Ross, D., 2006. The Economic and Evolutionary Basis of Selves. Cognitive Systems Research 7, 246-258.

Ross, D., 2007. *H. sapiens* as Ecologically Special: What Does Language Contribute? Language Sciences 29 (5), 710-731.

Ross, D., 2008. Two Styles of Neuroeconomics. Economics and Philosophy 24, 373-383.

Wilcox, N., 2008. Against Simplicity and Cognitive Individualism. Economics and Philosophy 24, 523-32.