Journal of Economic Literature 2018, 56(1), 234–264
https://doi.org/10.1257/jel.20171370

A Review Essay on *Social Neuroscience*: Can Research on the Social Brain and Economics Inform Each Other?†

Carlos Alós-Ferrer*

Social neuroscience studies the "social brain," conceived as the set of brain structures and functions supporting the perception and evaluation of the social environment. This article provides an overview of the field, using the book Social Neuroscience: Brain, Mind, and Society *(Russell K. Schutt, Larry J. Seidman, and Matcheri S. Keshavan, editors) as a starting point. Topics include the evolution of the social brain, the concept of "theory of mind," the relevant brain networks, and documented failures of the social brain. I argue that social neuroscience and economics can greatly benefit from each other because the social brain underlies interpersonal decision making, as studied in economics. (JEL* D11, D71, D87, Z13)

1. *Introduction*

Economics is a social science. As much as understanding individual decision making is crucial for its development, it is first through social interactions that markets and institutions appear. The economic world is a product of the social world, and the social world is a product of the human capability to interact in ever-larger groups, from the small hunter–gatherer tribes of the stone age to our heavily interconnected planetary society. Conversely, the feats an individual is capable of are enormously augmented by the existence of the social world, as less and less effort is required to satisfy basic needs and more time and resources are freed to pursue higher goals. It should come as no surprise that the human brain is uniquely adapted to the social world. While the dependence of human society on the capabilities and characteristics of the human brain is little more than a tautology, the recent decades have brought the recognition that the relation is *bidirectional*, that is, the converse link is equally strong, and the human brain's normal functioning is heavily dependent on the social world, to the extent that disturbances in the latter lead to disturbances in the former.

We are hopefully past the point where an argument is necessary as to whether

*Laboratory for Social and Neural Systems Research and Department of Economics, University of Zurich. This paper was written while the author was affiliated with the University of Cologne. The author thanks Anja Achtziger, Todd Hare, Sabine Hügelschäfer, and Christian Ruff for helpful comments. Financial support from the Research Unit "Psychoeconomics" (FOR 1882) of the German Research Foundation (DFG) is gratefully acknowledged.

†Go to https://doi.org/10.1257/jel.20171370 to visit the article page and view author disclosure statement(s).

economics should pay attention to the human brain. A human being is his or her brain. From toes to nose, everything else is a convenient accessory. It is, hence, clear that economics studies human brains, since it studies human decision makers and their behavior. Arguing whether economics should pay attention to neuroscientific studies is as fruitful as arguing whether economic theory should pay attention to mathematics. The question is not *whether* it should (the answer is of course yes, in both cases), but *how* and to what extent. The exact role played by potassium ions in brain processes is probably not very consequential for economic studies, exactly as the elegance of a correct mathematical proof has little bearing on the economic significance of the result. In both cases, that is hardly the point.

The bidirectionality of the link between brain and social world renders the recognition that economics should pay attention to neuroscience even more obvious. At the same time, it points out that the converse is equally true, and neuroscience should pay attention to the social sciences, including economics. This might sound surprising at first, because economists typically and implicitly think of their discipline as dealing with high-order constructs, while some other disciplines concentrate on more clearly delimited objects. One could naïvely think that neuroscience is exactly one such example. The argument would be that, after all, the only thing needed to understand the functioning of the brain is, well, the brain. This reductionistic argument, however, is fallacious, and amounts to attempting to develop a theory of planes by looking at what they are made of, ignoring their interaction with human society, starting with the fact that planes fly and ending with the fact that most planes fly to bring people to places they want to be, typically in exchange for money. The human brain evolved over time, and in doing so, certain characteristics were selected for. It is becoming increasingly clear that the brain and the first versions of the human social world coevolved over time. Understanding which characteristics this coevolution selected for and *why* they provided an evolutionary advantage provides a far better understanding of the nature and functioning of the brain than statistical studies on the relative proportion of neurons and glial cells in each brain region. This has been recognized in neuroscience (see, e.g., Krakauer et al. 2017), and is the starting point of the recent book of Schutt, Seidman, and Keshavan (2015), which this review essay will focus on. The following quote drives the point home.

> [...] as research on the brain accelerated, it became increasingly clear that internal examination alone was insufficient for understanding the development and functioning of [the brain].
>
> *Schutt, Seidman, and Keshavan* (2015, p. 2)

Social Neuroscience is the name of a nascent discipline that started with this realization. The term was coined by Cacioppo and Berntson (1992), who viewed it as the confluence of social psychology and neuroscience. Those authors defended an approach based on "multilevel integrative analysis" for the study of both brain and behavior, meaning a bidirectional exchange between the different levels of analysis that arise from the different structural scales targeted by different disciplines, that is, the biological (microscopic) substrate studied by neuroscience and the social (macroscopic) level proper of psychology or economics. The objective of the present work is to present a selective overview of developments and trends in this field. As a review essay, the scope will be roughly defined by the topics covered in the book by Schutt, Seidman, and Keshavan (2015), aptly titled *Social Neuroscience: Brain, Mind, and Society.* However, economics is notably absent from that text, and since the intention of the

overview is to pay special attention to the relevance of recent research in social neuroscience *for economics and economists*, I will take detours outside the scope of the book as needed in order to give the proper context.

To understand the internal dynamics of the social neuroscience field, and to put the remainder of the article in perspective, the reader might find it useful to keep in mind two key realizations and two practical observations. The two key realizations, already hinted at above, are, first, the bidirectionality between brain and social world, and second, their likely coevolution. The first explains the increasing interdisciplinarity of the field, as social scientists and neuroscientists realize that their respective disciplines are insufficient to foster progress beyond a certain point. The second helps us move beyond the merely descriptive and understand the origins and functioning of social and neurological processes. The first practical observation is that, to date, much research in this field has been driven by mental illnesses, exactly in the same way that much of the early progress in neuroscience occurred as a consequence of the study of brain lesions (see, e.g., chapter 2 of Cacioppo, Visser, and Pickett 2006). Many mental illnesses, and in particular schizophrenia, involve a breakdown of the "social brain," which is the object of study of social neuroscience. For this reason, mental illnesses, in addition to being a socially important problem worthy of study, are an invaluable source of data for the field. This is not to say that current social neuroscience focuses exclusively on such illnesses. To the contrary, in the last fifteen years, research in social neuroscience has increasingly focused on the social behavior of mentally healthy decision makers, allowing for the scope of research to encompass such social phenomena as, for instance, stereotyping, social exclusion, and social interactions, to mention just a few. The second practical observation, maybe linked to the first, is that economics is currently underrepresented in the field and (in the author's opinion) currently risks being left out.

The structure of the article is as follows. Section 2 briefly reviews some recent developments and trends within neuroscience that are necessary to understand both the goals of social neuroscience and the challenges faced by neuroeconomics. Section 3 provides a short historical perspective on the development of the field. Section 4 discusses the main insights of social neuroscience with respect to the nature, functions, and evolution of the social brain. Section 5 discusses the failures of the social brain that fuel and motivate large parts of the field, ranging from schizophrenia to data on orphanage-raised individuals. Section 6 offers a partial and probably biased discussion on what social neuroscience and economics can learn from each other. Along the way, the article will refer the reader to specific contributions (chapters) in Schutt, Seidman, and Keshavan (2015), some earlier texts such as Cacioppo, Visser, and Pickett (2006), and specific articles that go beyond those texts.

2. *Some Developments and Trends in Neuroscience*

Over a decade ago, Camerer, Loewenstein, and Prelec (2005) provided a review of developments in neuroscience and how they might inform economics. Although a large part of the early efforts were focused on individual decision making, it has to be remarked that, since the very beginning, a certain number of neuroeconomic studies considered paradigms centered on explicitly social situations (see, e.g., the review of Fehr and Camerer 2007 on social preferences). The discipline has experienced significant growth since then, as illustrated in the more recent review of Fehr and Rangel (2011) on the neural basis of economic choice. Likewise, neuroscience has started paying increasing

attention to economic decision making. It is, of course, impossible to review the many developments within this field here. A handful of those, however, are important both to understand the current trends in social neuroscience and to clarify possible misunderstandings among economists.

2.1 *Brain Localization versus Network Connectivity*

One frequent misunderstanding arising from early neuroeconomic studies is related to the *reverse inference fallacy*. Suppose differential activity in a given brain region is detected during a mental process X. Knowing from the literature that this particular brain region is also involved in processes of type Y, the fallacy leads the researcher to conclude that process X must be of type Y, which is clearly not justified. To be clear, careful reverse inference can be valuable, since it can be used, e.g., to arbitrate between competing hypotheses and find out which ones are *compatible* with the data. The brain, however, is a highly complex organ in whose structure no particular efficiency-based planning can be discerned. Practically each and every brain substructure is involved in a large number of different functions, the connections between them being often less than clear. It is of course tempting to draw a brain map and neatly label brain regions with purported functions in a one-to-one way. Such enticingly simple maps, however, must be seen as extreme simplifications, at best.

Indeed, the first important development to take into account is that neuroscience is steadily moving away from narrow localization ideas (and has been doing so for some time). Such ideas can be traced back to the concept of "brain modularity," which rests on the assumptions of domain specificity and fixed neural localization for many aspects of cognition (Fodor 1983). Such assumptions, however, have been repeatedly challenged in recent years. Certain brain functions (say, controlling breathing or heart activity) can clearly be traced back to specific brain regions, but higher-order activities such as decision making, reward processing, or social evaluations are not associated with a single, well-delimited brain region that researchers can then target as a region of interest. Although there is evidence of modularity to a certain extent, such functions typically involve interactions among a number of brain regions, each of which, in turn, is also involved in a large number of other functions. There is a clear shift in neuroscience, away from the study of brain structures in isolation and towards the study of interconnected *networks* of brain structures associated with certain broad kinds of functions. This requires, of course, a certain knowledge of anatomical connectivity, that is, of the neural pathways actually connecting distant brain regions. But the key concept is *functional connectivity*, which can be roughly defined as the temporal correlation of activity in different brain regions during a given (class of) tasks, and can be studied, e.g., by either correlating activity with a fixed region of interest or using model-free techniques as independent component analysis. The take-home message is that, as we study higher social cognitive functions, we might discover that there is little to learn from whether specific, localized regions of the brain are active or not. Such brain regions might be more constructively considered nodes in a series of interrelated networks, with the same region possibly playing different roles in different networks.[1]

[1] Some researchers go even further in their rejection of modular ideas and propose to focus on domain-general, highly distributed networks. For instance, in the field of memory, Fuster (2009) proposed a distributed-network paradigm where "nodes" would capture a far smaller scale than the brain regions usually discussed in neuroeconomics (see also Fuster and Bressler 2012). Following similar arguments, Lindquist and Barrett (2012) argue that human emotions such as, e.g., anger or disgust, are mental states resulting from the interaction of broadly distributed functional networks. In addition, there is increasing evi-

A basic example is the *default mode network* (DMN), associated with internal mentation and the brain's resting state, which includes the medial prefrontal cortex (MPFC), the posterior cingulate cortex (PCC), parts of the posterior temporal cortex around the temporoparietal junction (TPJ), and areas along the lateral temporal cortex (Buckner, Andrews-Hanna, and Schacter 2008). As we will see below, there might be just such a network corresponding to the "social brain," which, somewhat surprisingly, overlaps with the DMN. And of course, there has been much discussion in neuroeconomics about the *reward valuation network* (RVN). As a case in point, it is illustrative to have a brief look at this better-known case, which will also prove to be relevant when we turn to the discussion of the social brain (figure 1 provides a qualitative depiction of brain regions and networks mentioned in the text, as well as a reference for their often-abbreviated names).

Early brain-imaging studies showed a large variety of brain regions related to economic decision making. The picture was later clarified by a series of discoveries related to the neurotransmitter dopamine and the neural pathways dependent on it. Adopting a simplified view, the midbrain dopamine system encodes the difference between expected and realized rewards (the "reward prediction error"), which in turn is used as an input to construct a "value" (see Schultz 1998, 2010) that underlies decision making. The RVN is essentially the network which encodes that value, which is tempting to compare to neoclassical utility. Its key components are anatomically connected to the dopaminergic neurons and receive their input: the striatum (a deep brain structure, part of the basal ganglia, which includes the nucleus accumbens) and the medial prefrontal cortex (especially the ventromedial part). It is likely that the network also includes other frontal brain areas, especially the frontopolar and orbitofrontal cortices (Padoa-Schioppa and Assad 2006). Many studies have included several other regions in this network, but recent work has suggested that the actual RVN should be distinguished from other, related ones. In particular, it might make sense to speak of a *cognitive (self-)control network*, including the dorsolateral prefrontal cortex (DLPFC) and the anterior cingulate cortex (ACC) (Gläscher et al. 2012). This is especially interesting because these two areas are evolutionarily quite different: the DLPFC is a neocortex area, which has been consistently associated with higher cognitive functions, including executive control and working memory; the ACC is an older, limbic area, which performs more basic functions as the detection of an elementary conflict between competing impulses. Also, other studies have revealed that the process of choice implementation in the brain might be better understood as functionally separate from the encoding of valuation, and there might exist a specific network for this purpose (labeled as the *choice implementation network* in figure 1), including the posterior parietal cortex and the nearby intraparietal sulcus. There is a certain (evolutionary) logic to this, as the main functions of the intraparietal sulcus are related to visual attention and eye–hand coordination (Kable and Glimcher 2009), both necessary for the execution of simple choices.

In view of the evidence accumulated in the recent decades, a highly simplified picture of decision making in the brain might be as follows. First, the dopaminergic system (including the striatum) keeps track of the reward prediction error and the medial prefrontal area of the brain (the ventromedial prefrontal

dence that the more complex a function is, the more idiosyncratic its localization in the brain is. While there might be starting regions for basic functions, common to all individuals, higher-order complex processes might be organized differently in each individual brain (see Fehr 2013; Fehr and Herrmann 2015).

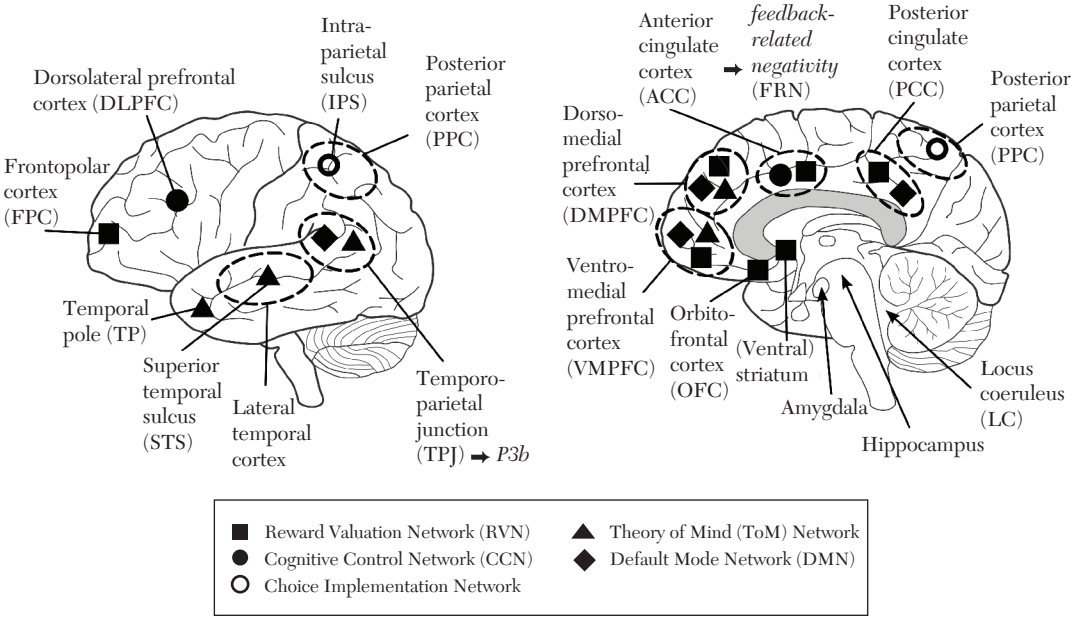

Figure 1

Notes: Qualitative representation of selected brain regions in the networks mentioned in the text, plus some additional, important structures. The left-hand and right-hand pictures present the lateral and medial views of the brain, respectively. The symbols are meant to represent an approximate orientation and not a precise area. Thanks are due to Alexander Jaudas for this figure.

cortex and possibly the orbitofrontal cortex, which are anatomically adjacent or overlapping, depending on naming conventions) then builds "decision values," in part through a reinforcement process on the basis of the dopaminergic feedback (but also relying on other sources, e.g., explicit labels or information from working or episodic memory). This reinforcement process is most likely the function of the RVN and might integrate different inputs into a single value. Second, the anterior cingulate cortex (which is anatomically just behind the medial prefrontal cortex) fulfills such crucial functions as action monitoring and error and conflict detection. At the same time, left and right of the medial prefrontal area, the dorsolateral prefrontal cortex provides the link to working memory and executive control. These two areas might form the core of a cognitive control network. Third, regions located in the parietal areas that originally governed visual attention and hand movements might form a choice implementation network executing and monitoring actual choices. Fourth, a number of deep brain regions naturally contribute to the process, ranging from the amygdala (which associates stimuli to motivational salience) to the hippocampus (which provides the link to episodic memory).

In summary, there is no specific region for decision making, not even for a well-defined component thereof. At the same time, not all regions identified in all decision-mak-

ing studies are part of the same network. Rather, there are functionally different but highly interconnected networks fulfilling different functions that contribute to decision making. It is likely that the same is true for virtually any higher-order brain function that economics might set its sights on. The take-home message at this point is that any future neuroeconomics study built on the premise that X regions do this and Y regions do that will be at least partly misleading.

2.2 *Epigenetics and Neuroplasticity*

The second important development is a further shift, this time away from the conception of the adult brain as a static structure and towards a more dynamic vision. Advances in genetics and neuroscience have made clear that, when studying brain processes and human behavior, drawing a clear separation between biological and social factors is not justified, and this realization underlies several of the chapters in Schutt, Seidman, and Keshavan (2015). On the one hand, genetics has shown that whether genes become active or not depends on environmental factors, with the consequence that identical genetic endowments might result in different phenotypical expressions. In particular, the link from the biological to the social is far from being a straightforward one. On the other hand, neurology has discovered that the brain can significantly change in response to experiences, including those arising from social interaction, providing a feedback link from the social to the biological. Further, neuroimaging studies are making increasingly clear that there is no physical separation of brain regions regulating social and "rational" processing. We will now turn to these developments.

The recent advances in research in genetics are particularly relevant for the long-run research agenda in microeconomics. In the search for the ultimate determinants of behavior, microeconomists turned to formal assumptions (derived from normative considerations or introspection) first, and to psychological foundations second. The advent of neuroeconomics can be understood as the third step, in which a neurobiological basis of behavioral processes is sought. It is easy to jump ahead and predict that the fourth step should be "genoeconomics," for our genes carry a complete blueprint of what we are to become, and hence of how we are to behave. Alas, this argument, seldom explicitly stated but often implicitly lurking in the background nowadays, is flawed. It is simply not true that our genetic blueprint fully determines who we are. The old "nature vs. nurture" argument has by now taken a huge step forward with the discovery of *epigenetics*. In a nutshell, genes determine who we are and what we do (the "phenotype") through the production of specific proteins. However, the physical and social environment to which we are exposed influences whether specific genes become activated or not. Insights from molecular biology have illuminated the mechanisms that modify gene function without altering the DNA itself. Certain molecules called microRNAs can act on the molecules that translate the information of the genes into protein production, effectively silencing the genes. More generally, we now know that one of the biochemical mechanisms altering how the instructions encoded in the DNA are translated is the addition of a methyl group to a DNA nucleotide, called DNA methylation. Methylation can be measured, delivering an indicator of the extents of modifications that the translation of a person's genetic blueprint has suffered. Crucially, it turns out that social experiences across the lifespan can activate all these processes and turn certain genes off or on. In other words, our social interactions, especially in early childhood, can very well influence whether a certain part of our genetic heritage determines our later behavior or becomes a dormant, irrelevant

marker in the background. For instance, it has been shown that orphans raised in state institutions present higher levels of DNA methylation than children raised by families (Naumova et al. 2012; see also section 5.2).

If epigenetics shows that the mapping from genes to behavior is less deterministic than originally assumed, the discovery of *neuroplasticity* delivers a similar caveat for the mapping between specific brain areas and modes of behavior (chapter 2 of Schutt, Seidman, and Keshavan 2015, written by Matcheri S. Keshavan, provides a short overview). Even ignoring the complications of epigenetics, it has been shown that experience (both in childhood and adulthood) directly changes the brain. Specifically, the brain responds to physical and social stimuli by growing new connections and letting others decay. Of course, the brain changes in response to early childhood experiences, which is hardly surprising, as it is still not completely developed at that stage. More important for economics, however, is the fact that neuroplasticity is not limited to such early phases. *Experience-dependent neuroplasticity* occurs throughout development, and is particularly important for activities related to reward and learning. Even beyond biological maturation, the brain is far from immutable. Daily activities and the input from our social environment can lead to long-lasting changes in brain structure. This has been known for a long time for more purely biological domains, as, e.g., Nobel Prize-winning studies on the development of the visual cortex (Wiesel 1982). A more recent example of activity-induced changes in brain structure established that adult London taxi drivers experienced posterior hippocampus growth as a consequence of the extensive memory training required by their job, in comparison to both age-matched controls (Maguire et al. 2000) and London bus drivers (Maguire, Woollett, and Spiers 2006) (a far better control group).[2] This is an example of *neurogenesis*, the growing of new brain tissue, a possibility which is well-documented only for the hippocampus and a region known as the olfactory bulb, related to the sense of smell (Ernst and Frisen 2015). Most examples of neuroplasticity correspond to neuronal reorganization and specialization, where, for instance, the functions of damaged brain regions are taken over by nearby regions.

While neuroplasticity typically has positive connotations, the brain is of course also susceptible to physical alterations and impairments even in the absence of lesions (neurodegeneration). An example of basic neuronal change in response to the social environment involves stress. If our environment's demands systematically exceed our capacity, chronic stress might result, and, through the release of cortisol, lead to the shrinkage of neuronal dendrites and a predisposition to psychiatric disorders such as depression (McEwen 2012). This and many other examples cast doubt on the idea of a "standard" brain, which underlies most brain localization studies, and remind us that the brain is dynamic and it adapts to our environment not only in an evolutionary sense, but also within the scope of every individual life.

2.3 *Beyond fMRI: Processes and the EEG*

It is probably safe to say that functional magnetic resonance imaging (fMRI) studies are the face of neuroeconomics, and most economists will immediately associate one with the other. Neuroscience, in contrast, relies on a mixture of methods, of which fMRI

[2] The hippocampus is related to memory in general and spatial memory in particular, and seems to be crucial for complex spatial representations. To emphasize the caveats raised in the previous section, however, it should be remarked that it is not the only brain region associated with such functions. In particular, Maguire et al. (1998) argue that it is part of a "human navigation network."

is an important one, but by no means the only one. Even though this point goes of course beyond the confines of social neuroscience, it is maybe important to make it clear at this point that not every neuroeconomic question must give rise to an fMRI study. Also, no discussion of developments and caveats for the future of neuroeconomics and social neuroscience would be complete without addressing the limits of fMRI studies.

At the risk of stating the obvious, it should be remembered that fMRI images do *not* represent neural activity, but rather a correlate thereof. The signal underlying such images, called "blood oxygen level dependent" or BOLD signal, represents changes in blood flow, which is assumed to occur after oxygen-consuming neural activity has taken place. Since blood flow is slow, fMRI images present indirect evidence of where activity might have occurred in a large time window (two seconds in the best of cases, which, for brain activity, is extremely long). That is, fMRI concentrates on *where*, not on *when*. Even leaving aside the possibility that blood flow does not always reliably track neural activity (see Schleim and Roiser 2009, for a discussion), the broad time window creates an important problem, since different functions of the same brain region, resulting in activity at different time points, will be necessarily lumped together in fMRI measurements. As a consequence, there are natural limits to what can be achieved with fMRI analyses in terms of mapping functions to brain regions or networks, and indeed some puzzling results in neuroeconomic studies might be due to the lumping nature of the BOLD signal.

It should also be remembered that the statistical analysis of fMRI data is far from trivial. Roughly, the colored renderings that illustrate such data represent levels of significance for coefficients estimated in particular voxels (three-dimensional pixels), say, for the comparison of a treatment group versus a control group. But the coefficients being compared are typically regression coefficients estimated at the particular voxel. A host of corrections and adjustments are performed at different stages of the process, and most researchers simply apply those considered standard and incorporated in standardized software. Alas, the statistics of such multistage analyses are not entirely settled, and occasional polemics flare up in the neuroscience literature regarding their reliability. For instance, Vul et al. (2009) targeted studies in social neuroscience and argued that certain kinds of multistage analysis might have artificially inflated significance levels in early studies in this field, especially when computing correlations between brain activity and personality-related scales. More recently, Eklund, Nichols, and Knutsson (2016) showed that possibly unwarranted assumptions on the spatial autocorrelation functions incorporated in certain software packages might have crucially affected a large number of fMRI studies in recent years, resulting in incorrectly low p-values.

These comments are merely meant as a reminder that, first, fMRI is an indirect measure of brain localization with high spatial but low temporal resolution, and, second, fMRI analyses are comparatively young and involved, and the underlying statistical methods are still evolving. While fMRI studies are extremely valuable and pave the way to further analysis by, e.g., identifying the regions of interest, economists should be aware of the fact that neuroscience has other (cheaper!) tools at its disposal, some of which hold particular promise for the field of social neuroscience. The simplest and oldest such technique is the analysis of electrocortical brain activity by means of the electroencephalograph (EEG). Essentially, the technique directly measures actual brain activity (electricity), but does so by placing electrodes on the scalp. Source localization is far less precise than in the case of fMRI,

but temporal resolution is measured in milliseconds. EEG studies have developed a catalogue of so-called *event-related potentials* (ERPs), which arise in response to specific classes of stimuli and whose latencies (times of occurrence) and peak amplitudes deliver information on the onset and intensity of brain processes. Different ERPs might originate in the same brain region, but the temporal difference still allows for differentiated analysis. Additionally, the statistical techniques necessary to analyze EEG data are often comparatively simple, as e.g., testing amplitude differences (measured in microvolts) across groups.

The seminal book on social neuroscience by Cacioppo, Visser, and Pickett (2006) included a detailed discussion of the use of ERPs for the study of social perception. Some more recent ERP studies have illustrated the potential interest of the technique for economic decision making, social and individual. As an example, an ERP component called *feedback-related negativity* (FRN), generated in the anterior cingulate cortex around 250 ms after win/loss stimuli presentation, has been reliably associated with processes of reinforcement learning. By analyzing this component, Achtziger et al. (2015) showed that increased incentives might result in increased reliance on reinforcement learning, but not necessarily in higher performance. Since the ACC is involved in many other functions, ACC activity cannot be uniquely associated with reinforcement learning (or vice versa, of course), and such a conclusion could not be directly drawn if one only knew that the ACC was active in a certain time window. Also interesting are applications pinning down when decisions are made, which can be done, e.g., by studying *lateralized readiness potentials*, which reflect the extent of motor preparation and indicate its onset, that is, when a decision maker is preparing to give one response or another by pressing certain keys. In a belief-updating study, Achtziger et al. (2014) showed that decision makers prone to the conservatism bias (overweighting the prior compared to new sample information) tend to prepare their response *before* the new information is presented, showing that conservatism bias might often arise from a base-rate-only heuristic where sample information is ignored (as proposed by Navon 1978 and Gigerenzer and Hoffrage 1995), and not from failures in information aggregation or retrieval (as argued by Edwards, 1982 or Dougherty, Gettys, and Ogden 1999). Obviously, such a conclusion (which helps separate different theories on faulty economic decision making) relies on the temporal precision of the EEG and cannot be obtained on the basis of choice data, response times, or fMRI measurements.

Neuroeconomics, through its focus on reward and valuation, has so far devoted a great deal of attention to the neurotransmitter dopamine and the dopaminergic pathways of the brain. As neuroeconomics advances, other neurotransmitters and pathways will become relevant. Although EEG studies in isolation can, of course, not distinguish among neurotransmitters, the temporal resolution they provide, as well as the possibility to focus on very early reactions, will make them invaluable. Consider processes of attention, which are closely related to motivation and are hence essential for both decision and social neuroscience studies. Motivationally relevant stimuli elicit the release of the neurotransmitter noradrenaline, which promotes alertness and focuses attention.[3] Noradrenaline release results

[3] Noradrenaline is chemically related to adrenaline, but it plays a role in the neural system. In the United States, they are called norepinephrine and epinephrine since the word "adrenalin" was trademarked by a pharmaceutical firm (Tansey 1995). The names are translations of each other. *Ad renes* and *epi nephros* are just references to the suprarenal glands (above the kidneys) synthesizing adrenaline, respectively in Latin and Greek.

in activity throughout a network of brain regions innervated by the noradrenergic pathways. The onset, however, is on the locus coeruleus (LC), an anatomically deep structure located in the brain stem. The initial reaction is extremely fast (100–150 ms), and the noradrenergic system (called locus coeruleus–noradrenergic or LC–NA system) is relatively broad, hence it is difficult to pinpoint natural targets for fMRI studies addressing specific processes of attention and arousal. However, the EEG provides natural, easy-to-measure correlates of such processes, in the form of two ERP components called P3a and P3b. Those have been argued to correspond to early activity in the LC–NA system (Nieuwenhuis, Aston-Jones, and Cohen 2005; Walz et al. 2013). The P3a, of lateral-prefrontal origin, is associated with novel or salient stimuli. The P3b, probably originating in or near the temporoparietal junction (which, as we will discuss below, plays an important role in the "social brain"), responds to infrequent or motivationally significant stimuli (e.g., large gains and losses; Yeung and Sanfey 2004) and is moderated by deliberate attention.

Since frequency and consistency are closely related, ERPs such as the P3b are particularly useful to obtain insights on social attitudes. A year after the seminal contribution on social neuroscience (Cacioppo and Berntson 1992), Cacioppo et al. (1993) pointed out that the P3b is more pronounced when a target word is inconsistent with context words, even if the inconsistency depends on subjective evaluations, for instance when a word with a subjectively negative evaluation appears within a series of positive words. This opened the door to a series of designs assessing social attitudes that do not rely on participants' self-reports. Hence, the P3b and related components can be used to estimate social attitudes when self-reports are likely to fail (e.g., racial attitudes; see Ito, Thompson, and Cacioppo 2004). This adds to a collection of ERPs that are used to evaluate social attitudes and their appearance. For instance, Ito and Urland (2003) used ERPs to show that both gender and race are differentiated within 200 ms (!) in the human brain, which suggests that attempts to build rationalistic models of social attitudes might be overoptimistic (see also Hügelschäfer, Jaudas, and Achtziger 2016).[4] Indeed, an especially interesting fact is that the ERPs employed in such studies are extremely fast (occurring way before conscious thought). For instance, the N170 (whose name refers to a negative deflection occurring around 170 ms after stimuli presentation) responds selectively to faces. Ratner and Amodio (2013) found larger N170 to ingroup compared with outgroup faces, where the groups were exogenously defined according to a minimal-group paradigm (Tajfel and Turner 1986). Such insights exemplify how the EEG technique, with its high temporal resolution, can contribute to social neuroscience along dimensions unreachable for fMRI research (for a recent review and other examples, see e.g., Amodio, Bartholow, and Ito 2014).

3. *A Brief Historical Perspective of the Field*

The previous section has attempted to provide context in the form of recent trends and developments, formulated as broad caveats. A further and equally important caveat arises from the past. Every scientific discipline naturally develops a certain path dependence, eventually redefining itself along the way. Schutt, Seidman, and Keshavan's (2015) book creates the impression that social neuroscience might be on the verge of just

[4] ERP research has consistently shown that certain kinds of economically relevant stimuli elicit extremely early neural responses. For instance, the FRN mentioned above, which typically peaks around or before 250 ms after stimuli presentation, responds to high versus low economic rewards.

such a redefinition. However, it is arguable whether the discipline covered by that text is actually the broad field of social neuroscience as envisioned by Cacioppo, Visser, and Pickett (2006). There are significant departures. While neuroeconomics and the more psychologically savvy parts of behavioral economics are converging into the broader "neuropsychoeconomics," which might easily be conceived to encompass social neuroscience, the developments covered in the book of Schutt, Seidman, and Keshavan (2015) sketch a field which, with apologies, might be more properly called "neurosociopsychiatry," understood as the confluence of sociology and the well-established field of neuropsychiatry. Of course, labels are far less important than the actual content, but in this case, the latter label would more appropriately convey the motivations (and the problems) underlying this trend. Sociology and psychiatry come with heavy historical baggage that is absent in other disciplines (such as economics or psychology) and which has strongly influenced their development, and so it is convenient to briefly recapitulate that baggage here, following, in part, the introductory chapter of Schutt, Seidman, and Keshavan (2015).

Sociology has been allergic to contact with the biological sciences for a long time. The reasons are understandable. Some early sociologists supported social Darwinism (Bannister 1973), which was in turn used to justify the racist policies of the German Third Reich. Postwar sociology, hence, attempted to steer clear of any discussion of biological foundations, a tendency only reinforced by the continued attempts to give pseudoscientific foundations to antisemitism in Europe and discrimination against blacks in the United States (Degler 1991). In the process, many sociologists developed a taboo mentality with respect to any attempt to examine possible biological foundations (genetic or neuroscientific) of their science.[5] While this might have been an understandable defense mechanism, it was not conducive to further the interdisciplinary exchange. Interestingly, this reaction also contributed to setting sociology apart from economics, as the former came to define sociological phenomena as strictly external to the individual, while the latter attempted to build models of both individual and aggregate behavior (micro- and macroeconomics, respectively). This focus on the social as surpassing all individual causes is especially strange since, contrary to economics, sociology has never abandoned mental illness as a subject of study. Sociological studies on mental health, however, have concentrated on external factors, deriving data, e.g., from community studies (Mirowsky and Ross 2003). It is only very recently that sociology has started to look beyond the social towards the neurobiological, and hence the new "neurosociopsychiatry" field is younger and less developed than, say, neuroeconomics.

Psychiatry, on the other hand, has moved back and forth between the purely social and the purely biological approach along its history (Grob 1994), consistently failing to find an equilibrium until very recently. Towards the end of the nineteenth century, the asylums (state psychiatric institutions) were born of the misconception that a socially ordered environment would suffice to cure mental illness. Instead, these institutions did active harm by isolating the mentally ill from the natural social environment that their brains were supposed to interact with

[5] As Russell K. Schutt explains in chapter 10 of Schutt, Seidman, and Keshavan (2015), an added factor was that the founding fathers of sociology, such as Emile Durkheim and Max Weber, understood psychology and evolutionary biology as focusing primarily on the individual, to the exclusion of explicitly social factors. It was first the polemic discussion of group selection in modern evolutionary sciences (see, e.g., Wilson and Sober 1994) which sparked the sociologists' interest, as exemplified by the work on cultural evolution (Boyd and Richerson 1988, 2009).

(Rothman 1990). Faced with an increasing population of mental patients and starved for results, from the 1920s to the 1950s psychiatrists turned to physically invasive but poorly founded interventions ranging from electroshocks to lobotomy, which in the long run did little to improve the reputation of psychiatry (or the mental health of patients). In the second half of the twentieth century, as the recklessness of previous approaches became better understood, psychiatry adopted a sociological view of human behavior where many mental illnesses such as depression were viewed as having social roots. The growth of psychoanalysis, with its flat-out rejection of biology and its reliance on anecdotal data, further led to the expansion of the nonbiological view also for other, less-understood maladies like schizophrenia, which were assumed to arise from the patient's "mind" and (childhood) experiences, but to have no biological basis. Progress along this path was, of course, slow, and the discipline had to wait until the 1980s to see significant advances in treatment, thanks to the advent of pharmacotherapy, which brought psychiatry and medicine closer together. As a consequence, in the last decades of the twentieth century, psychiatry distanced itself from the sociological view in favor of a reliance on psychopharmaca.

Around the turn of the century, a simplified view of the situation was that both sociology and psychiatry were interested in human mental illnesses, but they had taken fundamentally opposing views. Sociology was firmly entrenched in a "macro" view, rejecting any neurobiological approach, while psychiatry had become a subdiscipline of medicine, focusing on the biological approach with an emphasis on psychopharmaca. Then, both disciplines started converging towards a middle point. There were two big reasons for this movement.

First, as discussed in section 2.2, discoveries related to epigenetics and neuroplasticity made clear that biological and social factors could not be clearly separated, and progress required a more integrative view. Second, there was a widespread impression that the purely neurobiological approach had failed to deliver on its promises regarding mental illness. In the late 1990s and early 2000s, improved understanding of neurotransmitters led to the development of a so-called second generation of antipsychotic medications, which were then extensively tested under the auspices and control of the National Institute of Mental Health. Alas, the new drugs were found to be no more effective than the first-generation ones (e.g., Lieberman et al. 2005). Likewise, titanic efforts within the framework of the Human Genome Project have so far failed to fully identify the genetic causes of illnesses such as schizophrenia. The *heritability* of schizophrenia, that is, the proportion of the disease's variation explained by genetic factors, is currently estimated at 0.7, which is relatively high, but genetics alone seems unable to explain its onset. In both cases, it became clear that something was missing from the picture. The bottom line of Schutt, Seidman, and Keshavan (2015) is that the next step is the study of the interaction between the brain and the social environment, and how each influences the other. However, this step should not be seen as synonymous with the confluence of sociology and psychiatry, for such an approach would run the risk of reinventing wheels that have already been running for some time in economics and psychology.

4. *The Social Brain*

At a very general level, "social brain" is just a short expression for the set of brain structures and functions related to the perception and evaluation of the social environment,

and how that perception and evaluation affects social decision making. Its function can be roughly identified with "social cognition and emotion," a description whose meaning has changed over the decades but which is currently broadly taken to describe the set of cognitive abilities needed to process and interpret social and emotional information arising both from oneself and others. The social brain is an evolutionarily recent development, involving an extensive network of brain regions. Current consensus indicates that its basic structure is built around frontotemporal pathways connecting frontal executive regions (in the prefrontal cortex) to phylogenetically older regions in and near the temporal lobes. The following subsections summarize some key insights on the evolution and structure of the human social brain.

4.1 *Evolution*

Evolution is not a highly efficient process, nor is it a mechanism designed to achieve a result. It is merely the result of blind selection forces among the alternatives that happen to be available by sheer chance (mutation), with no guarantee that an optimum will be reached in any reasonably defined space. Our brain is not a wonderful machine: it is a patchwork organ made through kludges, an electrical machine that is permanently wet and where the most important parts are built on the surface, where damage is more likely. In other words, an idealized, efficient view of evolution would have the brain acquire characteristics maximizing survival chances given the environment, never investing any effort in anything else. However, the reality is that each characteristic that is selected for might and often does have "unintended" consequences, which then produce behavior that is not necessarily adaptive, let alone maximally so. This has two important consequences. First, it produces a *bidirectionality* between the environment and the brain. Given the environment (physical and social), evolutionary forces shape the brain. But changes in the brain produce changes in behavior that alter the physical and social reality, and not always in an adaptive way. Second, a particular kludge in the brain might make certain failures and breakdowns particularly likely, especially if the environment changes. Such breakdowns are the domain of psychiatry, and they are eminently informative about the functioning of the brain. It is then easy to argue that psychiatric malfunctions related to social behavior provide invaluable information on the functioning of the social brain.

It is instructive to briefly summarize the milestones of the evolution of the human brain, as far as our current understanding allows, especially as it will result in some rather counterintuitive claims, and hopefully increased understanding of how our emotional and social mindset came to be. Chapter 3 of Schutt, Seidman, and Keshavan (2015), written by sociologist Jonathan H. Turner, provides a compact overview. The first milestone has to be placed 65 million years ago, when a meteorite impact wiped out the dinosaurs and many other of our planet's lifeforms. Small rodent-like mammals survived and took refugee in the arboreal habitats of Africa. The key realization about arboreal habitats is that they are eminently three-dimensional, as opposed to the plains: predators can also get at you from above and below. Adaptation to these habitats required a switch and substantial brain rewiring placing vision as the dominant sense (as opposed to smell or hearing). Incidentally, and as an accidental byproduct, the rewiring resulted in a substantial development of the so-called association cortices around the inferior parietal lobe, where the lobes associated with different senses meet. It turns out that this development is the fundamental prerequisite for the capacity for language. This happy accident explains why some of our cousins, such as the great apes, have the capacity for

language even though anatomical constraints prevent them from actually developing it.

The second milestone comes from the fact that our ancestors were defeated, in evolutionary terms. We are, quite literally, the descendants of losers who had to go through quite radical changes as a consequence of their defeat. The rodent-like mammals eventually evolved into two broad classes of primates, the apes and the monkeys. Naturally, they competed for resources in the arboreal habitats. Around 23 million years ago, our ancestors the apes were defeated by the monkeys and pushed away from the core regions of the arboreal habitats. The reason was one random mutation that allowed monkeys (but not apes) to digest unripe fruit, thereby consuming this resource before apes could access it. Our distant forefathers were forced to the fringes of the habitat, where vegetation is less dense and branches offer less support. Gone was the possibility to sustain large groups and, as a direct consequence, natural selection created individualistic apes, lacking the intuitive group-formation tendencies that monkeys have. As a second consequence, since fringe areas are far more dangerous (one false step leads to a fall, and predators might abound), survival selected in favor of increased dexterity and increased intelligence.

Around 10–14 million years ago, the African forests began to recede, and apes were forced onto the plains. Random evolutionary events had endowed them with relative intelligence, dexterity, and the potential for language, but had taken away any tendencies for stable group formation. The problems created by a highly developed brain without the stabilizing factor of group-formation instincts persist to this day. Gorillas, chimpanzees, and orangutans build notoriously unstable group structures (compared to other animals), and it is likely that this was also true for our ancestors. Humans are descended from individualistic apes with very weak social tendencies. Hence, the often-spouted claim that humans are social animals is, simply put, at odds with all available evidence, and seems to arise from "wishful thinking" related to cultural (Western) values. This is a fundamental realization for social neuroscience *and* economics. Of course, the human brain has adapted to facilitate social interactions. But this is a forced and problematic adaptation, which goes against a previous adaptation that took away social group-formation tendencies. It does not take much to expect trouble arising from the "social brain," since, in evolutionary terms, it amounts to a recent patch on top of an individualistic brain.

Ever since the great apes left the forest, groups have had an evolutionary advantage over individuals, and indeed while group-forming monkeys flourished, individualistic great apes have become almost extinct (except for humans). Humans are simply the great apes for which evolutionary forces were able to partially undo what had been previously done. Comparative analysis of the brains of humans and contemporary great apes suggests that this adaptation took place by altering existing structures, rather than developing new ones. This is specially true for *subcortical* structures, mostly unrelated to the development of intelligence. For instance, the human amygdala (which in lower animals is essentially the center for fear and anger) is twice as large as in great apes (controlling for body size), and the human septum (where sexual pleasure arises) is twice as large as in chimpanzees, even though the latter are far more sex-crazed than we are. As Jonathan Turner speculates, a possible explanation is that these centers have been hijacked by evolution in order to create a complex pattern of emotions as a kludge to sustain social stability and group formation in a species that previously evolved away from such tendencies. This provides a (controversial) view of emotions

as feeble attempts to replace long-gone group-formation tendencies and strengthen group ties. Mental health and well-being are closely related to emotions, and hence it is easy to see that many of the problems of contemporary humans arise from the particular evolutionary path followed by our ancestors.

4.2 *Theory of Mind and the Social Brain*

In evolutionary terms, hence, it could be argued that the "social brain" developed to allow for group formation and reap the benefits of cooperation in a species which had previously gone down the individualistic path. This view, however, is incomplete, for along the way, something amazing happened. The great apes started developing the capacity to *mentalize*, that is, the capacity to understand other individuals' minds. A classical mentalizing test is the so-called "false-belief paradigm." In this experimental setting, subjects observe a different actor who sees an object being placed inside location A. Then, the observer sees how somebody else moves the object from location A to location B, but, crucially, the observed actor does not witness this change. The test is whether the subject is able to anticipate that the observed actor will act on the basis of false beliefs, looking for the object in location A even though the subject knows that it is actually in location B. Passing the test shows that the subject is capable of understanding that other individuals might have different perceptions about the world. This capacity is absent in monkeys (and in human children below age four), but is not an exclusively human trait. Rather, it seems to be present, in varying degrees, in all species of apes (Povinelli and Bering 2002). Indeed, a very recent study has shown that great apes pass the classical false-belief test (Krupenye et al. 2016). Hence, it seems very likely that the mentalizing capacity (like the capacity for language) is one of the many evolutionary accidents that occurred as a consequence of our ancestors' troubles dating back to their expulsion from the forests by the victorious monkeys.[6]

The study of mentalizing, or theory of mind, is one of the great pillars of social neuroscience. The importance of the concept is rather easy to understand for an economist, as it essentially parallels the difference between decision theory and game theory. In the former, an economic agent should simply optimize his or her preferences, while in the latter it is necessary to develop a theory regarding other agents' actions. This is precisely what is captured by the general concept of theory of mind. It can be defined as the capacity to understand somebody else's mind, that is, to construct a mental representation of somebody else's motivation, intentions, knowledge, and beliefs (see, e.g., Singer and Tusche 2014). Hence, when economists attempt to give epistemic foundations to equilibrium concepts in interpersonal situations, they are essentially appealing to theory of mind.

It is important to distinguish theory of mind from related but fundamentally different concepts as, e.g., *empathy*. The latter refers to the capacity to share the feelings of others. For instance, psychopaths lack empathy, but are very good at theory of mind. They can cognitively understand other people's motivations and hence manipulate them, but will do so without regard to the other's possible suffering (Blair 2005, 2008). In this sense, the classical *Homo economicus*, let loose in a social situation, is nothing else than a psychopath. Symmetrically, modern economic models of "social preferences," where the welfare of others is taken into account, can be seen as an attempt to incorporate empathy into economics (but not necessarily theory of mind).

[6] Interestingly, some species of dogs are able to consider what a human can or cannot see, which might point to a rudimentary theory of mind (Kaminski et al. 2009 and Kaminski, Pitsch, and Tomasello 2013).

The term "social brain," however, does not yet correspond to a clearly defined concept at this point. Refining our previous attempt at a definition, it refers to the collection of brain regions that have been associated with social cognition, that is, our capacity to understand others' emotions, intentions, and motivations. Early studies (Adolphs 2001) listed a large number of regions involved in social reasoning, including large parts of the prefrontal cortex, the amygdala, and lateral regions (in the temporal cortex), and this is briefly discussed in chapter 2 of Schutt, Seidman, and Keshavan (2015), written by Matcheri S. Keshavan (see also chapter 5 of Cacioppo, Visser, and Pickett 2006), followed by a more detailed exposition in chapter 4 (by Paul G. Nestor, Victoria Coate, and Ashley Shirai). Essentially, the social brain appears to encompass a "frontotemporal link," that is, it arises as the result of the interaction of regions in the temporal lobe and regions in the prefrontal cortex. Recent research, however, suggests that what is called "the social brain" is actually a set of several, differentiated networks accomplishing related but differentiated functions. The theory-of-mind network (see figure 1) comprises the medial prefrontal cortex (MPFC), which is also part of the reward valuation network as discussed above, and a collection of regions in the temporal cortex: the temporoparietal junction (TPJ; we encountered this region previously as the probable origin of the P3b measured in the EEG[7]), the superior temporal sulcus (STS), and the temporal poles (TP). A number of other regions and subregions have also been ascribed to the network, but, as argued by Spunt and Adolphs (2014), this might be due to the large variety of tasks used to measure theory of mind. In a recent meta-analysis encompassing seventy-three fMRI studies, Schurz et al. (2014) found that a particular "core network" engaged for all theory-of-mind tasks, including the MPFC and the bilateral TPJ. Comparable results were obtained by Spunt and Adolphs (2014) with a new "why/how" task designed to single out the basic idea of mentalizing (strangely, the evidence of Spunt and Adolphs 2014 and Schurz et al. 2014 is not mentioned in Schutt, Seidman, and Keshavan 2015). These developments make clear that mentalizing appears to be an identifiable, separable process in the human brain, distinct from other social processes. For instance, the "empathy network" appears to be quite different, including the medial anterior cingulate cortex (essentially located "behind" and next to the MPFC) and the anterior insula, among other regions (Singer and Tusche 2014).

This is precisely a point where a discipline of neurosociopsychiatry might miss out a crucial insight. Microeconomics in general and game theory in particular is ripe with paradigms requiring theory of mind, even if the terminology is usually different. Neuroeconomic studies employing games as the ultimatum game or the prisoner's dilemma have targeted prefrontal regions as regions of interest, and have frequently observed activation of the MPFC. This is, of course, consistent with the involvement of the theory-of-mind network. For instance, Rilling et al. (2004) examined subjects playing the ultimatum game and the prisoner's dilemma game with both (alleged) human and computer partners, and observed stronger activation for human partners in typical theory-of-mind areas. Coricelli and Nagel (2009) used the beauty-contest game, which requires iterative thinking involving others, and found that high-level reasoning correlated with activity in the MPFC. This connection, which is not made explicit in Schutt, Seidman, and Keshavan (2015),

[7] Mars et al. (2012b) argue that little is known about the neuroanatomy of the TPJ. In particular, they suggest that only the posterior part of the TPJ should be considered a node in the social brain, while the anterior part interacts with areas more associated with attentional control.

is sufficient to make clear why social neuroscience is important for economics and vice versa.

It is also worth noting that the relation of mentalizing to recursive or iterative thinking might lie at the basis of an open question in social neuroscience. Interestingly, it has been shown that there are major overlaps between the theory-of-mind network and the default mode network or DMN (see figure 1). Specifically (see Mars et al. 2012a), the medial frontal cortex (including the MPFC) and the (area around the) TPJ are considered nodes of the DMN, and a further node, the posterior cingulate cortex, plays a role in attributing mental states to others. It is worth remembering that the DMN is typically identified by looking at brain areas whose activity increases during "rest" blocks, compared to specific activities. The two networks might hence seem to have quite unrelated functions, begging the question of what the reason for the significant overlap might be. However, far from being a "resting network," the DMN is associated with internal mentation and recursive thinking (for example, the DMN is active when planning, thinking about one's future, etc), functions which are presumably also important for social cognition and mentalizing.

4.3 *Social Valuation and Social Preferences*

Mentalizing is one of the main functions of the social brain, and probably the most relevant for economics. However, several other broad categories of interest fall squarely within the domain of social neuroscience (see, e.g., Schutt, Seidman, and Keshavan 2015, chapter 5). Also relevant for economics is social perception, which includes the perception, interpretation, and evaluation of social stimuli.

Indeed, the presence of the MPFC (or specific parts thereof) in the theory-of-mind network and, more generally, among the regions forming the substrate of the social brain, is especially relevant for economics. As mentioned in section 2.1, parts of this brain region can probably be conceived as encoding value, akin to economic utility, after receiving reinforcement-based feedback from the dopaminergic pathways. The question is whether the medial prefrontal cortex generally encodes all kinds of decision-related value or just a specific, self-centered kind. In more economic terms, is social decision making separate from, say, consumption-based decision making? In classical economics, one would start with the appropriate space of outcomes and define (complete) preferences on it. If outcomes include distributional allocations affecting other individuals, that is just one more dimension in the space, causing no formal-mathematical difficulty whatsoever to the approach. Implicitly, that "economic view" postulates a unique type of value computation, where the different outcomes (whether they incorporate the welfare of others or not) are compared using a globally defined preference relation. The literature on social preferences is eminently orthodox in this respect, for it simply proposes specific functional forms for such global preferences.

The question, however, is unresolved at this point. On the one hand, a *single-preference* theory naturally arising from the classical microeconomic approach would postulate that social and nonsocial valuation processes should be implemented in the same neural structures. In this case, in view of evidence from nonsocial decision making, the ventromedial prefrontal cortex is the natural candidate as the region aggregating the various neural signals into a coherent preference (or value). On the other hand, a *separated-preferences* theory would postulate that social valuations are achieved by a differentiated set of neural structures, possibly overlapping but not identical with the ones managing nonsocial valuations. This important question is strangely absent in Schutt, Seidman, and Keshavan (2015), but a good recent

review can be found in Ruff and Fehr (2014). Certain aspects of social interactions, as, e.g., receiving social approval or disapproval, seem to be linked to the same brain regions involved in the processing of nonsocial rewards. Further, several studies show that behavior conforming with normative social principles (e.g., equity or cooperation) leads to increased activity in reward-related areas, larger than that observed in control conditions with similar payoffs but no social framing (e.g., Rilling et al. 2002). Overall, evidence suggests that social behavior is controlled by a network that is functionally very close to the RVN. However, it is unclear whether it is actually the same network or, rather, social and nonsocial values are managed by overlapping regions or different subregions of the same areas (different, specialized neurons within the same broad area). The latter view agrees with single-neuron analysis from nonhuman primates (see Ruff and Fehr 2014 for a detailed review of the evidence).

4.4 *Emotion Processing and Regulation*

Two other broad functions of the social brain are emotion processing (including recognition and expression) and self-regulation, understood as the process by which humans regulate both their own internal emotional states and the influence of social and emotional stimuli on them. Regarding emotion processing and regulation, one of the key nodes in the relevant network, that received much attention in early neuroeconomic studies, is the amygdala. As mentioned in section 4.1 above, the human amygdala greatly increased in size as the result of our particular evolutionary path. It is, hence, not surprising that the amygdala plays multiple roles in the social brain. A number of studies have shown that patients with amygdala damage suffer from difficulties in processing emotional information, especially fear. Further, if damage occurs during developmental age, patients develop many abnormal patterns of social behavior (see, e.g., chapter 2 of Cacioppo, Visser, and Pickett 2006).

Chapter 5 of Schutt, Seidman, and Keshavan (2015), written by Christine Hooker, is mostly devoted to a review of the author's research on the role of the amygdala in social cognition and social behavior. Interestingly, this research shows that the amygdala also plays an important role in social learning, for instance when learning the emotional value of stimuli (including threat and reward values). That is, the amygdala seems to be crucial for creating associations between stimuli and learned emotional values. Social learning, however, can be maladaptive, in the sense that if the amygdala "overreacts," exaggerated associations (e.g., fear) can be reinforced, leading to anxiety disorders. The evidence reviewed by Hooker shows that exaggerated amygdala response is associated with the neuroticism personality trait, that is, people with high neuroticism ratings exhibit higher amygdala activity during social learning and are, hence, most likely at risk of developing anxiety disorders. Also, the social environment interacts with amygdala reactions. For instance, an emotionally reactive significant other can become a negative externality, leading to a magnification of inappropriate responses contributing to anxiety disorders, as, e.g., PTSD or phobias.

The self-regulation of emotional states is, of course, closely related to the cognitive control network. Hooker concentrates on couples' studies showing that the lateral prefrontal cortex (especially its ventral part) is crucial to regulate and inhibit amygdala reactions, e.g., allowing to recover from interpersonal conflicts. Deficits in this region's activity might be associated with disorders as social anhedonia, a state where pleasure from social relationships is diminished, which itself is related to schizophrenia-related symptoms.

4.5 *Social Networks and the Social Brain*

The structure of our immediate social environment is obviously an important determinant of our behavior and an important factor in our (mental) health. An extreme example is discussed in chapter 11 of Schutt, Seidman, and Keshavan (2015), where Bernice Pescosolido makes the point that differences in social network structure and connectedness are a major determinant in suicidal behavior. The natural question, however, is to what extent the feedback link from our social world to the social brain can be demonstrated and quantified. As mentioned in section 2.2, idiosyncratic experiences can result in structural changes to the brain. Is this also true for the social brain?

At the evolutionary level, the social brain hypothesis (Dunbar 1998; Dunbar and Shultz 2007) states that human brains have expanded to their current size precisely due to the demands of social decision making and social cognition. Supporting this hypothesis, it is known that, across primate species, the size of social groups correlates with frontal lobe volume. Further, if brain growth is linked (in evolutionary terms) to the demands of social cognition, it is maybe easier to understand that, as discussed above, the human default mode network might have evolved precisely to satisfy the demands of functions as mentalizing. Recent evidence, however, indicates that similar correlations might also hold across individuals, especially for specific brain regions corresponding to the theory-of-mind network. For instance, evidence from research on adult macaques involving pseudo-random assignment suggests that living in larger social groups might causally result in structural brain changes, affecting the STS, the ACC, and the (rostral) prefrontal cortex (Sallet et al. 2011). In a study with adult humans, Bickart et al. (2011) found that amygdala volume correlates with the size and complexity of individual social networks. Also, Lewis et al. (2011) showed that gray-matter volume in the posterior frontal poles, the STS, and the (left) TPJ co-vary with mentalizing capacity, while gray-matter volume in the medial orbitofrontal cortex and other areas vary both with mentalizing capacity and social-network size. These examples suggest a tight, bidirectional link between our immediate social environment and (social) brain structure.

As argued at length by Lieberman (2013) and others, as a result of the coevolution of our brain and our social environment, social connections are fundamental for mental health. Even if, as argued by Turner (see section 4.1), evolution needed to tinker quite a bit to turn individualistic apes into social humans, in our current state, social proximity and interaction appear to be a necessary condition for a well-functioning brain.

5. *The Failure of the Social Brain*

Given the bidirectional link between the social world and the social brain, a simple thought experiment then leads to the conclusion that large dysfunctions in social behavior are most likely associated with sharp structural brain anomalies, and going to these extreme cases might deliver information on the functioning and structure of the social brain. Hence we enter the domain of mental illnesses, where the social environment is perturbed because of a dysfunction of the social brain, and society's failures, where the social brain is damaged because of a dysfunction in the social world.

5.1 *Schizophrenia and Other Mental Illnesses*

Schizophrenia is a complex mental disorder that affects roughly one percent of the world population. Its most visible symptoms are of psychotic nature. Patients' thinking patterns become disorganized, they might suffer from hallucinations (e.g., hearing

voices), and they often lose contact with reality and develop delusions and false beliefs. Beyond such visible symptoms, schizophrenia is also associated with severe cognitive impairments affecting attention, working memory, and executive functions. Crucially, however, schizophrenia is characterized by what could be described as a breakdown of the social brain. Social cognition, including mentalizing, is severely impaired (patients might even experience difficulties interpreting others' facial expressions), followed by withdrawal and isolation. Cognitive impairments and deficits in social cognition appear to be interconnected, and are largely unresponsive to the medications used to treat the psychotic symptoms of the illness.

Schizophrenia is not the only mental illness related to dysfunctions of the social brain. Major depressive disorder is related to failures in the regulation of negative emotions, while mania is related to failures in the regulation of positive emotions. Autism and Asperger's Syndrome are associated with an extreme failure of theory of mind known as "mind blindness," the inability to attribute mental states to others. However, schizophrenia is particularly relevant for the study of the social brain due to the pervasive deficits in social cognition it entails. As a result, a rather large part of Schutt, Seidman, and Keshavan (2015) is devoted to schizophrenia and its symptoms.[8] In chapter 7 of that work, Junghe Lee, William P. Horan, and Michael F. Green describe the deficits in social cognition associated with this illness in detail. In chapter 8, Jean Addington and Mariapaola Barbato undertake a comparable endeavor for those at high risk of developing psychotic disorders (including schizophrenia). In chapter 5, Christine Hooker argues that a main process facilitating mentalizing is simulation, which, in lay terms, corresponds to putting yourself in somebody else's shoes. Simulation is known to be supported by the ventromedial prefrontal cortex (VMPFC), and evidence by Hooker shows that, in schizophrenia, the reduction of mentalizing skills is associated with a reduction of gray-matter volume in the VMPFC. Causality, however, is difficult to establish, especially in view of the bidirectionality between social behavior and the structure of the social brain.

Recent research seems to indicate that, at a higher level, schizophrenia is associated with altered connectivity patterns among brain regions part of the "social brain" described above. Chapter 4 of Schutt, Seidman, and Keshavan (2015), written by Paul G. Nestor, Victoria Coate, and Ashley Shirai, reviews the neuroanatomic origins of psychopathological symptoms associated with schizophrenia and other mental disorders. Comparing patients with healthy controls performing different social and nonsocial tasks, and analyzing gray versus white matter in schizophrenic patients, a number of abnormalities have been identified. It is, of course, not possible to point to a single region as the origin of the disorder, because schizophrenia exists in many different grades and patients' symptoms are far from uniform. However, two regularities are worth mentioning here, for they might provide insights toward a better understanding of the social brain.

The first important regularity is that schizophrenia seems to be associated with microstructural damage in the white matter of a region known as the cingulum bundle, which very rarely features as a region of interest in neuroeconomic studies. The reason is that the cingulum bundle is not a real "actor" in the brain, but rather an anatomical bridge, a physical white-matter pathway connecting the anterior cingulate cortex and the frontal regions of the brain (VMPFC, OFC, and DLPFC). In particular, the cingulum

[8] It should be mentioned, however, that this text's focus on schizophrenia might not be representative of the field. For example, the study of this illnesss was absent in the earlier text of Cacioppo, Visser, and Pickett (2006).

bundle is one of the pathways supporting the frontal-temporal communications that build the backbone of the social brain. The disturbance in the cingulum bundle might mean that what fails in this disease is not a particular region–function pair, but rather network connectivity. In schizophrenic patients, the ACC might fail to fully fulfill its role assisting value formation as derived from the dopaminergic pathways. This is consistent with evidence studying neural responses in the striatum, a central brain structure where dopaminergic pathways from the substantia nigra end. As discussed in section 2.1, the striatum is a key node in the RVN, and activity in this region has been shown to correlate with reward prediction. Studies with schizophrenic patients have shown abnormal striatum-activity patterns in response to reward prediction errors. The link to the social brain is that reward learning through prediction and prediction errors is fundamental for understanding the intentions and actions of others, and hence a compromised processing of reward prediction errors might lead to a disturbance in theory-of-mind processes. This line of reasoning might become part of an ultimate explanation for some of the most important symptoms of schizophrenia in the near future. However, the argument is far from entirely clear at this point. One problem is that activity in the striatum does not particularly reflect *social* prediction errors. Evidence by Behrens et al. (2008) shows that, while dopaminergic activity in the striatum is indeed related to nonsocial reward prediction, prediction errors for social tasks elicit heightened activity in different areas, most notably the superior temporal sulcus and the temporoparietal junction, which, as mentioned above, are candidate nodes for the theory-of-mind network. A major problem with the thesis defended by Nestor and coauthors in chapter 4 of Schutt, Seidman, and Keshavan (2015) is that it ignores this nuance. However, this most likely simply reflects the current gaps in our knowledge. First, schizophrenia implies cognitive deficits, and not only social-cognitive ones, possibly arising from general problems in processing reward prediction. Second, the absence of specific and systematic evidence at this point on similar abnormalities in regions specifically associated with social reward prediction is not evidence of absence, but merely points at a natural avenue for future research. Nestor and coauthors, however, might be (inadvertently) guilty of obscuring this point by basing their argument on the claim that "reinforcement learning via prediction error represents an all-purpose mechanism that is engaged in the processing of both social and nonsocial salient information" (Schutt, Seidman, and Keshavan 2015, pp. 91–2). As discussed in the previous section, this might be an excessive simplification at this point.

The second important regularity is related to a basic dichotomy in the symptoms of schizophrenia, which are divided into "positive" and "negative" ones. The terminology can be confusing, because all the symptoms are quite negative from a personal-welfare point of view. Positive ones are those that refer to behavior and perceptions present in patients but not in healthy humans, and vice versa for negative ones. For instance, hallucinations, delusions, and thought derailment are positive symptoms, while diminished levels of motivation and social responsiveness are negative symptoms. Evidence reviewed by Nestor and coauthors suggests that the dichotomy can be traced back to schizophrenia-induced volume reductions in two different regions of the temporal lobes. Positive symptoms might be associated with volume reductions in the superior temporal gyrus, which is just a brain-surface ridge along the wedge called superior temporal sulcus that we have already discussed. This is particularly interesting because, since such temporal structures are integral to the

theory-of-mind network, the observed volume reduction might provide a bridge to the mechanisms by which schizophrenia ultimately disrupts the social brain. In particular, the superior temporal gyrus might be crucial for correctly attributing the source of our own speech to ourselves, and not to an external source. Negative symptoms, on the other hand, might be associated with volume reductions in the fusiform gyrus, which is also part of the temporal lobe. This structure has been consistently associated with facial recognition, and hence evidence gives rise to the natural hypothesis that deficits in facial recognition are at the root of many of the social-brain disruptions induced by schizophrenia. Simply put (probably too simply), if stimuli as prediction errors cannot be reliably associated with social actors, the very process by which we successfully interact with our social environment might break down.

5.2 *Orphans and Orphanages: Society's Failure*

Schizophrenia and other mental disorders serve as case studies for how the breakdown of the social brain results in impaired social behavior. An unfortunate series of natural experiments serves as illustration of the reverse link, that is, how a disruption of the social environment results in brain abnormalities. It has been known for a long time that, as a whole, children raised in orphanages exhibit cognitive impairments (IQs lower than the normative mean, impaired global cognition, language difficulties) and behavioral problems (e.g., inattention and overactivity). This is, in itself, not surprising, for the link between reduced maternal care and abnormal behavior in adult age has been demonstrated even in rodents (see the review by Meaney 2001 or the one-page summary in chapter 11, p. 249, of Schutt, Seidman, and Keshavan 2015). Chapter 12 of Schutt, Seidman, and Keshavan (2015), written by Michael E. Behen and Harry T. Chugani, reviews evidence on brain development of human children raised in orphanages, focusing on their own work, and including analysis of a relatively large sample ($N = 156$) of children raised in Eastern European and Asian orphanages and later adopted in the United States. Other studies work with specific, "standard" samples such as the European and Romanian Adoptees (ERA) sample or the Bucharest Early Intervention Project (BEIP). Across samples, the evidence speaks a clear language. There is a small window after birth (no more than six months) in which where the child is raised has little or no effect. After that, negative effects appear and become more severe with the length of time spent in the orphanage.

Effects are of two kinds: general cognitive impairment and specific neurocognitive deficits ranging from attention and impulse-control problems to language-processing, memory, and manual-dexterity impairments. For example, roughly one third of the children in Behen and Chugani's sample had severely impaired global intellect, one third had no or minor global-intellect impairment but exhibited at least one specific neurocognitive deficit, and only one third displayed no problems. However, this classification was correlated with time spent in the orphanage, with a later age of adoption associated with a larger likelihood of developing problems. A battery of brain measurements (PET, MRI, EEG) has been carried out in different subsamples compared to different control groups. The evidence reviewed by Behen and Chugani, mostly derived from PET and MRI studies of adopted children at sample-average ages eight to ten, shows dysfunctions and structural anomalies in several brain regions. Those dysfunctions affect brain regions including the orbitofrontal cortex and lateral and medial temporal structures (the latter are involved in language functions). Volumetric studies show increased amygdala volume and reduced

cortical thickness. MRI studies allowing to estimate the integrity of white-matter pathways show aberrant connectivities and abnormalities in limbic neural pathways. The areas involved are consistent with the brain functions affected by known impairments in orphanage-reared children. Further, results usually follow patterns showing a monotonicity in length of institutionalization. Taken together, the evidence suggests that the cognitive and behavioral problems arising from being raised in an orphanage result directly from changes in the children's brains at early age.

Evidence from the EEG is especially interesting, because, unlike the fMRI, measurements in the EEG are perfectly feasible for very young children. Parker and Nelson (2005) and Parker, Nelson, and the Bucharest Early Intervention Project Core Group (2005) conducted ERP studies on facial recognition and discrimination (recall the N170 discussed in section 2.3) and found atypical patterns in orphanage-raised children (age thirty months). Since facial recognition can be seen as a basic building block for our perception of the social world, this evidence points to a very early dysfunction in the social brain. Further, in addition to ERP studies, the analysis of EEG frequencies is also informative. Different studies in different samples have found abnormalities in EEG frequency patterns, which have in turn been associated with inattentive-overactive symptoms.

It is unclear at this point whether the damage is reversible. Different studies (see Behen and Chugani's chapter for a detailed discussion) have shown that, for instance, EEG patterns seem to approach normality after adoption for children placed earlier in foster care. Using a randomized controlled trial, Nelson et al. (2007) showed that Romanian orphans placed into foster care experienced improvements in cognition and behavior, compared to children who remained institutionalized. Such results point out that the plasticity of the brain might help repair some of the damage. However, neuroplasticity has its limits, and it seems likely that, in many cases, at least part of the effects of orphanage rearing in the children's brains persist until adult age.

There are two main causes for the abnormal development of orphans' brains, and each of them entails a lesson. The first is simply that orphanage rearing entails early social deprivation. As part of normal brain development, synapses that are not used often enough in the children's brain are pruned. A socially deprived child will not develop a healthy social brain. Absence of a consistent primary caregiver after six months of age has severe negative consequences, revealing the folly of various social-design experiments and ideologies along human history. The second cause is stress. Even the best-run orphanage places children under prolonged stress, compared to a stable family environment. Many of the brain regions and pathways showing dysfunctions in the studies mentioned above are known to be damaged by prolonged stress.[9] Of course, unstable family environments will also generate stress, showing that the decision to bring a child into a problem-plagued family might require careful consideration.[10]

[9] Increased stress levels and exaggerated responses to stress might go hand-in-hand in this case. Evidence from rodents (Meaney and Szyf 2005) shows that maternal grooming causes differences in DNA methylation patterns and affects a glucocorticoid receptor gene promoter in the hippocampus, whose expression in turn affects responses to stress.

[10] There is, however, a caveat to all studies on orphanage-raised children. Such children are often affected by other factors that might impair brain development, such as pre- and postnatal malnutrition or alcohol exposure. Appropriate statistical controls (for instance, through body-mass index) do not, in general, affect the findings, but, by the very nature of the data, the potential confound cannot always be ruled out.

5.3 *Experience-Induced Changes in the Social Brain*

The example of orphanage-reared children is certainly extreme, but also highly provocative. A natural question is whether the feedback link from the social world to the social brain might also be demonstrated in less extreme cases. A look at some of the more sociological chapters in Schutt, Seidman, and Keshavan (2015) provides some insight.

Chapter 13, by Patrick Sharkey and Robert Sampson, reviews the evidence on negative effects of neighborhood violence on the cognitive skills of US schoolchildren. It is natural to speculate that neighborhood violence is associated with exposure to chronic environmental stress, which is, e.g., known to lead to an elevated risk of mental-health problems. Given the relatively high levels of everyday violence in the United States compared to other Western countries and the large variation across neighborhoods, this is a promising example for analyzing (negative) feedback links from the social environment to the social brain. Relying on the National Longitudinal Study of Adolescent Health, Harding (2009) found that the negative effects of living in a disadvantaged neighborhood on developmental outcomes (e.g., high-school graduation or teenage pregnancy) were mediated by the level of violence in the neighborhood. This is aligned with recent work by Patrick Sharkey and coauthors reviewed in the chapter, which analyzed variation in the timing of exposure to violent incidents among children within given communities. Children evaluated within a week of a homicide occurring close to their homes performed substantially worse in assessments of cognitive skills and were less likely to pass standardized exams. This evidence is thought provoking, but the relation to actual alterations of the social brain remains speculative since there is so far no direct, systematic neurological evidence on children exposed to violence incidents.

Chapter 14, by sociologist Allan V. Horwitz, raises a caveat. In a mostly historical review, he argues that some intense but "normal" human reactions might temporarily lead to symptoms typically associated with mental disorders, but not actually constitute disorders. An example is bereavement, i.e. sadness due to the death of a family member or friend. Although in some cases bereavement might actually lead to pathological depression, for most people it is a temporary condition. The point of the chapter is that the "normal" social brain is prepared to deal with intense negative experiences without necessarily experiencing long-run effects or, by extension, structural changes. This is a reasonable warning for further developments in this area but, as before, actual neurological studies are scarce. In stark contrast to this position, in chapter 9, psychologist Jill M. Hooley warns that even a mild aversive stimulus such as personal criticism is associated with increased relapse risk for patients with depression or schizophrenia. Of course, both positions are compatible, pointing just to the difference between a "normal" brain and those of individuals vulnerable to psychopathological disorders.

On a more positive note, the psychiatry-oriented chapter 15, by Shaun Eack and Matcheri Keshavan, focuses on the possibility of using neuroplasticity to improve cognitive and social-cognitive functions in patients with schizophrenia. Previous uses of neuroplasticity have already produced promising therapies for other patients. For instance, stroke patients with impaired motor functions can benefit from repeated forced movements of affected limbs, and patients with dyslexia show improvements in language skills following repeated auditory treatment. Eack and Keshavan report on a specific eighteen-month treatment known as "cognitive enhancement therapy" (Hogarty

et al. 2004), in which schizophrenic patients receive cognitive training (computer-based exercises to enhance attention, memory, and problem solving) and social-cognitive group therapy addressing such key elements of social cognition as, e.g., emotional processing, perspective taking, or the reading of nonverbal cues. The results of randomized controlled trials indicate improvement in social and nonsocial cognition. Part of the patients completed MRI procedures, and for those it could be observed that gray-matter volume in medial–temporal areas associated with social cognition was maintained, while in the case of controls it exhibited a reduction that is typical for schizophrenia patients. These results illustrate that a supportive, carefully structured social environment can, to a certain extent, repair the neurological problems related to social cognition in schizophrenic patients.

6. *Discussion: Where is Economics?*

6.1 *What Social Neuroscience Can Learn from Economics*

The study of the social mind naturally attracts psychiatrists and sociologists—the former due to the importance of certain brain disorders for social neuroscience, and the latter for their focus on the social environment. The cross-disciplinarity exemplified by Schutt, Seidman, and Keshavan (2015) is a welcome development, but its focus is markedly different from the one envisioned by Cacioppo and Berntson (1992) and Cacioppo, Visser, and Pickett (2006). The latter naturally placed social neuroscience in a habitat that has given rise to natural developments such as neuroeconomics or the tighter integration of psychology and behavioral economics. The former seems to seek a new, self-contained identity in which economics (or even, say, social and cognitive psychology) play little or no role.

This is regrettable, for economics has much to offer to social neuroscience, on several fronts. First, as mentioned above, there is a remarkable parallel between the study of interpersonal decision making, that is, game theory, and the development of theory of mind, one of the main pillars of social neuroscience. Social neuroscience is to decision neuroscience as game theory is to decision theory. It is no coincidence that the regions of interest in neuroeconomic studies employing game-theoretic paradigms and studies of theory of mind are pointing to the same network, for the very starting point of both is simply "thinking about thinking." As commented above, theory-of-mind studies have suffered from the lack of a standardized tasks (Spunt and Adolphs 2014; Schurz et al. 2014), to the point that the exact nature and components of the underlying neural network remain unclear. One of the strengths of behavioral game theory is precisely the discipline in paradigm development. Formulating a paradigm as a well-defined game allows one to extract its essence and naturally disciplines research. Further, once the essence is extracted, it is a simple matter to take the next natural step. For instance, while social neuroscience is still "thinking about thinking," standard game-theoretic paradigms already capture "thinking about thinking about thinking about. . ." and so on. Researchers of the theory of mind might find it useful to turn their attention to well-established paradigms as the prisoner's dilemma, the beauty-contest game, or even coordination games.

Second, many of the ideas being informally discussed in social neuroscience can be given a formal-analytical foundation that would help disentangle speculation from deduction. Indeed, although social neuroscientists seem to be mostly unaware of them, such models already exist. Consider, for instance, the evolution of the brain. Social neuroscience attempts to understand the appearance

of intelligence and the social brain in terms of evolution, depicted as a series of responses to exogenously given changes in the environment. However, evolution has its own dynamics, which is not always easy to pin down in the absence of a formal model. For instance, Robson and Kaplan (2003) explained the evolution of intelligence in hunter–gatherer societies as a purely economic phenomenon. Young humans consume much more than they produce, creating a debt that has to be repaid in a later life phase with enhanced food production. This "life-cycle strategy" involves significant learning, which is made more economically efficient by increased intelligence. Robson and Kaplan (2003) consider growth models in which the brain is treated as somatic capital, with evolution selecting for the size leading to the steady-state growth rate. Simple comparative statics then explains how the expansion of the African Savannah resulted in increased brain productivity and led to increases in both intelligence and life expectancy. Further, evolutionary arguments also help explain the coevolution of intelligence and theory of mind. As argued by Robson (2002), strategic intelligence could be caught in a "red-queen race" where being able to predict the actions and intentions of others confers an advantage, but since this is true for others as well, higher and higher levels of sophistication become necessary. At a different level, the axiomatic approach typical of microeconomics can be fruitfully applied to understand the difference between "as if" concepts such as "value" and the actual properties of decision making in the brain, both individual and interpersonal. For instance, Caplin and Dean (2008) provide an axiomatic foundation of "dopamine release functions" capturing the link between dopamine and the reward prediction error as the basis for decision making.

Last but not least, economics can contribute to social neuroscience at the technical level. The analysis of brain data is becoming increasingly complex, as illustrated by the difficulties encountered by fMRI researchers. New techniques and the combination of existing ones (such as, e.g., recent studies integrating EEG and fMRI measurements) will require even more sophisticated analyses and explicit statistical models. Structural models from econometrics might prove invaluable in this field. Further, it is obvious that network analysis will play an important part in the future of neuroscience (and social neuroscience is already making extensive use of it), and economists can hit the ground running in this particular race, as analytical models of networks have already been integrated into several subdisciplines within microeconomics.

6.2 *What Economics Can Learn from Social Neuroscience*

Even though a large part of economic behavior can be understood through decision theory, the ultimate objects of study of economics are markets, institutions, and groups. The study of reward valuations is an important milestone for (neuro)economics, but most economic decisions are made in interpersonal settings. Independently of whether social neuroscientists pay attention to economics or not, economists need to pay attention to social neuroscience. The simplest definition of the social brain might be the set of brain functions underlying interpersonal decisions and hence, the object of study of social neuroscience provides the foundation of economic decisions.

The conceptual links between both disciplines have hopefully been made clear in the previous sections. There are, however, more proximate payoffs to the study of social neuroscience. Recent trends in microeconomics and behavioral economics have focused on individual differences, ranging from gender to cognitive sophistication and personality traits, as a way to improve our understanding

of economic decisions and especially deviations from normative rationality. At the very least, social neuroscience can be an invaluable source of information on individual differences on interpersonal economic decision making, and such insights can be of interest even for economists not interested in the neural foundations of behavior (recall, for instance, the link between trait neuroticism, easily measured through the Big-Five questionnaire, and amygdala function mentioned in section 4.4).

It should be kept in mind that developments in neuroscience are providing a refreshing foundation for classical economic concepts such as utility. It would be naïve to expect, though, that the full neoclassical apparatus will be ultimately justified. Rather, evidence from neuroscience will be the ultimate judge as to the direction in which basic economic models should be developed, changed, and adapted to increase their realism. For instance, a key question at this point is whether the same value can be postulated for social and nonsocial decisions. While one can assume either a single, unified value or two orthogonal ones and develop economic theories accordingly, the answer should come from the data, but choice data alone cannot settle this particular question.

In spite of popular beliefs, economics has already moved past the neoclassical dream of all-knowing, optimizing, self-interested agents. Several decades have passed since the first microeconomic models incorporating bounded rationality found their way into the mainstream. Psychology played a fundamental role in identifying heuristics, biases, and behavioral rules of interest. Later, behavioral economics challenged the self-interest assumption, and started opening the black box of human motivation, often while essentially keeping the parsimonious assumption of preference-maximizing agents. Insights from social neuroscience can be invaluable for the next steps as economics advances towards fully descriptive models of human behavior, incorporating bounded rationality, departures from self-interest, and strategic behavior. After all, if economics is a social science, then its object of study is, directly or indirectly, the social brain.

References

Achtziger, Anja, Carlos Alós-Ferrer, Sabine Hügelschäfer, and Marco Steinhauser. 2014. "The Neural Basis of Belief Updating and Rational Decision Making." *Social Cognitive and Affective Neuroscience* 9 (1): 55–62.

Achtziger, Anja, Carlos Alós-Ferrer, Sabine Hügelschäfer, and Marco Steinhauser. 2015. "Higher Incentives Can Impair Performance: Neural Evidence on Reinforcement and Rationality." *Social Cognitive and Affective Neuroscience* 10 (11): 1477–83.

Adolphs, Ralph. 2001. "The Neurobiology of Social Cognition." *Current Opinion in Neurobiology* 11 (2): 231–39.

Amodio, David M., Bruce D. Bartholow, and Tiffany A. Ito. 2014. "Tracking the Dynamics of the Social Brain: ERP Approaches for Social Cognitive and Affective Neuroscience." *Social Cognitive and Affective Neuroscience* 9 (3): 385–93.

Bannister, Robert C., Jr. 1973. "William Graham Sumner's Social Darwinism: A Reconsideration." *History of Political Economy* 5 (1) 89–109.

Behrens, Timothy E. J., Laurence T. Hunt, Mark W. Woolrich, and Matthew F. S. Rushworth. 2008. "Associative Learning of Social Value." *Nature* 456 (7219) 245–49.

Bickart, Kevin C., Christopher I. Wright, Rebecca J. Dautoff, Bradford C. Dickerson, and Lisa Feldman Barrett. 2011. "Amygdala Volume and Social Network Size in Humans." *Nature Neuroscience* 14 (2): 163–64.

Blair, Robert James R. 2005. "Responding to the Emotions of Others: Dissociating Forms of Empathy through the Study of Typical and Psychiatric Populations." *Consciousness and Cognition* 14 (4): 698–718.

Blair, Robert James R. 2008. "Fine Cuts of Empathy and the Amygdala: Dissociable Deficits in Psychopathy and Autism." *Quarterly Journal of Experimental Psychology* 61 (1): 157–70.

Boyd, Robert, and Peter J. Richerson. 1988. *Culture and the Evolutionary Process*. Chicago and London: University of Chicago Press.

Boyd, Robert, and Peter J. Richerson. 2009. "Culture and the Evolution of Human Cooperation." *Philosophical Transactions of the Royal Society B: Biological Sciences* 364 (1533): 3281–88.

Buckner, Randy L., Jessica R. Andrews-Hanna, and Daniel L. Schacter. 2008. "The Brain's Default Network: Anatomy, Function, and Relevance to Disease." *Annals of the New York Academy of Sciences*

1124 (1): 1–38.
Cacioppo, John T. and Gary G. Berntson. 1992. "Social Psychological Contributions to the Decade of the Brain: Doctrine of Multilevel Analysis. *American Psychologist* 47 (8) 1019–28.
Cacioppo, John T., Stephen L. Crites, Jr., Gary G. Berntson, and Michael G. H. Coles. 1993. "If Attitudes Affect How Stimuli Are Processed, Should They Not Affect the Event-Related Brain Potential?" *Psychological Science* 4 (2) 108–12.
Cacioppo, John T., Penny S. Visser, and Cynthia L. Pickett, eds. 2006. *Social Neuroscience: People Thinking about Thinking People.* Cambridge and London: MIT Press.
Camerer, Colin, George Loewenstein, and Drazen Prelec. 2005. "Neuroeconomics: How Neuroscience Can Inform Economics." *Journal of Economic Literature* 43 (1): 9–64.
Caplin, Andrew, and Mark Dean. 2008. "Dopamine, Reward Prediction Error, and Economics." *Quarterly Journal of Economics* 123 (2): 663–701.
Coricelli, Giorgio, and Rosemarie Nagel. 2009. "Neural Correlates of Depth of Strategic Reasoning in Medial Prefrontal Cortex." *Proceedings of the National Academy of Sciences* 106 (23): 9163–68.
Degler, Carl N. 1991. *In Search of Human Nature: The Decline and Revival of Darwinism in American Social Thought*. New York and Oxford: Oxford University Press.
Dougherty, Michael R. P., Charles F. Gettys, and Eve E. Ogden. 1999. "MINERVA–DM: A Memory Processes Model for Judgments of Likelihood." *Psychological Review* 106 (1): 180–209.
Dunbar, Robin I. M. 1998. "The Social Brain Hypothesis." *Evolutionary Anthropology* 6 (5): 178–90.
Dunbar, Robin I. M., and Susanne Shultz. 2007. "Evolution in the Social Brain." *Science* 317 (5843): 1344–47.
Edwards, Ward. 1982. "Conservatism in Human Information Processing." In *Judgment Under Uncertainty: Heuristics and Biases*, edited by Daniel Kahneman, Paul Slovic, and Amos Tversky, 359–69. Cambridge and New York: Cambridge University Press.
Eklund, Anders, Thomas E. Nichols, and Hans Knutsson. 2016. "Cluster Failure: Why fMRI Inferences for Spatial Extent Have Inflated False-Positive Rates." *Proceedings of the National Academy of Sciences* 113 (28): 7900–7905.
Ernst, Aurélie, and Jonas Frisén. 2015. "Adult Neurogenesis in Humans—Common and Unique Traits in Mammals." *PLoS Biology* 13 (1): 1–12.
Fehr, Ernst, and Colin F. Camerer. 2007. "Social Neuroeconomics: The Neural Circuitry of Social Preferences." *Trends in Cognitive Sciences* 11 (10): 419–27.
Fehr, Ernst, and Antonio Rangel. 2011. "Neuroeconomic Foundations of Economic Choice—Recent Advances." *Journal of Economic Perspectives* 25 (4): 3–30.
Fehr, Thorsten. 2013. "A Hybrid Model for the Neural Representation of Complex Mental Processing in the Human Brain." *Cognitive Neurodynamics* 7 (2): 89–103.
Fehr, Thorsten, and Manfred Herrmann. 2015. "Can Modular Psychological Concepts Like Affect and Emotion Be Assigned to a Distinct Subset of Regional Neural Circuits?: Comment on 'The Quartet Theory of Human Emotions: An Integrative and Neurofunctional Model' by S. Koelsch et al." *Physics of Life Reviews* 13: 47–49.
Fodor, Jerry A. 1983. *The Modularity of Mind*. Cambridge and London: MIT Press.
Fuster, Joaquín M. 2009. "Cortex and Memory: Emergence of a New Paradigm." *Journal of Cognitive Neuroscience* 21 (11): 2047–72.
Fuster, Joaquín M., and Steven L. Bressler. 2012. "Cognit Activation: A Mechanism Enabling Temporal Integration in Working Memory." *Trends in Cognitive Sciences* 16 (4): 207–18.
Gigerenzer, Gerd, and Ulrich Hoffrage. 1995. "How To Improve Bayesian Reasoning without Instruction: Frequency Formats." *Psychological Review* 102 (4): 684–704.
Gläscher, Jan, et al. 2012. "Lesion Mapping of Cognitive Control and Value-Based Decision Making in the Prefrontal Cortex." *Proceedings of the National Academy of Sciences* 109 (36): 14681–86.
Grob, Gerald N. 1994. *The Mad Among Us: A History of the Care of America's Mentally Ill*. New York: Simon and Schuster, Free Press.
Harding, David J. 2009. "Collateral Consequences of Violence in Disadvantaged Neighborhoods." *Social Forces* 88 (2): 757–84.
Hogarty, Gerard E., et al. 2004. "Cognitive Enhancement Therapy for Schizophrenia: Effects of a 2–Year Randomized Trial on Cognition and Behavior." *Archives of General Psychiatry* 61 (9): 866–76.
Hügelschäfer, Sabine, Alexander Jaudas, and Anja Achtziger. 2016. "Detecting Gender Before You Know It: How Implementation Intentions Control Early Gender Categorization." *Brain Research* 1649 (Part A): 9–22.
Ito, Tiffany A., and Geoffrey R. Urland. 2003. "Race and Gender on the Brain: Electrocortical Measures of Attention to the Race and Gender of Multiply Categorizable Individuals." *Journal of Personality and Social Psychology* 85 (4): 616–26.
Ito, Tiffany A., Erin Thompson, and John T. Cacioppo. 2004. "Tracking the Timecourse of Social Perception: The Effects of Racial Cues on Event-Related Brain Potentials." *Personality and Social Psychology Bulletin* 30 (10): 1267–80.
Kable, Joseph W., and Paul W. Glimcher. 2009. "The Neurobiology of Decision: Consensus and Controversy." *Neuron* 63 (6): 733–45.
Kaminski, Juliane, Andrea Pitsch, and Michael Tomasello. 2013. "Dogs Steal in the Dark." *Animal Cognition* 16 (3): 385–94.
Kaminski, Juliane, Juliane Bräuer, Josep Call, and Michael Tomasello. 2009. "Domestic Dogs Are Sensitive to a Human's Perspective." *Behaviour* 146 (7): 979–98.

Krakauer, John W., Asif A. Ghazanfar, Alex Gomez-Marin, Malcolm A. MacIver, and David Poeppel. 2017. "Neuroscience Needs Behavior: Correcting a Reductionist Bias." *Neuron* 93 (3): 480–90.

Krupenye, Christopher, Fumihiro Kano, Satoshi Hirata, Josep Call, and Michael Tomasello. 2016. "Great Apes Anticipate that Other Individuals Will Act According to False Beliefs." *Science* 354 (6308): 110–14.

Lewis, Penelope A., Roozbeh Rezaie, Rachel Brown, Neil Roberts, and R. I. M. Dunbar. 2011. "Ventromedial Prefrontal Volume Predicts Understanding of Others and Social Network Size." *NeuroImage* 57 (4): 1624–29.

Lieberman, Jeffrey A., et al. 2005. "Effectiveness of Antipsychotic Drugs in Patients with Chronic Schizophrenia" *New England Journal of Medicine* 353 (12): 1209–23.

Lieberman, Matthew D. 2013. *Social: Why Our Brains Are Wired to Connect*. New York: Random House, Crown.

Lindquist, Kristen A., and Lisa Feldman Barrett. 2012. "A Functional Architecture of the Human Brain: Emerging Insights from the Science of Emotion." *Trends in Cognitive Sciences* 16 (11): 533–40.

Maguire, Eleanor A., Neil Burgess, James G. Donnett, Richard S. J. Frackowiak, Christopher D. Frith, and John O'Keefe. 1998. "Knowing Where and Getting There: A Human Navigation Network." *Science* 280 (5365): 921–24.

Maguire, Eleanor A., et al. 2000. "Navigation-Related Structural Change in the Hippocampi of Taxi Drivers." *Proceedings of the National Academy of Sciences* 97 (8): 4398–403.

Maguire, Eleanor A., Katherine Woollett, and Hugo J. Spiers. 2006. "London Taxi Drivers and Bus Drivers: A Structural MRI and Neuropsychological Analysis." *Hippocampus* 16 (12): 1091–101.

Mars, Rogier B., Franz-Xaver Neubert, MaryAnn P. Noonan, Jérôme Sallet, Ivan Toni, and Matthew Rushworth. 2012a. "On the Relationship Between the 'Default Mode Network' and the 'Social Brain'." *Frontiers in Human Neuroscience* 6 (189): 1–9.

Mars, Rogier B., Jérôme Sallet, Urs Schüffelgen, Saad Jbabdi, Ivan Toni, and Matthew F. S. Rushworth. 2012b. "Connectivity-Based Subdivisions of the Human Right 'Temporoparietal Junction Area': Evidence for Different Areas Participating in Different Cortical Networks." *Cerebral Cortex* 22 (8): 1894–903.

McEwen, Bruce S. 2012. "Brain on Stress: How the Social Environment Gets under the Skin." *Proceedings of the National Academy of Sciences* 109 (Supplement 2): 17180–85.

Meaney, Michael J. 2001. "Maternal Care, Gene Expression, and the Transmission of Individual Differences in Stress Reactivity across Generations." *Annual Review of Neuroscience* 24 (1): 1161–92.

Meaney, Michael J., and Moshe Szyf. 2005. "Environmental Programming of Stress Responses through DNA Methylation: Life at the Interface between a Dynamic Environment and a Fixed Genome." *Dialogues in Clinical Neuroscience* 7 (2): 103–23.

Mirowsky, John, and Catherine E. Ross. 2003. *Social Causes of Psychological Distress*, Second edition. New York: Aldine de Gruyter.

Naumova, Oksana Yu, Maria Lee, Roman Koposov, Moshe Szyf, Mary Dozier, and Elena L. Grigorenko. 2012. "Differential Patterns of Whole-Genome DNA Methylation in Institutionalized Children and Children Raised by Their Biological Parents." *Development and Psychopathology* 24 (1): 143–55.

Navon, David. 1978. "The Importance of Being Conservative: Some Reflections on Human Bayesian Behaviour." *British Journal of Mathematical and Statistical Psychology* 31 (1): 33–48.

Nelson, Charles A., III, Charles H. Zeanah, Nathan A. Fox, Peter J. Marshall, Anna T. Smyke, and Donald Guthrie. 2007. "Cognitive Recovery in Socially Deprived Young Children: The Bucharest Early Intervention Project." *Science* 318 (5858): 1937–40.

Nieuwenhuis, Sander, Gary Aston-Jones, and Jonathan D. Cohen. 2005. "Decision Making, the P3, and the Locus Coeruleus–Norepinephrine System." *Psychological Bulletin* 131 (4): 510–32.

Padoa-Schioppa, Camillo, and John A. Assad. 2006. "Neurons in the Orbitofrontal Cortex Encode Economic Value." *Nature* 441 (7090): 223–26.

Parker, Susan W., and Charles A. Nelson. 2005. "An Event-Related Potential Study of the Impact of Institutional Rearing on Face Recognition." *Development and Psychopathology* 17 (3): 621–39.

Parker, Susan W., Charles A. Nelson, and the Bucharest Early Intervention Project Core Group. 2005. "The Impact of Early Institutional Rearing on the Ability to Discriminate Facial Expressions of Emotion: An Event-Related Potential Study." *Child Development* 76 (1): 54–72.

Povinelli, Daniel J., and Jesse M. Bering. 2002. "The Mentality of Apes Revisited." *Current Directions in Psychological Science* 11 (4): 115–19.

Ratner, Kyle G., and David M. Amodio. 2013. "Seeing 'Us vs. Them': Minimal Group Effects on the Neural Encoding of Faces." *Journal of Experimental Social Psychology* 49 (2): 298–301.

Rilling, James K., David A. Gutman, Thorsten R. Zeh, Giuseppe Pagnoni, Gregory S. Berns, and Clinton D. Kilts. 2002. "A Neural Basis for Social Cooperation." *Neuron* 35 (2): 395–405.

Rilling, James K., Alan G. Sanfey, Jessica A. Aronson, Leigh E. Nystrom, and Jonathan D. Cohen. 2004. "The Neural Correlates of Theory of Mind within Interpersonal Interactions." *NeuroImage* 22 (4): 1694–703.

Robson, Arthur, J. 2002. "Evolution and Human Nature." *Journal of Economic Perspectives* 16 (2): 89–106.

Robson, Arthur, J., and Hillard S. Kaplan. 2003. "The Evolution of Human Life Expectancy and Intelligence in Hunter–Gatherer Economies." *American Economic Review* 93 (1): 150–69.

Rothman, David J. 1990. *The Discovery of the Asylum: Social Order and Disorder in the New Republic*, Second edition. Boston: Little, Brown.
Ruff, Christian C., and Ernst Fehr. 2014. "The Neurobiology of Rewards and Values in Social Decision Making." *Nature Reviews: Neuroscience* 15 (8) 549–62.
Sallet, Jérôme, et al. 2011. "Social Network Size Affects Neural Circuits in Macaques." *Science* 334 (6056): 697–700.
Schleim, Stephan, and Jonathan P. Roiser. 2009. "fMRI in Translation: The Challenges Facing Real-World Applications." *Frontiers in Human Neuroscience* 3: 1–7.
Schultz, Wolfram. 1998. "Predictive Reward Signal of Dopamine Neurons." *Journal of Neurophysiology* 80 (1): 1–27.
Schultz, Wolfram. 2010. "Dopamine Signals for Reward Value and Risk: Basic and Recent Data." *Behavioral and Brain Functions* 6 (24): 1–9.
Schurz, Matthias, Joaquim Radua, Markus Aichhorn, Fabio Richlan, and Josef Perner. 2014. "Fractionating Theory of Mind: A Meta-analysis of Functional Brain Imaging Studies." *Neuroscience and Biobehavioral Reviews* 42: 9–34.
Schutt, Russell K., Larry J. Seidman, and Matcheri S. Keshavan, eds. 2015. *Social Neuroscience: Brain, Mind, and Society*. Cambridge and London: Harvard University Press.
Singer, Tania, and Anita Tusche. 2014. "Understanding Others: Brain Mechanisms of Theory of Mind and Empathy." In *Neuroeconomics: Decision Making and the Brain*, Second Edition, edited by Paul W. Glimcher and Ernst Fehr, 513–32. Cambridge: Elsevier, Academic Press.
Spunt, Robert P., and Ralph Adolphs. 2014. "Validating the Why/How Contrast for Functional MRI Studies of Theory of Mind." *NeuroImage* 99 (1): 301–11.
Tajfel, Henri, and John C. Turner. 1986. "The Social Identity Theory of Intergroup Behavior." In *Psychology of Intergroup Relations*, edited by Stephen Worchel and William G. Austin, 7–24. Chicago: Nelson-Hall.
Tansey, Elizabeth M. 1995. "What's in a Name? Henry Dale and Adrenaline, 1906." *Medical History* 39 (4): 459–76.
Vul, Edwad, Christine Harris, Piotr Winkielman, and Harold Pashler. 2009. "Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition." *Perspectives on Psychological Science* 4 (3): 274–90.
Walz, Jennifer M., Robin I. Goldman, Michael Carapezza, Jordan Muraskin, Truman R. Brown, and Paul Sajda. 2013. "Simultaneous EEG–fMRI Reveals Temporal Evolution of Coupling between Supramodal Cortical Attention Networks and the Brainstem." *Journal of Neuroscience* 33 (49): 19212–22.
Wiesel, Torsten N. 1982. "Postnatal Development of the Visual Cortex and the Influence of Environment." *Nature* 299: 583–91.
Wilson, David Sloan, and Elliott Sober. 1994. "Reintroducing Group Selection to the Human Behavioral Sciences." *Behavioral and Brain Sciences* 17 (4): 585–608.
Yeung, Nick, and Alan G. Sanfey. 2004. "Independent Coding of Reward Magnitude and Valence in the Human Brain." *Journal of Neuroscience* 24 (28): 6258–64.