Is neuroeconomics doomed by the reverse inference fallacy?

Sacha Bourgeois-Gironde

Mind Soc
DOI 10.1007/s11299-010-0076-z

Is neuroeconomics doomed by the reverse inference fallacy?

Sacha Bourgeois-Gironde

Received: 22 January 2010 / Accepted: 7 September 2010

Abstract Neuroeconomic studies are liable to fall into the reverse inference fallacy, a form of affirmation of the consequent. More generally neuroeconomics relies on two problematic steps, namely the inference from brain activities to the engagement of cognitive processes in experimental tasks, and the presupposition that such inferred cognitive processes are relevant to economic theorizing. The first step only constitutes the reverse inference fallacy proper and ways to correct it include a better sense of the neural response selectivity of the targeted brain areas and a better definition of relevant cognitive ontologies for neuroeconomics. This second way also allows increased coherence between the cognitive processes actually involved in neuroeconomics experiments and the theoretical constructs of economics. We suggest means of increasing neural response selectivity in neuroeconomic experimental paradigms. We also discuss how the choice of cognitive ontologies can both avoid implicit reductionist strategies (from economic constructs to neural patterns) and irrelevance, as cognitive processes engaged in experimental tasks may lack immediate bearing on the study of economic behavior. With these joint improvements neuroeconomics can be a progressive science.

Keywords Neuroeconomics · Reverese inference fallacy · Cognitive ontologies · Reductionism · Cultural neural recycling

1 Introduction

Neuroeconomics has been defined as the investigation of neural correlates of decision-making, in choice situations which may be of interest to the economist and with the prospect of enriching or revising some theoretical assumptions of the

S. Bourgeois-Gironde (✉)
Institut Jean-Nicod, Pavillon Jardin, 29, rue d'Ulm, Ecole Normale Supérieure, 75005 Paris, France
e-mail: sbgironde@gmail.com

Published online: 19 September 2010

economic science (Sanfey et al. 2006). It is not our purpose here to assess the likeliness that neural data really can change the theoretical foundations of economics. This problem has been addressed elsewhere (Gul and Pesendorfer 2008; Bourgeois-Gironde and Schoonover 2008). In particular it has been doubted that theoretical constructs of neuroscience, economics and psychology are commensurable. Our more specific concern here is with the possibility to infer cognitive processes and mental states—which are assumed themselves to be relevant to an understanding of economic behavior—from observed neural activities. This inference constitutes a crucial conceptual link for neuroeconomics. Without postulating conceptual identity or overlap between constructs of different disciplines, inferential steps between neural observations and mental or behavioral concepts would give sense to frequently stated three-tiered conclusions in neuroeconomic studies which from the observation of given neural patterns conclude to the presence of cognitive processes and/or mental states and, eventually, to the incorporation of a new psychological feature in a utility-maximization functional form.

Our discussion in the following bears essentially on the first inferential link. We will start by stating in what way the problem of reverse inference, which looms over psychological interpretations of brain-imaging data in general, affects some important neuroeconomic studies and then consider under what conditions and constraints this problem can be bypassed. Only at the end will we raise the more general issue of the relevance of drawing and validating such inferences as forming the core research program in neuroeconomics. This issue will connect back to the main epistemological difficulty met by neuroeconomists which is to make sense of integrating neural data in economic models. One way of doing so would be to assert that economic behaviors are ultimately reducible—by means of intermediating valid inferences—to neural data. But most generally neuroeconomists do not care about holding a reductionist view of the mind and are happy to correlate inferred mental states or cognitive processes based on a functional analysis of some observed neural activities with behavioral patterns that make sense with respect to some intended economic model. But those correlations are not innocent. Even though they do not necessarily entail reductionism, they might rely on a naïve view of the reliability of correlations between observed neural activity and cognitive processes or mental states. This may actually amount to a reification problem[1] which appears when neuroeconomists dogmatically assume that their theoretical constructs successfully capture behavioral patterns and cognitive processes and that neural bases observed in correlations with these patterns and processes are neurobiological realizations of those constructs. So even if neuroeconomists do not really philosophically care about the ultimate biological constituents of behavior, they should nevertheless feel concerned about the validity of these basic correlations in order not to fall into unwanted reductions. The reification problem is independent from the reverse inference fallacy but the confidence that researchers may have in having observed the neural bases of their constructs may nevertheless bias them toward its unreflective use.

[1] I thank Nicholas Bardsley for having suggested this label and for many comments and improvements on this article.

Epistemological safety may be restored in two ways: either by complying with the constraints that license functional inferences from observed neural patterns to cognitive processes and states, or by relying on an altogether different way of interpreting neurobiological mechanisms which underlie economic behavior. We discuss the nature of those constraints and how some extent neuroeconomic studies meet with them with variable success. The use of cognitive ontologies is the most straightforward way to address those informational constraints and improve the likeliness of drawn reverse inferences. They consist in databases of extant correlations between neural activities and functional and psychological descriptions. In that way they can also be considered architectures which reflect current assumptions about human cognition. In the final section we give a hint of an alternative approach to apprehend the connection between neural data and economic constructs and of its potential epistemological relevance for neuroeconomics.

In Sect. 2 we recall what the general problem of reverse inference is and how it threatens the neuroeconomic enterprise. In Sect. 3 we address the special point of neural selectivity which, when sufficiently high, allows for a Bayesian inference from a given observed neural pattern to a particular cognitive process. We give examples of salient neuroeconomic studies which, we deem, do not sufficiently meet this criterion of neural selectivity. In Sect. 4 we indicate how neuroeconomics may resort to cognitive ontologies to improve the information about the selectivity of observed neural activities with respect to intended cognitive processes in the run experimental tasks. We also consider whether the way conclusions of neuroeconomic studies are usually stated does not point to implicit and often unwanted reductionist strategies. An alternative strategy can be explicitly stated which makes the choice of cognitive ontologies possibly relevant to neuroeconomics dependent on background evolutionary hypotheses about how some specific brain areas may have evolved in order to fit with regular features of economic environments. Given a three-fold caveat on methodological improvements on control of neural selectivity, choice of cognitive ontologies, and theoretical relevance of the intended cognitive processes and their presumably associated neural ones, the position advocated here is that neuroeconomics can escape the reverse inference fallacy.

2 The problem of reverse inference in neuroeconomic studies

There is much current interest in using brain-imaging techniques in order to obtain a better understanding of the nature of cognition. Several techniques can be used. Some of them, like MEG and EEG, capture the information that is carried on by the brain's electric activities; others such as fMRI and NIRS rely on hemodynamic changes in the cortex.[2] Given the accumulation of brain imaging data during the past two decades we presumably have acquired a stable functional cartography of

[2] Magneto-encephalography (MEG) and electro-encephalograhy (EEG) are techniques for mapping brain activity by recording respectively magnetic or electric fields occurring naturally in the brain. Functional magnetic resonance imaging (fMRI) measures the change in blood flows in the brain due to neural activity. Near-infrared spectroscopy is an optical method using the correlation between infrared light-waves transmission and metabolic activity in the brain.

the brain, namely the possibility to localize functions in areas with a high level of accuracy and precision. However, the incremental process through which such a mapping can grow up in the course of a systematic functional exploration of the brain is far from inferentially sound. Namely, the transition from one set of functional observations to another may be more dependent on analogical reasoning than on deductive validity. A slippery logical practice has indeed pervaded the field which consists in inferring to the engagement of a particular cognitive process from the activation of a particular brain region on the sheer basis of similar past inferences. This logical move is called a reverse inference. It is not deductively valid even though, as we shall see, it can provide some information under some precisely defined conditions.

Many neuroeconomic studies, in different ways and to variable extents, seem to us to follow this pattern of reasoning. We provide a set of 5 examples from the recent neuroeconomic literature, which we have schematized in order to make salient this underlying logical structure. We start with the most widespread and obvious reverse inference fallacy, which in study 1 below may be rather innocuous to the extent that the intended correlation between a given neural pattern and a cognitive process is explicitly based on previous similar studies. In the following studies 2–5 the reverse inference fallacy is less benign to the extent that authors seem to directly look for the neural realization of some economic theoretical constructs. In studies 2 and 3 authors seem to infer from neural activity to their intended theoretical constructs on the doubly frail basis of too weak support by extant correlations between the same neural activities and similar or sufficiently close cognitive processes, and of available possible alternative interpretations of the observed brain activities. In example 4 the problem is with the gross-grainedness and, therefore, lack of selectivity of the neural systems said to correlate with an economic modeling of behavior. In 5 most of those defects seem to be under control except for the fact that we can point to a local strategy of drawing inferences on the basis of previous reverse inferences made in analogous neuroeconomic investigations.

1. Padoa-Schioppia and Assad (2006) have identified a population of neurons in the orbitofrontal cortex that, they claim, assigns values to economic goods and, they argue, represents their subjective utility independently of the action needed for their acquisition. Value processing means that single goods are attributed anticipated values that will be realized at the time of their consumption. The suggestion is that the orbitofrontal cortex might contain a map for expected subjective cardinal utility. These findings seem to bear directly on the conceptual foundations of economics by presenting a potentially promising three-tiered inference from the observation of neural patterns, the involvement of a cognitive process (value processing) and its understanding as the psycho-biological realization of a fundamental theoretical construct of economics, namely utility. The reverse inference fallacy consists, in this example, to infer from the activity of some neurons in the orbitofrontal cortex to the cognitive process of valuation. This reverse inference simply inverts, as we can see, the hypothesis that performing the task at hand would activate neurons in the

orbitofrontal cortex which are associated, by previous studies, to the cognitive process of valuation. This inversion, however, may not be fatal to the asserted conclusion depending on the confirmation by previous studies of the correlation between the targeted population of neurons and the intended cognitive process (value processing). An extra step, beyond the reverse inference fallacy proper, is to conclude to the observation of the neural basis of the concept of utility. This study may present the weakest and most widespread form of the reverse inference fallacy.

2. Ellsberg (1961) evidenced, through a famous paradox of decision-theory, the different attitudes that are elicited by risky vs. ambiguous situations. Huettel et al. (2006) detected individual differences in brain activations depending on a subject's preferences for risky or ambiguous decision contexts. People who prefer ambiguity demonstrate increased activity in the prefrontal cortex while those who prefer risk have increased activity in the parietal cortex. Here we have a clear instance of neuroeconomics defined as the investigation of the neural correlates of decision-making. Huettel et al.'s conclusion is underdetermined in the sense that the differentiated brain activities correlated with the proposed decision contexts may be attributed to the processing of distinct levels of information rather than of probability contexts proper. The inference from neural activity to relevant theoretical constructs for economics is more direct in this case than in the former example. The cognitive process which is inferred to be correlated with brain activity (processing distinct uncertainty context) and the economic concepts (risk and ambiguity) at stake are the same. In Padoa-Schioppia and Assad's study, "value processing" had a biological sense of its own which takes a further inference or reduction to make tantamount to the notion of utility.
3. Camille et al. (2004) and Coricelli et al. (2005) have tested, respectively with brain-lesioned patients and in an fMRI setting, the reaction to two types of feedback to participants' choices over pairs of fortune wheels. In one case they had complete feedback for the wheel they had chosen and the one they had not chosen. In another case they got only the partial feedback corresponding to the wheel they had chosen. This difference is supposed to capture the conceptual distinction between, respectively, regret and disappointment. Notably, regret and disappointment have been theorized and modeled in decision-theory (Loomes and Sugden 1982) and considered to be possible explanations of fundamental decision-theoretic behavioral anomalies such as the Allais paradox. The brain activities documented by Coricelli et al. (2005) lead them to conclude that the orbitofrontal cortex has a fundamental role in experiencing regret.

   Here again the inference connects three items, as they report correlations between neural activities in the orbitofrontal cortex and an experimental construct they call regret and intend, on that basis, to bridge the gap between those neural activities and the decision-theoretical notions of regret and disappointment. But there is a specificity in this three-step construction compared to the clear three-step inference in Assad and Padoa-Schioppia's study (neural paterns—psychobiological concept—economic theoretical

construct) and the merging of the two last steps in Huettel's study (neural patterns—and then an equivalence between the inferred cognitive process and the theoretical construct, namely ambiguity processing). In Coricelli's study there may be a further doubt that the targeted cognitive process labeled in terms of regret is really involved in performing the task. Indeterminacy looms in general when one infers from observed neural activities to a specific cognitive process. However, in the present context, there are obvious other ways to interpret the alleged psychological processes involved in performing the task (comparing wheels of fortune) than in the phenomenologically rich term of an emotion of regret. This would not be a problem if it were a sheer terminological issue. What Coricelli et al. calls regret is, in a simplified version, what Sugden and Loomes themselves call regret (and if this was the end of the story we would have here a case of conceptual collapse similar to the one in Huettel's study). The problem is that we can alternatively label the psychological process that takes place when performing the task in terms of information processing not associated with any particular emotion. Rationalizing this process in terms of regret may be unwarranted, even though it fits with the theoretical construct intended in decision-theory.

4. In a widely discussed article, McClure et al. (2004) examined the brain activity of participants while they were making a series of intertemporal choices between small proximal rewards ($\$ R$ available at delay d) and larger delayed rewards ($\$ R'$ available at delay d'), where $\$ R < \$ R'$ and $d < d'$. Rewards ranged from $ 5 to $ 40 Amazon.com gift certificates, and the delay ranged from the day of the experiment to 6 weeks later. They found that time discounting is associated with the engagement of two neural systems. Limbic and paralimbic cortical structures are preferentially recruited for choices involving immediately available rewards; and fronto-parietal regions, which support higher cognitive functions, are recruited for all other intertemporal choices. Moreover, the authors find that when choices involved an opportunity for a sooner (not strictly immediate) and a later reward, both neural systems are engaged but an activity in fronto-parietal regions greater than in limbic regions is associated with choosing larger delayed rewards. This dual, and potentially conflicting, system in the brain, given its relative activation according to kinds of delays displayed to the participants, is interpreted by the authors to support a quasi-hyperbolic functional account of utility-discounting over time. The quasi-hyperbolic discount function presents a curve with a steep hyperbolic declivity in the present and very short term and a more constant and less acute exponential slope from a point in the near future to the far future. McClure et al.'s dual neural system apparently supports this dual functional form. One noticeable feature of this original study, though, is its coarse-grainedness in terms of the neural activities, divided between two large systems of the brain, that are supposed to account for intertemporal choice behavior (Laibson 1997). This feature has been targeted by rival accounts of the neural bases of utility discounting (see Kable and Glimcher 2007) which point to more scattered and complex neural-subsystems underpinning intertemporal choice behavior. By

opting for large neural structures, McClure and his colleagues decrease the level of specificity with which they are actually correlated with intertemporal decision-making rather than other psychological activities. In that field of study a safer strategy would be to decompose intertemporal choice behavior into its various cognitive components (valuation, prospective thought, etc.) and study their corresponding neural subsystems.

5. Fundamental issues in economics such as equity/efficiency trade-offs have been directly tackled by means of brain-imaging techniques. Hsu et al. (2008) combined choices over distributive justice situations, implementing trade-offs between equitable and efficient donations to actual Ugandese orphans, with fMRI recordings of overall brain activity during those trade-offs. They were able to observe salient correlations between neural activities and experimental situations in which dilemmas between equity and efficiency were implemented. They report that differentials of equity were processed by the insula, differentials of efficiency by the putamen, and differentials of utility (defined as a combination of equity and efficiency) by the caudate nucleus. They manage to make sense of those correlations by inferring cognitive processes from prior functional mappings of those brain areas, namely inequity aversion may be associated with insular activity, calculation with the putamen and reward processing with caudate nucleus activation. The theoretical constructs (equity and efficiency) are much plausible labels given the trade-offs involved in the tasks. There is also no doubt that observed neural activities are correlated with the distinct typical responses to those tasks. The only concern is with the specificity of the observed neural activities with respect to the cognitive processes involved. Inferences from these activities to further characterizations of the engaged cognitive process are too loose. For instance, the fact that from what we previously know about which affective or cognitive states are associated with insular activity we are tempted to infer that dealing with inequity may tap into deep affective and proprioceptive states such as disgust essentially relies on Sanfey et al. (2003). There may be a sort of mutual reinforcement between the reverse inferences drawn in these thematically related studies but this is still insufficient to remove all qualms about the level of specificity of the observed neural activities with respect to the intended cognitive processes.

The slippery inferential move in which the reverse inference fallacy consists and which was found to a variable extent in the above mentioned studies can be plainly described as (1)–(2)–(3) below. It is a reverse inference by contrast with a properly ordered one which simply states that if a cognitive process or a mental state X is engaged, then brain area Z is activated. But what we rather regularly find in reports of fMRI studies is the alternative sequence:

(1) In the current study at hand, we observe that when task A is performed brain area Z is activated.
(2) In former studies, when cognitive process X is supposed to be involved, brain area Z is activated.
(3) Therefore, brain area Z activities in the current study at hand demonstrate the involvement of cognitive process X by the performance of task A.

Compared with the regular inference from cognitive processes to brain activities which is strategically used in (2), (1)–(3) follows a reverse inference pattern as it aims at establishing the involvement of a cognitive process on the observational basis of given brain activities. The inference is invalid as it simply affirms the consequent. The inference would be valid if (2) was rather asserting a strict equivalence such as: brain area Z is activated if and only if cognitive process X is involved in a task. Poldrack (2006) makes a very cogent point about the use of this fallacious reasoning in a vast array of fMRI studies. It would be tempting to interpret this regular move in conclusions of fMRI studies as a token of abductive reasoning which naturally accompanies scientific progress. Abduction can be described as the attempt to conclude that A given that we observe B and that we know that A entails B. In quest for an explanation A of an observed phenomenon B, it is clear that the simple presence of B in nothing more than a probable indication that A is also currently the case. This reasoning pattern is tantamount to the fallacy of affirmation of the consequent in logic. As Peirce has originally conceived of it abduction was supposed to be the inferential process that leads to the formulation of a hypothesis, not to a logically certain conclusion (Peirce 1878). Such quasi-logical procedures would then be expected reasoning patterns in an emerging scientific field such as neuroeconomics and what we often find conclusively stated in that field would rather be taken, according to Peirce's recommendations of scientific methodology, as the hypotheses to test. Some constraints may transform these logical fallacies into informational processes, given prior background knowledge.

In absence of a lucid assessment of these informational constraints neuroeconomic investigations may rely on what we label a "same-areas strategy": the *same areas* that are usually dedicated to the processing of X are also observed to be activated in the processing of Y; X being a biological or cognitive function previously mapped onto some brain structures, and Y the currently examined economic construct. The same are strategy consists then in the mere conjunction of (1) and (2) above without regard to neural selectivity matters. However, X already possesses some domain specificity, whereas an important question, which is far from being solved, is whether Ys—i.e. conceptual constructs or cognitive processes theoretically associated with the description of economic behavior—can have any. We analyze in the following section one particular example of a neuroeconomic study in the light of the informational constraints that bear on the validity of its conclusions that could help it escape from a seemingly "same-areas strategy".

3 Bayesian constraints on the informativeness of reverse inferences

De Quervain's article on the neural basis of altruistic punishment (De Quervain et al. 2004) is a sufficiently elaborate example in order to unravel the sort of internal inferential links neuroeconomic conclusions rely on. Altruistic punishment is a concept of evolution theory (see Fehr and Gächter 2002) and one important way of envisioning neuroeconomics as a promising and growing scientific field is certainly, as we will show in the final section, to consider that it helps addressing questions relating to the evolution of social and economic behaviors. Altruistic punishment is

the fact, reported in laboratory experiments and confirmed by field observations, that people tend to punish free riders and non cooperators, when they have a chance to do so in experimental or life settings even if they incur a personal cost of doing so and even in one-shot situations in which there cannot be a corrective effect of the administered punition on subsequent social interactions. Investigating the neurobiological basis of altruistic punishment is important in order to shed light on the evolutionary mechanisms that selected such an apparently individually counterproductive behavior.

Moves (1), (2) and (3) of the above inferential sequence occur in De Quervain's presentation of his study. Move (1) is the far most complex one as it has to contain the description of a task and a report of correlated neural activities. It proceeds by stating the rule and conditions of an experimental game between two players. Players are anonymously matched and play a repeated trust game. Punishment is introduced in the ultimate phase of each period of the repeated game. Namely, each time the first player is returned a share of a multiplied amount of the money he has himself previously entrusted to the second player, he can express his potential discontent over the return by punishing that second player. By spending one monetary unit on punishment, the first player may deprive the second player of two monetary units. The first player can spend up to 20 monetary units on punishment. In De Quervain's study, the brains of first players are scanned while their trust is abused by second players and they ponder whether they will inflict punishment to second players, under various experimental conditions that correspond to the way punishment can be inflicted. The treatments were Costly Punishment (CP) just as we described it, Free Punishment (FP) in the sense that the penalties inflicted to the second players cost nothing to first players, and "Symbolic" Punishment (SP), which in De Quervain's particular terminology only meant that players may have wished to punish free riders but were not actually given the possibility to do so.

Having defined their experimental conditions, the authors state their hypotheses, which bear on the resulting brain activities when the usual subtractive fMRI methodology is applied. This consists in making apparent by subtraction (or by superposition of the average maps of activities in the brain obtained through the repetition of the tasks) the contrasts between brain activities associated with these experimental conditions which are designed to be as close as possible except for the manipulated variable of interest. The resulting contrasts are supposed to be specific to the first terms of the successive subtractions below:

FP-SP is supposed to activate reward related brain regions.
CP-SP is also hypothesized to activate reward related brain regions.

Now what is observed is that FP-SP and CP-SP activate the caudate nucleus, which is a reward related brain region. Punishment, at a cost or at no cost, activates this region of the brain.

Move (2) is simply the reminder that there is well documented evidence of reward related areas in the brain (nucleus accumbens, nucleus caudate, etc.) showing that they are activated when subjects get reward in the form of e.g. money, beautiful faces, or psychoactive drugs.

Move (3) allegedly concludes that people derive utility from altruistic punishment. Altruistic punishment is rewarding for subjects. When they have the opportunity to punish, they experience reward, "as shown" by related brain activities.

This reverse inferential step is embedded within a broader argument about the investigation of the proximate neural phenomena that may account for the evolutionary selection of supposedly widespread altruistic punishment behavior. The problem is that, as it stands, this inferential step is fallacious and threatens a broader valuable endeavor.

The initial issue was to assess whether reward related areas in the brain were activated when subjects have the opportunity to punish. The result is that they are. But why was this issue intended in the first place? The stated conclusion is that free and costly punishments are rewarding given the activity of reward related brain areas when punishment is performed. Our point is not to target a phenomenological slip, from brain-activities correlated with some experimental constructs to a feeling of reward, which is actually not very present in that study. The far more problematic point is that it is fallacious to conclude from the correlation of performances that have been described in terms of kinds of punishment with reward related brain activities, to the idea that punishment itself is a reward related behavior. As we said that logical slip is embedded in an attempt at providing a broader account of the neurobiological mechanisms that have fostered altruistic punishment and eventually human cooperation (my emphasis): "Our study is part of recent attempts in "neuroeconomics" and the "cognitive neuroscience of social behavior" to understand the social brain and the associated moral emotions. However, this study sought to identify the neural basis of the altruistic punishment of defectors. The ability to develop social norms that apply to large groups of genetically unrelated individuals and to enforce these norms through altruistic sanctions is one of the distinguishing characteristics of the human species. Altruistic punishment is probably a key element in explaining the unprecedented level of cooperation in human societies. *We hypothesize that altruistic punishment provides relief or satisfaction to the punisher and activates, therefore, reward-related brain regions*. Our design generates five contrasts in which this hypothesis can be tested, *and the anterior dorsal striatum is activated in all five contrasts, which suggests that the caudate plays a decisive role in altruistic punishment*."

It is clear that the crucial move (3) of the reverse inference fallacy is bluntly assumed in this introductory assertion. The conclusion is preempted by the hypothesis that if punishment is rewarding then it activates typical brain regions that have been previously associated with reward processing. And if a significant activation of these regions is observed in connection with tasks that implement altruistic punishment in a unambiguous way, then the assertion that punishment provides relief or satisfaction seems within reach. But saying that the caudate plays a role in altruistic punishment does not yet mean that that apparent correlation between caudate activities and the performance of altruistic punishment anchors the latter in reward-related activities. It, on the one hand, depends on the level of functional specificity of that area and, on the other hand, of a better characterization of the specific sense in which punishment may involve reward rather than some

other more general theoretical psychological construct that would be common to reward proper and punishment.

Another example would be Sanfey and his colleagues' article on the neural basis of economic decision-making in the ultimatum games which paved the way for neuroeconomic studies of human cooperation (Sanfey et al. 2003). The argumentative structure of this article is as follows. Involvements of the anterior insula and the dorsolateral prefrontal cortex are associated to the performance of the ultimatum game. This is not problematic as this move (corresponding to move (1) above) consists in a plain assertion of neural activities). It is more problematic of course when these activities are subsequently assumed to represent the twin demands of the Ultimatum Game task, namely the emotional goal of resisting unfairness and the cognitive goal of accumulating money. But in the light of past functional investigations of those same brain areas (move (2)), the performance of these tasks is re-described in terms of some cognitive processes and mental states that have been associated with those areas. Activities in the insula, in particular, were associated with autonomic arousal, negative emotional state, anger, physical disgust (taste and odor) and emotional disgust. Hence a swift inference (move (3)) from observed activities in the insular cortex to experienced feelings of anger and disgust when the subject is facing low offers in the game. When such low offers happen to be accepted, higher activities are observed in the dorsolateral prefrontal cortex, which has been associated *inter alia* with the engagement of cognitive control and is interpreted as indicating that an automatic emotional response has been inhibited.

In spite of this inherent logical failure, neuroeconomic studies that succumb to it are not necessarily wrongheaded in proceeding in that way. It might be the case that the intended conclusion is justifiably asserted. But it sometimes seems to be out of epistemic luck rather than the outcome of a controlled methodology and consideration of the conditions under which one can confidently infer the engagement of cognitive states on the basis of an observed correlation between some neural patterns and the performance of given economic decision-making tasks. Reverse inferences should be considered as spreading over a continuous confidence scale between sheer hypothetical statements and definite conclusions, depending on some background conditions. More exactly, reverse inferences vary over the dual dimension of certainty and informativeness. Given a cognitive process X, a neuronal activity in a targeted area A, and a performed task T, the probability that X is engaged while A is observed in T can be classically formalized using Bayes' theorem:

$$P(X|A) = \frac{P(A|X)P(X)}{P(A|X)P(X) + P(A|\sim X)P(\sim X)}$$

P(A|X) and P(A|$\sim$X) constitute the prior information from the existing evidence base, which inform about selectivity. We focus here on the conditionalization of the engagement of a cognitive process X (or a mental state, cognitive or affective) on the observation of some neural patterns, because the inference from the latter to the former is the main and most common source of incautious reasoning in neuroeconomics and elsewhere in neuro-studies. One should note that the presence of X is itself conditioned on the hypothesis that the performance of task T in the

experiment under scrutiny engages X. Neuroeconomic researchers expect that the task T they experimentally run will involve a certain cognitive process X. On the basis of past studies they know that when X was engaged by a task, some brain activities A were observed. The reverse inference fallacy consists exclusively in this further step from the observation of brain activities A to the conclusion that X is engaged. The background expectation that T engages X is part of the overall research strategy and seems to be retrospectively confirmed by the observation of brain activities A.

More subtly, it may also be the case that observed brain activities actually "reveal" the presence of cognitive processes alternative to the ones that are expected to be involved given the task at hand. The point of neuroeconomics is sometimes to deliver this sort of "scientific surprise". But this clearly relies on the prior belief that a certain cognitive process is engaged given a certain task. Let's note that if the expected cognitive process X given task T is the same cognitive process X which is inferred from brain activities A, we have what we define as a set of coherent cognitive expectations. More precisely, a tacit expectation (the correlation between cognitive process X and task T) is confirmed by the observation of brain activities A, given that such activities have been previously correlated with the similar hypothetical presence of cognitive process X. However, as we just said, investigative strategies in neuroeconomics may rely on actually discrepant cognitive expectations when task-performance and brain activities are alternatively considered. The usual strategy is to tacitly or explicitly assume that cognitive process X is associated with task T at hand, and, then, to put retrospectively this assumption in doubt when it is observed that actual brain activities A rather indicate (on the basis of a reverse inference) that cognition process Y or complex cognitive process X + Z seem to be involved. Now given that both types of correlation between T and X, or A and Y, or A and X + Z are based on uncertain conditionalization procedures it makes the intended surprise more or less genuine.

This overall strategy and the conditional series it contains (X given T, as a background hypothesis, followed by a confirming or disconfirming X given A, on the basis of a reverse inference) can be made sounder by unraveling the informational priors contained in this reasoning pattern. The main problem, as we can see, is the conditionalization of an expected X (or of any other less expected cognitive process) over a pattern of observed neural activities A. The degree of belief that we can soundly attach to the reverse inference from A to X depends on what is called the selectivity of the neural response (see Poldrack 2006). The selectivity of brain activation is inversely correlated with its involvement across all possible experimental tasks, and therefore possible cognitive processes. The more a neural response is involved in a vast array of tasks and processes, the less the inference from that neural response to the engagement of a specific cognitive process or mental state is plausible.

The amount of selectivity of the neural response is then crucial in order to propagate certainty between the two levels of inference that we have outlined. Low selectivity of A will retrospectively weaken the likeliness of X given T. This will of course be a problem for the neuroeconomist whose strategy is based upon what we have called a coherent set of cognitive expectations, hypothesizing that X/T and

then citing evidence that X/A. But for the one whose tacit or explicit agenda is the disconfirmation of the generally admitted X/T by showing that in fact T $\Longrightarrow$ A and A $\Longrightarrow$ Y, a low selectivity of A will also ruin his strategy as the inference of Y from A will lack significance and will be orthogonal to the presence of X. Basically, according to the degree to which a region of interest in the brain is selectively activated when the engagement of cognitive process X is hypothesized, the more confident one can be in the reverse inference that cognitive process X is engaged when that region of the brain is activated. The problem is that the measure of neural response selectivity is itself far from simple, stable and reliable.

One would rather have already well-considered notions of where to look in the brain in order to use the criterion of neural response selectivity as an enhancing factor of one's intended reverse inferences. In social neuroscience and, more especially, in neuroeconomics, the point is seldom raised and the level of "brain region" and its functional associations is generally adopted without any further discriminative ado. However, in cases in which the question at hand has been structured over the activation of a specific set of neurons the issue of the neural response selectivity more naturally arises. Let's take an example of a critical assessment of the neural response selectivity factor in a context where the authors seek to establish new functional correlations between brain areas and some human cognitive capacities. Bastiaansen et al. (2009) have wondered whether the mirror neurons system which allows us to anticipate others' motor behavior is also involved in the simulation of their emotions. This supposes that the selectivity of an identified set of neurons to motor tasks is strong enough and sufficiently stable across several similar tasks. But it also implies that this same set of neurons will equally be activated over tasks of a very different nature, related now to the decoding of others' emotions. A seeming paradox immediately arises: if this same set of neurons is both involved in the mirroring of motor and emotional behaviors, its selectivity will logically decrease. Bypassing such difficulties is crucial to the pursuit of one of the currently active programs in neuroeconomics which aims to anchor our understanding of cognitive processes involved in coordination games in precise mirroring and mentalizing neuronal systems in the brain (Coricelli and Nagel 2009).

Recent neuroeconomic studies of coordination games have thus taken advantage of the precisely established correlations between populations of neurons in several brain areas and mental processes in an attempt to validate theoretical models of how people solve coordination problems. But the gap between game-theoretical models and specific populations of neurons remains unbridgeable if one cannot validate all the methodological intermediate steps from the latter to the former. For example, Coricelli and Nagel (2009) report having identified the neural substrates of strategizing in a "beauty contest" game. They present their data as showing that successful strategic reasoning in that game correlates with neural activity in the medial prefrontal cortex. Adequate reasoning is inferred when behavioral answers happen to conform to a theoretical model that accounts for rational solutions in "beauty contest" games. As they actually observe increased activities in the prefrontal cortex when people exhibit that type of behavior, authors interpret their data as supporting that theoretical model. These steps may look bold by comparison

with the methodological meticulousness that has been displayed in neural studies of mentalization and social cognition in general. But it does not mean that those results and conclusions are to be a priori discarded upon the motive of their partial reliance on these anterior studies, as it may precisely be the case that neuronal activities in the medial prefrontal cortex enjoy a high level of selectivity under the performance of such coordination games. The problem is that the question is not explicitly addressed in the discussion of the results and we generally remain uncertain as to the selectivity of an observed neural response with respect to a given task, weakening the inference that can subsequently be drawn to the fact that we have actually observed the neural substrates of an intended cognitive process.

One way for neuroeconomic studies to avoid external reliance on previously acquired data with respect to the correlation between pinpointed neural activities and cognitive and affective processes that are important for neuroeconomics would be that experimental paradigms in that discipline generate their own measure of neural selectivity. So-called repetition suppression paradigms have been used, in the context of the investigation of low level brain processes such as perception, in order to determine the selectivity of a region. Repetition suppression is a reduction of neural response that can be observed when stimuli are presented several times. Many functional investigations of brain areas have taken advantage of the phenomenon of repetition suppression to probe the sensitivity of those areas to variable stimuli.

This methodology is not confined to the investigation of most basic brain functions and can be applied to assess the involvement of affective and cognitive processes that have relevance in neuroeconomics. Jeankins et al. (2008) examine repetition suppression of a brain region, the ventromedial prefrontal cortex (vMPC), in elaborate cognitive tasks that consist in the introspection of one's own mental states vs. inferences about whether other persons are having similar mental states. The neural bases of the human mentalizing ability have immediate neuroeconomic importance in order to clarify the cognitive basis of strategizing as it is captured by game-theory (Singer and Fehr 2005). In particular it is interesting to know whether the understanding of others' thoughts and intentions crucially depends on a projection onto others of our own thoughts and intentions. Jenkins and her colleagues use alternatively self vs. other-directed judgment tasks systematically followed by an other-directed judgment task. They note that a repetition-related suppression of neural activity in the vMPC occurs when the second other-directed judgment is elicited, whether it has been preceded by a self- or by an other-directed judgment. This result shows that thinking about others depends on introspective capacities. An assessment of the selectivity of the functional vMPC response vis-à-vis the targeted dual cognitive process on the basis of previous extant studies is here optional as the suppression phenomenon provides an internal criterion of that correlation.

4 Ontological problems for neuroeconomists

The assessment of the selectivity of neural responses depends among other things on the access to databases and meta-analyses. One has only partial access to such

databases and some cognitive and mental processes may be over-represented while others are only scarcely studied. Those databases have been labeled "cognitive ontologies" and can be simply defined as functional mappings between cognitive processes and functions, on the one side, and anatomical structures of the brain, on the other side Price and Friston 2005; Poldrack 2006; Christoff and Owen 2006). Extant 'cognitive ontologies'[3] do not contain entries for economic theoretical constructs. Their elaboration was guided by the unfolding of a functional architecture of the brain, reflecting current working assumptions about human cognition. Cognitive ontologies connect an anatomical level with a cognitive functional level: a problem is then is the granularity level at which a neural response is envisioned. What is the required scale? Is it the neuron or the functionally defined area? The more physiologically fine-grained the nature of the response, one could presume, the more reliable the prediction of a given mental process, given that selectivity, intuitively, co-varies with the size of the neural sample. However, this may amount to a naïve view of functional organization in the brain as a high level of response variability at the neuron scale may not thwart a high level of functional selectivity at a more regional scale. This is at least a question that is meticulously raised by leading researchers of the issue of neural response selectivity for the type of mental processes about which a large set of data has been gathered. Logothetis and his team, in particular, have even investigated whether neurons in primate inferotemporal cortex respond selectively to complex, often meaningful, stimuli such as faces and objects and are stable from 1 day to the other (Bondar et al. 2009). They found that those neurons maintained their selectivity in both response magnitude and patterns across a large array of visual images throughout periods that sometimes exceeded 2 weeks.

One complication in view of the functional cartography of the brain is that there is no one–one correspondence between brain activities—at whatever scale they are envisioned—and cognitive processes. Functional brain-imaging has fostered a view of the brain which is quite in line with the multi-realizability of mental states thesis that had been advanced in philosophy of mind in the 1970s (e.g. Fodor 1974). There is a wide degree of overlap among the different neural systems that are activated by tasks that have no apparent cognitive components in common, which suggests that a given neural system can "realize" several functions. This could be considered an extra source of indeterminacy when a specific area is known to underpin several functions. In that case, even if the selectivity of that neural response is granted for those functions, the particular association of that response with a cognitive process will remain uncertain. To solve this problem some neuroscientists use a further criterion of connectivity (Price and Friston 2005). Neural connectivity refers to a pattern of anatomical links ("anatomical connectivity"), of statistical dependencies ("functional connectivity") or of causal interactions ("effective connectivity") between distinct units within a nervous system. The units may vary in scale, from individual neurons, to neuronal populations, or anatomically segregated brain regions. Neural activity, and by extension brain functions, are constrained by

[3] As a relevant example one can refer to the ongoing database project brainmap.org, which can compute activation likelihood estimations.

connectivity.There will indeed be a limited range of functions that an area can perform if we fix its internal and external connectivity. In other terms, the set of stable coincident activations within and without a targeted area across a specific task may provide a wider observational basis to map more finely cognitive processes and brain functioning. Fixed states of neural connectivity in the brain help to increase the accuracy of cognitive ontologies. The point, then, for a neuroeconomist to avoid the reverse inference fallacy, should be to check that his conclusion from a brain activity to a cognitive process falls within the scope of a local ontology. One requisite for this approach is then that neuroeconomics have as clear a preliminary idea as possible of what cognitive processes are worth studying in their perspective. Neuroeconomics could be defined as the investigation of the neural bases of economic decision-making. But the use of cognitive ontologies may refine this plain definition and incite neuroeconomists to reach an understanding of how some brain structures may have become specifically involved in aspects of economically relevant behavior. How the brain adapted to economic environments and how, if ever, it became partly functionally specialized to deal with those features of environments and related behaviors that economic science models should be, in my view, the main concern of neuroeconomics, But the pursuit of that goal depends on tight intermediary correlations between precise cognitive ontologies and a high selectivity of associated neural responses. The usual investigative strategy in neuroeconomics which consists in observing presumable neural correlates of some hypothesized cognitive process or, even more directly, to some theoretical construct, remains defensible under the dual condition that the neural response selectivity and a cognitive ontology are granted.

So when a prediction is formed about which area of the brain will be activated under the performance of task T, or which cognitive process is involved given brain activity A, the functional associations that have been previously set with respect to that brain area or those brain activities are determinant. They constitute a background knowledge, which is sometimes cursorily referred to but still guides the neuroeconomist's explicit investigative strategy and give potential weight to the more implicit conditional inferences on which this strategy relies. So the characterization of task T (which among the three place-holders A, X, T of the inference is the one we have least discussed so far) according to its alleged cognitive domain increases the viability of an intended reverse inference. However, does it make sense to associate economically relevant tasks with specific cognitive domains, and is it methodologically and strategically sound to do so if the purpose of economics is precisely to investigate the neural bases of *economic* decision-making and hope to reach a high level of specificity?

Tasks are swiftly assigned cognitive domains: memory tasks, linguistic tasks, attention tasks, and so on. This initial assignment helps form expectations about which area of the brain will be activated by their performance given that it is established that some regions show selective neural responses with respect to such cognitive domains. Broca's area is for example well known to be selectively activated in language tasks by comparison with tasks which do not involve linguistic performance. By contrast the insular cortex is known to be involved in negative emotional and autonomic arousal contexts but it probably crosses over emotional

domains whose boundaries are hard to define. Even more clearly, some regions of the brain, like the rostrolateral prefrontal cortex, have little domain-specificity. So tasks that would be defined only by their presumed cognitive domain and that would activate such low selective regions would constitute a poor starting point in view of a credible reverse inference. The case is even worse when no specific cognitive domain can be assigned to the task and the strategy is purely explorative in terms of correlation of type of performance and brain areas which have a variable degree of selectivity.

Examples of such loose associations abound in neuroeconomics and have given rise to legitimate attacks by authors who could easily pinpoint abuses of conceptual overlaps or reductionist attempts between theoretically separate fields (Gul and Pesendorfer 2008). Some of the neuroeconomic studies we have mentioned in Sect. 2 may happen to be sufficiently grounded if the neural selectivity and a fine connective functional mapping of the targeted area grant a reverse inference from those activities to the cognitive processes. But besides granting the likelihood of reverse inferences, cognitive ontologies fulfill two further roles in the typical neuroeconomic strategy: (1) selecting a cognitive process for which it will make sense to presume its engagement in the task and its correlation with the observed neural response, and (2) ensuring that this intended cognitive process has relevance in neuroeconomic and, plainly, economic theorizing. The difficulty that may arise from these demands is that their joint fulfillment may yield the impression, sometimes not intended, of an attempt at metaphysically reducing theoretical economic constructs to neural activities.

Not only a reminder of methodological safety, but one of metaphysical clarity, is then in order. Reductionism claims not only that all natural kinds are co-extensive with physical natural kinds, but that those co-extensions are nomologically necessary. In Fodor's terms and example: bridge laws between separate scientific descriptions of reality are actual laws; so, if Gresham's law is true, it follows that there is a (bridge) law of nature such that 'x is a monetary exchange $\Leftrightarrow$ x is P', where P is a term for a physical natural kind" (Fodor 1974). An instance of Gresham's law is obviously a worldly fact which is an aggregate of physical phenomena. However, an economic law is not reducible to physics in the proprietary sense of reduction involved in claims for the unity of science for the same reasons that psychology can not be reduced to neurology. Fodor argues that psychological laws, first, have exceptions and, second, are multi-realizable, resulting in uninformative potentially infinitely disjunctive bridge laws of reduction of psychological types to neural types. Now if one takes behavioral economics as a mix of economics and psychology the argument, a fortiori, applies.

The choice of a cognitive ontology in neuroeconomics is tricky, first, because the intended cognitive processes do not a priori belong to one single specific cognitive domain, and, second, because they are high-level processes whose neural underpinnings may be more difficult to selectively assign compared with those of lower mental functions. Another pressing issue is that those cognitive processes, whatever theoretical domain they fall into, have a high level of complexity which may complicate the assessment of their neural response selectivity. But it has also been suggested that an estimate of cognitive complexity for a given task may solve the

reverse inference problem when this task triggers a lowly selective neural response (Christoff and Owen 2006). Cognitive complexity has been defined in several ways: parallel processing of sub-tasks, load of working memory, temporal and hierarchical unfolding of a plan, and so on. The analysis of the rostrolateral prefrontal cortex activations, for instance, which is an area of apparent low selectivity, show that those activations are positively correlated with task-complexity. Brain regions can be differentiated by their selectivity not only to types of tasks, when the cognitive characteristics of these tasks are clearly identified, but also to particular aspects or properties of tasks such as their complexity, especially for those which do not a priori fall within a well determined cognitive domain.

Finally it could be stated that solving the reverse inference issue consists in developing a clear preview of the structural organization of the brain in response to tasks that are of interest for the behavioral economist. Only provided these preliminary functional hypotheses, can drawing inferences from brain activities to the engagement of specific cognitive processes in the performance of economic tasks make sense. But it should also be the aim of brain-imaging studies in economics to contribute an answer to the more fundamental question of whether our brain has developed responses that became specific to the economic domain. De Quervain et al. (2004) article on the neural basis of altruistic punishment which we discussed above actually goes into that direction, its main lesson being that to deal with unfair economic exchanges and reinstate cooperative norms an efficient strategy is to punish free riders even at a personal cost. This is one behavioral device that helped human survival because it was crucial to install norms of cooperation in environments too complex to be individually coped with. The fact that the administration of such altruistic punishment correlates with neural activities in some structures of the reward system is seen as optimal from an evolutionary point of view, in the sense that something painful—incurring a financial cost—would be processed as something pleasurable.

This relates to the second aspect of why cognitive ontologies are especially problematic in neuroeconomics. The cognitive processes that are involved by tasks that make up neuroeconomic experimental paradigms not only should have enough neural selectivity, but also some relevance in view of economic theorizing. In the vein of De Quervain's or other studies, we want to emphasize a possible strategy in neuroeconomics that would be driven by a self-conscious concern on how neural activities became specialized to deal with behaviors and cognitive processes that would clearly mark the cognitive ontology of neuroeconomics. It is nevertheless the case that modern economic environments and artifacts (such as money or labor contracts) are too recent to have influenced brain anatomy by evolution. If they receive a specialized and invariant treatment in the brain—as it is the tacit assumption in a lot of neuroeconomic studies—one should find a plausible explanation of that apparent evolutionary riddle according to which selective neural responses may have been evolutionarily selected in correlation with such artificial and recent behavioral and cognitive patterns. Or if one says that those neural responses simply correlate with mental processes that *happen* to be engaged in the performance of tasks that have some degree of relevance to neuroeconomics, one eschews the question of the epistemological purpose of connecting, along reverse

inferential logical patterns, neural responses, cognitive processes, and economic theoretical constructs.

Brain mechanisms, however, are 'plastic' enough to have adapted to some recently appeared stimuli, at the scale of human history. Brain plasticity refers to the ability of the brain to undergo structural and functional changes. It is a necessary process that allows our brain to learn from our environment and implement adaptive functional revisions. The anatomy of the brain could not have been influenced by recent economic environments given their too recent historical apparition. But we can conceive that the brain recycled some of its anciently implemented circuits in order to address the new set of stimuli provided by economic contexts and to functionally adapt to these novel cultural artifacts and settings. There is a shortage of data to currently support this hypothesis in the case of economic artifacts but it has been amply demonstrated in other cultural contexts such as reading and arithmetic (Dehaene and Cohen 2008). This evolutionary perspective may shift the most common research strategy in neuroeconomics from the reporting of alleged correlations between some neural patterns and some economic theoretical constructs to an inquiry on how some cognitive processes typically associated with solving economic tasks may have become associated with selective neural responses.

The hypothesis of a cultural recycling of cortical maps was put forward to make sense of a seemingly neurobiological paradox. As Dehaene and Cohen (2007) put it: "Part of the human cortex is specialized for cultural domains such as reading and arithmetic, whose invention is too recent to have influenced the evolution of our species. (…) To explain this paradoxical cerebral invariance of cultural maps, we propose a neuronal recycling hypothesis, according to which cultural inventions invade evolutionarily older brain circuits and inherit many of their structural constraints". In what does the facilitation consist and what sort of inherited constraints can affect the neural processing of cultural inventions? Central to Dehaene and Cohen's formulation of their hypothesis is the concept of a cortical map. Maps are invariant brain structures. At various possible scales, cortical maps reflect in an isomorphic way the representational structure of the targeted cultural item. These neuronal layouts have been shaped by evolution and are genetically constrained. Epigenetic factors in the early phase of the individual's development finalize the cortical structures that will react in an invariant way to some external stimuli. There thus occurs a compromise between genetic constraints, cortical relative plasticity, and the frequency and tractable structure of encountered stimuli. When neural observations collected through particular neuroeconomic studies can be related to such a hypothesis about the specific functional evolution of targeted brain areas, there is a possibility to reduce the inferential gaps between neural selectivity and the engagement of cognitive processes, on the one hand, and between the cognitive processes and their relevance to economic theoretical constructs on the other hand.

5 Conclusion

Relative lack of caution with the reverse inference fallacy leads some neuroeconomic investigations to seemingly rely on what was labeled above a "same-areas strategy":

the *same areas* that are usually dedicated to the processing of X are also observed to be activated in the processing of Y; X being a biological or cognitive function previously mapped onto some brain structures, and Y the currently examined economic construct. The shift between X and Y is problematic and if Y belongs to the description of economic behavior, in most cases it will be objectionable to say that Y and X share a common neural basis. This assertion would imply a (most often unwanted) reductionist approach, but also, most likely, a poor assessment of the selectivity of the observed neural response in view of granting the engagement of Y in the task at hand. X already possesses some domain specificity, whereas an important question, which is far from being solved, is whether Ys—i.e. conceptual constructs or cognitive processes theoretically associated with the description of economic behavior—can have any. Neuroeconomics may be seen as the study of neural correlates of behaviors that involve some cognitive and affective functions which are not specific to economic cognition, for the simple reason that there could not be such a functional description of those processes involved in economic behavior. Nevertheless, it is not vain to try to correlate economic theoretical constructs to identified neural patterns through the performance of certain cognitive or affective functions as the concerned areas may have shown enough plasticity across recent brain evolution to encode economically relevant behavior.

In the first section of this article, we have recalled what the reverse inference fallacy amounts to. In the second section, specific Bayesian constraints were outlined on the use of reverse inferences in neuroeconomics to make its conclusions more compelling. The argumentative strategy of De Quervain's investigation of the neural bases of altruistic punishment was detailed and informational constraints were spelled out in order for it to acquire more conclusiveness. An alternative experimental strategy—the use of repetition suppression paradigms—was also suggested, which presumably avoids overreliance on preliminary data about a given neural response's selectivity. In Sect. 4 the use and relevance of cognitive ontologies in neuroeconomics were discussed. Cognitive ontologies involve a double difficulty: the complexity of the engaged cognitive processes and the irrelevance of those processes vis-à-vis economic theorizing. Solutions to these two problems were suggested. Respectively, complexity may be an asset in order to increase our confidence in the selectivity of certain brain regions, and brain areas or neural responses that have been designed through long term evolution to process ancient human cognitive functions and behaviors may have optimally adapted to deal with modern economic situations. Some future neuroeconomics studies may be aligned with this "cultural recycling of cortical niches" approach. Granted this approach, possible alternative experimental strategies, and an increased attention to cognitive ontologies and neural selectivity, neuroeconomics may indeed grow as a sound scientific field.

References

Bastiaansen J, Thioux M, Keysers C (2009) Evidence for mirror systems in emotions. Phil Trans R Soc B 364:2391–2404

Bondar I, Leopold D, Richmond D, Victor J, Logothetis N (2009) Long-term stability of visual patterns selective responses of monkey temporal lobe neurons. PLOS One 4:1–10

Bourgeois-Gironde S, Schoonover C (2008) Cross-talks in economics and neuroscience. Rev Econ Polit 118:35–50
Camille N, Coricelli G, Sallet J, Pradat P, Duhamel JR, Sirigu A (2004) The involvement of the orbitofrontal cortex in the experience of regret. Science 304:1167–1170
Christoff K, Owen A (2006) Improving reverse neuroimaging inference: cognitive domain versus cognitive complexity. Trends Cogn Sci 10:352–353
Coricelli G, Nagel R (2009) Neural correlates of strategic reasoning in medial prefrontal cortex. PNAS 106:9163–9168
Coricelli G, Critchley H, Joffily M, O'Doherty L, Sirigu A, Dolan R (2005) Regret and its avoidance: a neuroimaging study of choice behavior. Nat Neurosci 8:1255–1262
De Quervain D, Fischbacher U, Treyer V, Schellhammer M, Schnyder U, Buck A, Fehr E (2004) The neural basis of altruistic punishment. Science 305:1254–1258
Dehaene S, Cohen L (2008) Cultural recycling of cortical maps. Neuron 56:384–398
Ellsberg D (1961) Risk, ambiguity, and the savage axioms. Q J Econ 75:643–699
Fehr E, Gächter S (2002) Altruistic punishment in humans. Nature 415:137–140
Fodor Jerry (1974) Special sciences: or the disunity of science as a working hypothesis. Synthese 28:97–115
Gul F, Pesendorfer W (2008) The case for mindless economics. In: Andrew C, Andrew S (eds) The foundations of positive and normative economics. Oxford University Press, Oxford
Hsu M, Ahnen C, Quartz S (2008) The right and the good: distributive justice and neural encoding of equity and efficiency. Science 320:1092–1095
Huettel S, Stowe C, Gordon E, Warner B, Platt M (2006) Neural signatures of economic preferences for risk and ambiguity. Neuron 49(5):765–775
Jeankins A, Neil Macrae C, Mitchell J (2008) Repetition suppression of ventromedial prefrontal activity during judgments of self and others. PNAS 105(11):4507–4512
Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nat Neurosci 10(12):1625–1633
Laibson D (1997) Golden eggs and hyperbolic discounting. Q J Econ 112(2):443–477 MIT Press
Loomes G, Sugden R (1982) Regret theory: an alternative theory of rational choice under uncertainty. Econ J 92:805–824
McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004) Separate neural systems value immediate and delayed monetary rewards. Science 306(5695):503–507
Padoa-Schioppia C, Assad J (2006) Neurons in orbitofrontal cortex encode economic value. Nature 441:223–226
Peirce CS (1878) Deduction, induction, and hypothesis. Pop Sci Mon 13:470–482
Poldrack RA (2006) Can cognitive processes be inferred from neuroimaging data? Trends Cogn Sci 10(2):59–63
Price CJ, Friston KJ (2005) Functional ontologies for cognition: the systematic definition of structure and function. Cogn Neuropsychol 22:262–275
Sanfey A, Rilling J, Aronson J, Nystrom L, Cohen J (2003) The Neural basis of economic decision-making in the ultimatum game. Science 300:1755–1758
Sanfey A, Loewenstein G, McClure S, Cohen J (2006) Neuroeconomics: cross-currents in research on decision-making. Trends Cogn Sci 30:108–117
Singer T, Fehr E (2005) The neuroeconomics of mind-reading and empathy. Am Econ Rev 95:340–345