DOI:
10.1177%2F0048393114530948

Neural Findings and Economic Models:

Why Brains have Limited Relevance for Economics

Acknowledgments

I wish to thank J. McKenzie Alexander, Richard Bradley, Jaakko Kuorikoski, Uskali Mäki, Caterina Marchionni, Petri Ylikoski, and three anonymous referees for their comments on earlier versions of this paper. I also received helpful feedback from audiences at the Finnish Centre for Excellence (Helsinki), the London School of Economics, and the first joint conference of the European Network for the Philosophy of the Social Sciences (ENPOSS) and the Philosophy of Social Sciences Roundtable (RPOSS).

Abstract

The proponents of neuroeconomics often argue that better knowledge of the human neural architecture enables economists to improve standard models of choice. In their view, these improvements provide compelling reasons to use neural findings in constructing and evaluating economic models. In a recent paper, I criticized this view by pointing to the trade-offs between the modelling desiderata valued by neuroeconomists and other economists respectively. The present article complements my former critique by focusing on three modelling desiderata that figure prominently in economic and neuroeconomic modelling. For each desideratum, I examine findings that neuroeconomists deem to be especially relevant for economists and argue that neuroeconomists have failed to substantiate their calls to use these findings in constructing and evaluating economic models. In doing so, I identify methodological and evidential constraints that will continue to hinder neuroeconomists' attempts to improve such models. Moreover, I draw on the literature on scientific modelling to advance the ongoing philosophical discussion regarding the prospects of interdisciplinary models of choice.

Word Count: 7095

Key Words:
Scientific Modelling; Modelling Desiderata; Interdisciplinary Relations; Modelling Pluralism; Neuroeconomics.

Introduction

The advocates of neuroeconomics (henceforth, NE) frequently argue that better knowledge of the human neural architecture enables economists to improve standard models of choice. In their view (see e.g. Camerer et al., 2005, and Glimcher, 2010, ch.4-6), these improvements provide compelling reasons to use neural findings in constructing and evaluating economic models. For their part, several economists question the relevance of neuroeconomists' (henceforth, NEs) contributions for the economic modelling of choice. In particular, some put forward methodological considerations to doubt that neural findings do (e.g. Rubinstein, 2008) or even can possibly (e.g. Gul and Pesendorfer, 2008) inform their models. Others (e.g. Harrison, 2008) contend that the evidential limitations inherent in current NE studies make it at least premature to employ neural findings in constructing and evaluating economic models.[1]

The ongoing debate between NEs and other economists has attracted increasing attention by philosophers (see e.g. Hausman, 2008, Mäki, 2010, Ross, 2008, and Vromen, 2010). In a recent article (Fumagalli, 2011), I developed a philosophical critique of NE that builds on the pragmatic and epistemic goals governing the construction and the evaluation of economic models. My critique in that article points to the trade-offs between the modelling desiderata valued by NEs and other economists respectively. The idea is that distinct desiderata often make dissimilar demands on modellers and that these trade-offs significantly constrain NEs' attempts to improve economic models. The present paper complements my former critique by focusing on three modelling desiderata that figure prominently in economic and NE modelling.

More specifically, I consider in turn descriptive accuracy, predictive power, and explanatory relevance. For each desideratum, I articulate several conceptual distinctions, examine findings that NEs deem to be especially relevant for economists, and assess the relevance of NEs' contributions for the economic modelling of choice. I shall argue for the following claims. Economists can improve some models of choice with regard to specific desiderata by using neural findings. However, NEs systematically overstate the extent to which neural findings help economists to improve their models. Moreover, even leading NEs gloss over methodological and evidential concerns that will continue to hinder their attempts to build models that supersede those developed by other economists. These considerations do not license unqualified

[1] Different approaches to NE have been developed (see e.g. Fumagalli, 2010, and Ross, 2008, for some conceptual distinctions). In this article, I speak generally of NE unless the differences between distinct approaches are material to my evaluation of NEs' proposals. Also, I employ the expressions 'economic models of choice' and 'standard economic models' broadly to indicate both individual and strategic decision theoretic models.

scepticism regarding the prospects of NE. Still, taken together, they cast serious doubts on NEs' calls to use neural findings in constructing and evaluating economic models.

Before proceeding, three preliminary remarks are in order. Firstly, the list of desiderata I examine does not encompass all the respects in which neural findings are claimed to inform standard economic modelling. In particular, it does not include NEs' attempts to provide normative evaluations of decisions. I am not concerned here with assessing normative NE analyses. For present purposes, it suffices to note that these analyses face additional difficulties besides those affecting positive NE contributions (see e.g. Fumagalli, 2013a, for a critique of recent calls to use neural findings to measure and enhance individuals' well-being).[2]

Secondly, each of the desiderata I consider has been defined in dissimilar ways in the literature on scientific modelling. This occasionally complicates the assessment of neural findings' relevance for model construction and model evaluation in economics. However, it does not preclude us from providing precise and relatively uncontroversial characterizations of the desiderata we examine. In what follows, I comment in various places on such characterizations and their interrelations. In doing so, I do not take economists' modelling practices to constitute an unquestionable normative benchmark. On the contrary, I prevalently regard such practices as a useful point of reference for explicating and assessing NEs' calls to use neural findings in economic modelling.

Finally, this article aims to advance the ongoing discussion regarding the prospects of interdisciplinary models of choice in at least three respects of general interest to philosophers of the social sciences. Firstly, it provides a philosophically informed evaluation of recent advances at the interface between economics, psychology and neuroscience that are potentially relevant to social scientific research. Secondly, it draws novel connections between parallel debates about the modelling of choice that are still insufficiently integrated across distinct decision sciences. And thirdly, it combines considerations from neuro-psychological studies, economic methodology and philosophy of science to develop a systematic critique of prominent calls in favour of interdisciplinary models of choice.

[2] NEs occasionally advocate using neural findings to support - rather than improve or substitute - standard economic models of choice (see e.g. Camerer, 2008a). I shall focus predominantly on NEs' attempts to improve or substitute standard economic models, as most debates between NEs and other economists target these contributions.

1. Descriptive Accuracy

A model's descriptive accuracy can be evaluated along several dimensions. Two senses of 'descriptive accuracy' are often distinguished in the literature on scientific modelling (see e.g. Weisberg, 2007). On the one hand, there is the question whether a model represents the relevant properties and entities figuring in its target system.[3] On the other hand, one can assess how detailed a model's characterization of each of these properties and entities is. These two senses of descriptive accuracy are conceptually distinct. For instance, a model may represent most of the properties of its target system, but fail to characterize these properties in detail. Another model, instead, may represent just a small subset of those properties, yet provide a detailed characterization of them.[4]

NEs frequently claim that neural findings enable economists to increase the descriptive accuracy of their models in both of these respects. In particular, they argue that economists can obtain significant predictive and explanatory benefits by building models that provide accurate characterizations of the neural substrates of decisions (see e.g. Camerer et al., 2005, and Glimcher, 2010, ch.5-6). Their reasoning goes as follows. Standard economic models omit neural entities and properties that are known to influence decisions. Moreover, they make implausible assumptions concerning agents' cognitive and computational abilities. Due to these limitations, several economic models fail to be predictive and explanatory. Fortunately, neural findings enable economists to build models that represent neural entities and properties that are known to influence decisions. Furthermore, they facilitate the construction of models that make more plausible assumptions concerning agents' cognitive and computational abilities. These modelling benefits, in turn, help economists to construct predictive and explanatory models. Hence, economists should employ neural findings in building their models.

These considerations point to some of the alleged benefits NEs associate with models that accurately characterize the neural substrates of decisions. However, they do not provide economists with convincing grounds to use neural findings in constructing their models. Below I

[3] What entities and properties are deemed to be relevant in a given context may importantly depend both on the modellers' aims and on what notion of relevance they presuppose (see e.g. Mäki, 2009, and Matthewson and Weisberg, 2009). I shall explicate how these two issues bear on the merits of NEs' contributions in various places throughout the paper. For now, it suffices to note that the modelling practices of NEs and other economists differ remarkably with regard to both of those issues.

[4] I speak of models 'representing' or 'omitting' entities and properties without specifying what modelling devices (e.g. adding or removing variables, altering functional forms) modellers may use to do so. I shall comment on the differences between these modelling devices whenever the cogency of my claims rests on such differences.

identify and discuss two reasons in support of this claim. Firstly, building predictive and explanatory models of choice does not require economists to provide accurate characterizations of the neural substrates of decisions. And secondly, it is questionable whether NEs are in the epistemic position to provide accurate characterizations of these substrates.

i) Economists investigate their target systems at different levels of detail and for dissimilar purposes. In particular, they frequently build predictive and explanatory models by isolating and distorting specific features of their phenomena of interest (see e.g. Cartwright, 2009, and Mäki, 2009). Moreover, they often acquire predictive and explanatory insights by relying on models which do not accurately characterize any actual phenomena, but rather target hypothetical or even counterfactual states of affairs (see e.g. Gibbard and Varian, 1978, and Sugden, 2000 and 2009). For these reasons, the mere fact that a choice model does not provide accurate characterizations of the neural substrates of decisions does not preclude such model from yielding predictive and explanatory insights to economists.

A proponent of NE may concede that choice models that do not provide accurate characterizations of the neural substrates of decisions *can* be predictive and explanatory for economists. At the same time, she might object that building predictive and explanatory models *typically* requires economists to provide accurate characterizations of such substrates. Unfortunately, even this claim fails to withstand scrutiny. By way of illustration, consider the often-made assumption that agents calculate the expected utility of the available options using perfect Bayesian updating with negligible cognitive costs. This assumption clearly exaggerates human individuals' cognitive and computational abilities. Yet, it enables economists to predict decisions across a wide range of settings (e.g. think of repetitive choices). Furthermore, economists have made significant predictive advances by building models that, while relaxing such assumption, do not make any claim concerning the neural substrates of decisions (see e.g. Machina, 1987, and Starmer, 2000, for a review).[5]

Analogous remarks apply to economic models that deliberately omit neural entities and properties that are known to influence decisions. To see this, let us focus on the issue of preference change. Standard economic models posit agents having exogenously fixed preferences, which do not vary unless such agents learn novel information about their choice options. However, the preferences of real-life individuals are frequently constructed during decision-making (see e.g. Guala,

[5] A proponent of NE might object that these predictive advances do not exclude that NEs could further improve economists' predictions by constructing models that provide accurate characterizations of the neural substrates of decisions. However, NEs still have to substantiate this conjecture (see *Section 2* on some major limitations affecting most NEs' predictions).

2005, ch.5). Moreover, not all instances of preference change seem prompted by information learning. For instance, an agent's preferences may vary depending on which of her options' features are salient, with changes in what features are salient occurring in the absence of variations in the agent's information set.

Now, economists' reliance on models positing exogenously fixed preferences is rarely due to their alleged ignorance of preferences' variability. On the contrary, this practice is usually motivated by specific modelling goals and considerations (e.g. simplicity). Moreover, several economic models of preference change have been developed to deal with cases where standard models fail to be predictive and explanatory. For example, Dietrich and List (2011) model situations where preferences can vary without agents learning any novel information about their choice options. Their contribution generalizes standard economic models by providing a representation theorem for both information-based and non-informational preference change.

The aforementioned examples point to a more general reason why building predictive and explanatory models does not require economists to provide accurate characterizations of the neural substrates of decisions. This reason relates to the so-called derivational robustness of economic models' implications with respect to variations in assumptions concerning the neural substrates of decisions. The expression 'derivational robustness' indicates the degree to which a model's implications hold under variations in the assumptions used to derive them (see e.g. Woodward, 2006). The implications of economic models are often robust under significant variations in assumptions concerning the neural substrates of decisions. To give one example, consider a famous series of computerized market experiments by Gode and Sunder (1993), who replace human traders with 'zero-intelligence' programs that submit random bids and offers. As shown by Gode and Sunder, imposing a simple budget constraint on trades robustly leads to market allocative efficiency irrespective of traders' cognitive and computational abilities (see Lehtinen and Kuorikoski, 2007, for additional examples).

A proponent of NE may concede that the implications of several economic models are robust under variations in assumptions concerning the neural substrates of decisions. At the same time, she might insist that economic models that neglect information about such substrates often fail to be predictive and explanatory. In particular, she might contend that the implications of economic models of choice frequently change in response to variations in behavioural or psychological assumptions. This latter contention is not without merit (see e.g. Goeree and Holt, 2001, on how alterations in payoff structures can generate inconsistencies between game theoretic predictions and observed strategic behaviour). Still, it does not *per se* support the claim that building predictive and explanatory models requires economists to provide accurate characterizations of the neural substrates of decisions.

In this respect, it is telling that NEs rarely attempt to demonstrate that limitations in the predictive and explanatory performance of economic models are due to these models' failure to provide accurate characterizations of those substrates.[6]

ii) Suppose - for the sake of argument - that NEs demonstrated that building predictive and explanatory models of choice requires economists to provide accurate characterizations of the neural substrates of decisions. Even so, the question would remain as to whether NEs are in the epistemic position to provide accurate characterizations of such substrates. Over the last few years, NEs have claimed to measure several variables (e.g. neural correlates of specific choices) that economists previously regarded as unobserved or even unobservable. As Camerer puts it, NE "is not in opposition to rational choice theory, but sees potential in extending its scope by observing variables that are considered inherently unobservable in [it]" (2008b, 45).

These alleged advances may be interesting to choice modellers, but do not substantiate the claim that NEs are in the epistemic position to provide accurate characterizations of the neural substrates of decisions. There are at least two reasons to doubt such claim. The first reason relates to well-known limitations in the accuracy and reliability of the tools used to collect and interpret the data on which NEs' models are based (see e.g. Logothetis, 2008, on the haemodynamic proxies used to estimate neural activity in fMRI studies). The idea is that the limitations inherent in NEs' observational tools hamper their attempts to provide accurate characterizations of the neural substrates of decisions and that providing these characterizations requires NEs to alleviate or circumvent such limitations (see e.g. Fumagalli, 2013b).

The second reason to doubt that NEs are in the epistemic position to provide accurate characterizations of the neural substrates of decisions concerns the complexity of these substrates.[7] This point can be

---

[6] Intense debates have taken place in the philosophy of science concerning the epistemic import of robustness analysis. In particular, some (e.g. Odenbaugh and Alexandrova, 2011) allege that robustness analysis is best regarded as a method of discovery of robust theorems. Others (e.g. Kuorikoski et al., 2010 and 2012) argue that robustness analysis can justifiably increase modellers' degree of confidence in the robust theorems which connect their models' substantial assumptions to specific modelling results. I do not take a position in this debate for the purpose of my article.

[7] Several notions of complexity have been distinguished by philosophers of science. For instance, Mitchell alleges that a target system is: *compositionally* complex, if it is "constituted by a [high number of] nonrandomly structured parts"; *dynamically* complex, if it is "the location for multiple, interacting causes some of which are represented as […] non-linear functions"; and *evolutionarily* complex, if it displays "a variety of historically contingent, adaptive responses to environmental challenges" (2002, 58). The neural substrates of decisions can be regarded as a rather complex target system in all these three senses.

explicated as follows. The workings of the human neural architecture are influenced by a number of factors, ranging from the presence of particular environmental cues to the momentary availability of metabolic resources (see e.g. Wilcox, 2008). Interactions between brain regions exhibit high degrees of anatomical and functional interdependence in terms of involved neural populations, connectivity patterns, and so on (see e.g. Cisek, 2012). For these reasons, specific NE models can endogenize a small subset of the neural entities and properties that influence individuals' decisions (e.g. think of choices between long-term, non-stimulus-bound alternatives).

To be sure, advances in scanner technology and experimental design may help NEs to build models that endogenize a higher number of neural entities and properties. Moreover, NEs might provide increasingly accurate characterizations of the neural substrates of decisions by using clusters of models that yield complementary insights about such substrates (see e.g. Bechtel, 2002, on the opportunity to combine brain-imaging and brain-stimulation instruments to support conjectures about the human neural architecture). Still, the point remains that specific NE models are typically bound to represent only a small subset of the entities and properties figuring in their target systems. In this respect, it is worth noting that even models of basic sensorimotor tasks make deliberately inaccurate assumptions about their targets and include several free parameters (see e.g. Fernandes and Kording, 2010).[8]

2. Predictive Power

The predictive power of a model can be evaluated in several respects. Choice modellers ordinarily distinguish between: predictive *accuracy*, which refers to how close a model's predictions are to the actual value of the examined variables; predictive *robustness*, which concerns the stability of a model's predictive performance across distinct settings; and predictive *reliability*, which relates to the degree to which a model's predictions can be expected to hold in a particular context. These three notions are not unrelated, yet point to different aspects of models' predictive power. For instance, a model may yield very accurate predictions, but only in specific settings. Another model, instead, may make less accurate predictions that hold across a wider range of contexts. Furthermore, each notion of predictive power can be

[8] NE models that represent a small subset of the entities and properties figuring in their target systems may still count as descriptively accurate, provided that these entities and properties are relevant to the modelling task at hand and are given a detailed characterization by such models. However, as I argue in *Sections 2* and *3*, NEs rarely put forward models that, while representing just a few neural entities and properties, are predictive and explanatory for other economists.

evaluated along different dimensions. For example, one may deem a model's predictions to be more or less accurate depending on whether she considers the occurrence of the predicted events, the magnitude of these events, or the time at which those events occur.

NEs often criticize economic models for failing to provide accurate, robust and reliable predictions of choices (see e.g. Loewenstein et al., 2008, and McCabe, 2008). Moreover, they argue that economists could obtain substantial predictive gains in these respects by using neural findings. Some even take these purported gains to constitute NE's most significant contribution to the economic modelling of choice. As Camerer puts it, "the largest payoff from [NE] may come from pointing to biological variables which have a large influence on behaviour and are underweighted or ignored in standard theory" (2007, C35). To be sure, some authors (e.g. Rustichini, 2009) acknowledge that economists may make accurate predictions without having detailed knowledge of the neural substrates of decisions. Yet, most proponents of NE insist that models that do not provide accurate characterizations of these substrates frequently fail to predict choices (see e.g. Schotter, 2008).

At first glance, one may expect economists to welcome NEs' insistence on predictive gains. After all - the thought would be - economists of all stripes ascribe a prominent relevance to predictive considerations. Nonetheless, at least two major limitations almost invariably constrain the relevance of NEs' predictions for other economists. Firstly, most NE studies make predictions over exceedingly limited time spans to be valuable to economists. And secondly, NEs' attempts to improve economists' predictions face some severe challenges. Let us examine these two limitations in turn.

i) If NEs are to substantiate their calls to use neural findings to make economic predictions, their rationale should go beyond the econometric platitude that when one expands the set of explanatory variables in a model, the correlation between these variables and the dependent variable is almost bound to increase. For one could increase this correlation by using variables that few economists would employ to predict choices.[9] To give one example, the alleged fact that sunshine significantly correlates with daily stock returns in several countries (see e.g. Hirshleifer and Shumway, 2003) falls short of implying that economists should include meteorological variables in their models of the financial market. Over the last decade, NEs have provided some persuasive illustrations of how neural findings may help choice

[9] Such correlation may not increase if the explanatory variable one includes is statistically independent from the dependent variable. Indeed, one might even worsen a model's predictive performance by incorporating more variables into it. The idea, which relates to the so-called problem of overfitting, is that including more variables into a model could render it exceedingly sensitive to idiosyncrasies in the initial data set, thereby leading it to make predictions that are inferior to those it would have otherwise made in additional data sets (see e.g. Hitchcock and Sober, 2004).

modellers to predict individuals' decisions. Three classes of contributions seem particularly significant in this context.

The first class of contributions attempt to predict choices by means of experimentally regimented observations of specific areas' activations. For example, Knutson et al. (2007) illustrate that activations observed in the nucleus accumbens, the insula, and the mesial prefrontal cortex occasionally predict purchase decisions even better than self-report variables. The second class of contributions purport to improve choice modellers' out of sample predictions thanks to more detailed knowledge of the neural substrates of decisions. The idea is to use the neural data gathered in specific experimental settings to better predict decisions in other experiments and real-life situations. For instance, in a study of the neural basis of altruistic punishment, de Quervain et al. (2004) predict variations in punishers' willingness to pay for punishment when such punishment is costly for them on the basis of activations observed in the caudate when punishment is costless for these punishers.

The third class of contributions aim to induce predictable variations in individuals' choices by stimulating or disrupting activations in neural areas that putatively contribute to the execution of such choices. By way of illustration, let us focus on some studies of the neural substrates of strategic decision-making. In an often-quoted fMRI study, Sanfey et al. (2003) document that the right dorsolateral prefrontal cortex (DLPFC), an area frequently taken to serve cognitive control and goal maintenance, undergoes increased activation in presence of unfair offers in ultimatum games. Building on this finding, Wout et al. (2005) demonstrate that disrupting activation in the right DLPFC via repetitive TMS predictably alters both subjects' propensity to reject unfair offers and their reaction times for rejecting unfair offers. This study extends previous correlational results by corroborating hypotheses about the causal role of the right DLPFC in strategic decision-making.

These predictive accomplishments are promising, but are rarely shown by NEs to yield significant predictive benefits to other economists. My point is not just that NEs' use of small experimental samples and multi-stage data manipulation techniques constrains the reliability and the robustness of their predictions (see e.g. Harrison, 2008). Rather, my main concern relates to the limited time spans covered by such predictions. As noted in *Section 1*, individuals' choices result from the dynamic interactions of several neural populations and factors operating at multiple levels of organization. For this reason, monitoring specific areas' activations rarely enables NEs to reliably predict individuals' choices beyond a range of several hundred milliseconds. This temporal interval is exceedingly limited for NEs' predictions to be relevant for other economists, who typically aim to predict choices over much longer time spans. Yet, even leading NEs recurrently gloss over this limitation when presenting their findings.[10]

A proponent of NE may concede that NEs rarely succeed in predicting individuals' choices beyond limited time spans. At the same time, she might claim that NEs can make significant predictive advances in this respect by categorizing agents into different types. This claim relates to the following two-step procedure (see e.g. Rubinstein, 2008). Firstly, neural findings are used to identify characteristics (e.g. attitudes to risk) that supposedly predict agents' behaviour across choice domains. And secondly, the distribution of agents' types is employed as a primitive in models that predict choices probabilistically over non-negligible temporal intervals. Regrettably, NEs have hitherto failed to make significant predictive advances in this respect. Moreover, it is questionable whether much progress is to be expected on this issue. For the characteristics examined by NEs rarely seem stable enough across individuals, time spans and choice domains to provide robust agent-type categorizations (see e.g. Holt and Laury, 2002, on the influence of various incentives on individuals' risk preferences).

ii) Suppose that NEs identify some neural area whose activations exhibit statistically significant and robust correlations with observed decisions. Assume, further, that NEs can reliably predict choices over extended time spans by examining this area's activations. This would be a remarkable predictive achievement for NEs. In this respect, it would be of little import to object that predicting choices from endogenous neural activity is "largely orthogonal" to economists' predictive goals (Bernheim, 2009, 23). For although few economists currently attempt to predict choices from endogenous neural activity, the aforementioned predictive achievement would provide them with a convincing reason to use neural findings for predictive purposes. Unfortunately, no such predictive achievement is in sight for NEs. Moreover, NEs' attempts to improve economists' predictions face some severe challenges. To illustrate this, let us consider briefly a few respects in which neural findings are claimed to yield predictive benefits to economists.

NEs frequently maintain to have identified anatomically delimited neural populations whose activations predict variables of economic interest. For instance, Levy and Glimcher (2012) contend that specific

[10] Some NE studies report correlations between specific areas' activations and economic variables over extended temporal intervals. For instance, after monitoring a few adolescents who listen to songs of unknown artists, Berns and Moore (2012) find that ventral striatum's activations significantly correlate with these songs' sales among the general public several months after scanning. Regrettably, these results are not shown to hold across samples and choice situations. Moreover, the authors do not specify how much testing was implemented before obtaining those results, and provide no reason to exclude that their study reports a rather arbitrary selection of unrepresentative findings.

sub-regions in the ventromedial prefrontal cortex and the orbitofrontal cortex encode the subjective values of individuals' choice options on a common neural scale. This, in turn, allegedly enables NEs to predict choices across various types of stimulus-bound rewards and experimental conditions. Even so, one should not overemphasize the predictive relevance that observed correlations between choices and neural activations have for other economists. For NEs' studies have predominantly focused on experimental tasks that are far less sophisticated and computationally demanding than decision problems commonly faced by human individuals (see e.g. Fumagalli, 2013a). In particular, there is no compelling reason to think that the same neural mechanisms that contribute to stimulus-bound reward valuation determine choice in circumstances where individuals face non-stimulus-bound alternatives (see e.g. Fumagalli, 2013b).

Analogous considerations hold for NEs' calls to employ neural findings to predict aggregate choice patterns or impute missing economic variables. For *in primis*, it appears to be prohibitively difficult to collect the population-level neural findings (e.g. large-scale interpersonal distributions of specific areas' activations) needed to reliably predict aggregate choice patterns across diverse decision settings. And secondly, neural findings are especially hard to obtain in situations where also choice data are missing or inaccurate. To put it differently, it seems rather unlikely that we can perform detailed investigations of subjects' brain activity, but cannot collect data about their choices and preferences. More generally, NEs have not offered yet any convincing reason to expect that NEs will build "a mechanistically accurate economic theory which is by necessity predictive" (Glimcher et al., 2005, 221).

3. Explanatory Relevance

NEs frequently claim that neural findings provide economists with highly explanatory insights. For instance, Zak contends that NE research "will allow economists to answer fundamental questions they are unable to address" (2004, 1738). Similarly, Brocas and Carrillo allege that NE studies yield "new reliable theories capable of explaining […] individual behaviour and strategic choices" (2010). Substantiating these claims would require NEs to specify what features they take to be constitutive of adequate explanations, or at least indicate what account of explanation they presuppose. NEs do not always indicate what account of explanation they presuppose when claiming that their findings are explanatory for economists. However, as illustrated below, three accounts of explanation - namely the unificationist, mechanistic, and interventionist accounts - figure prominently in NEs' works. Let us assess the explanatory relevance of NEs' contributions with regard to each of these three accounts in turn.

i) According to the *unificationist* account, explanation consists in disclosing connections between phenomena that were formerly regarded as unrelated (see e.g. Friedman, 1974, and Kitcher, 1981). The aim is to demonstrate that seemingly unrelated phenomena are, at bottom, manifestations of a common system of entities, processes, and so on. In recent years, several NEs have advocated using neural findings to account for diverse economic phenomena in terms of some basic neuro-cognitive and neuro-computational principles (see below). Some (e.g. McCabe, 2003) maintain that more accurate knowledge of the human neural architecture enables economists to develop more general models of choice, which explain a wider variety of decisions than standard economic models. In their view, the "ultimate goal" of NE is "to produce [a] detailed computational and neurobiological […] common foundation for understanding human behavior across the natural and social sciences" (Fehr and Rangel, 2011, 4).[11]

Do NEs substantiate their unificationist contentions? It does not seem so. Indeed, even leading NEs appear to grossly overstate the explanatory relevance of their unificationist contributions for economic modellers. The following example, which concerns putative NE explanations of macroeconomic phenomena, nicely illustrates this point. NEs often conjecture that neural findings help economists to account for aggregate choice patterns (see e.g. Camerer, 2007, C29). McCabe goes as far as to assert that such findings may enable economists to better explain "the disparity of economic growth, and material welfare, both between and within nations" (2008, 348).

Now, one may certainly tell some story of how neural activations in individuals' brains indirectly influence macroeconomic phenomena by shaping individuals' choices and preferences. Yet, when it comes to accounting for macroeconomic growth and welfare, NEs' findings appear to have marginal explanatory relevance for other economists. For the influence of neural activations on these phenomena is mediated by a number of factors - ranging from institutional incentives to individuals' axiological commitments - that can be only partly controlled for in NE studies. That is to say, too many and overly speculative inferential steps are required to account for macroeconomic phenomena in terms of the neural activations observed in individuals' brains.

A proponent of NE may concede that some NEs overstate the explanatory relevance of their findings for economic modellers. Still,

[11] Models' generality can be appraised with regard to both actual and hypothetical targets. In particular, one may distinguish between *a-generality* - which concerns how many actual targets a model applies to - and *p-generality* - which relates to how many logically, nomologically or physically possible targets a model applies to (Matthewson and Weisberg, 2009, 182). NEs prevalently speak of generality in the former sense, but occasionally use this term in the latter sense as well.

she might insist that NE models can be employed to represent a wide variety of choices and therefore serve as useful springboards for explanatory unifications. This claim invites the following two-fold rejoinder. Firstly, it is doubtful that NE models can be employed to represent a greater variety of choices than standard economic models (e.g. think of the flexibility of economists' constrained optimization techniques). And secondly, the mere fact that a model enables one to represent a vast array of phenomena by no means implies that such model advances the development of explanatory - as opposed to merely descriptive - accounts of those phenomena (see e.g. Kitcher, 1989, sec.4-5, for similar remarks).

ii) Mechanistic explanations account for why a target system of interest exhibits particular features by pointing to the features and the interactions of its components (see e.g. Bechtel, 2008, ch.1, and Bechtel and Richardson, 1993, ch.2). On this account, a choice model is explanatory if it accurately identifies the possibly multilevel (e.g. psychological, neural) mechanisms whose interactions influence agents' decisions in the investigated settings (see e.g. Craver, 2006, and Craver and Alexandrova, 2008). Several NEs urge economists to "ground economic theory in detailed neural mechanisms" for both predictive and explanatory purposes (Camerer, 2007, C26). Some even claim to have identified the basic mechanistic components of reward valuation systems across primates (see e.g. Levy and Glimcher, 2012). I am not concerned here with evaluating these claims. For present purposes, I critically examine the grounds on which NEs take their mechanistic insights to be explanatory for other economists.[12]

NEs often appear to presuppose that information concerning the neural aetiology and antecedents of decisions is explanatory for other economists. Some seemingly take it for granted that acquiring mechanistic insights about the neural substrates of decisions *ipso facto* increases their ability to account for individuals' choice behaviour. For example, after noting that "all economic activity flows through the brain" Camerer conjectures that "understanding brain function [will] be useful for understanding […] economic choice" (2008b, 47). Similarly, Zak takes it to be "self-evident" that since "economic decisions are made in the brain […] economic regularities can be understood, and economic models are based, on cognitive science" (2008, 301). Let us assess the cogency of these claims.

[12] Various definitions of mechanism have been proposed. For instance, Machamer et al. characterize mechanisms as "entities and activities organized such that they are productive of regular changes" in the examined target systems (2000, 3). For his part, Glennan defines them as "complex system[s]" whose components produce the phenomena of interest via interactions that "can be characterized by direct, invariant, change-relating generalizations" (2005, 445). These definitions are not equivalent, yet are usually taken to be sufficiently similar to speak of a mechanistic account of explanation in a unified sense.

NEs' studies of the neural substrates of decisions provide economists with several mechanistic insights that were previously unavailable to them. This, however, falls short of implying that NEs' mechanistic insights are explanatory for economists. To see why, consider how variations in criteria of explanatory relevance across economics and neuroscience bear on the explanatory import of NEs' contributions. What insights are explanatory for a modeller typically depends on a number of factors, ranging from the theoretical constructs she employs to her modelling purposes. NEs' and other economists' contributions frequently differ in these respects. In particular, their models are often premised on dissimilar assumptions as to whether mechanistic information is required to provide adequate explanations of choice and, if so, at what levels of description such mechanistic information is to be obtained. For these reasons, mechanistic insights concerning the neural substrates of decisions may simultaneously be highly explanatory for NEs, yet fail to be explanatory for other economists. To illustrate this, let us focus on some NE accounts of preference transitivity.

Standard economic models of choice posit agents having transitive preferences. Yet, real-life individuals occasionally exhibit preferences that violate this transitivity requirement (see e.g. Loomes and Sugden, 1982). These violations can occur not just due to individuals' mistakes, but also because the subjective values individuals assign to choice options vary depending on what alternatives are available to them. Over the last few years, some studies indicated that value representations in the orbitofrontal cortex, whose activations allegedly predict choices across various rewards (see *Section 2*), are invariant for changes of menu (see e.g. Padoa-Schioppa and Assad, 2008). Furthermore, the stated preferences of individuals with lesions in this region have been shown to violate transitivity more often than those of control subjects (see e.g. Fellows and Farah, 2007). These findings have led some authors to conjecture that when choices are based on values encoded in the orbitofrontal cortex, menu invariance constitutes "the neurobiological origin of preference transitivity" (Padoa-Schioppa, 2011, 344).

The aforementioned findings seem explanatory for NEs on the mechanistic account of explanation we are considering. At the same time, they do not appear to be explanatory for other economists. After all, standard economic theory remains agnostic regarding the neuro-biological substrates of preference transitivity. Hence, information about such substrates does not directly speak to economists' explanatory concerns. This, in turn, forces NEs to explicate why exactly findings about the neuro-biological substrates of preference transitivity would be explanatorily relevant for other economists. In this respect, it is worth emphasizing that the mere fact that some neural areas' activations respect menu invariance by no means implies that the corresponding individuals will exhibit transitive preferences. For individuals' preferences are typically influenced by the interactions of several areas, and only some of these areas' activations are invariant for changes of menu (see e.g. Kable and Glimcher, 2009, and Padoa-Schioppa, 2011).

iii) The *interventionist* account relates explanation to the identification of patterns of counterfactual dependence between one's explanandum variable and explanans variables (see e.g. Woodward, 1997 and 2003). On this account, a model's explanatory relevance depends on the range of what-if-things-had-been-different questions (w-questions) it answers regarding how hypothetical or counterfactual variations in the values of the explanans variables affect the value of the explanandum variable. In particular, a model's generalizations are explanatory to the extent that they are invariant under interventions, i.e. manipulations that directly affect only the explanans variables of interest and do not correlate with any other cause of the explanandum variable.

Prima facie, several NE contributions appear to fare quite well on the interventionist account. For instance, Kosfeld et al. (2005) document how inhaling oxytocin can significantly increase individuals' propensity to trust others, yet without altering their risk preferences. Similarly, Pessiglione et al. (2006) show how administering drugs that respectively enhance (e.g. L-DOPA) and reduce (e.g. haloperidol) dopaminergic function modulates subjects' ability to choose the most rewarding options across decision settings. These studies provide NEs with informative insights regarding how individuals' choices can be altered by means of selective and reversible neurochemical manipulations. Even so, it is questionable whether such insights are explanatory for other economists. For disclosing patterns of counterfactual dependence between observed choices and neural activations does not *per se* enable one to identify invariant generalizations between economic variables.

A proponent of NE may rebut that neural findings help economists to answer a wide range of w-questions concerning hypothetical and counterfactual variations in individuals' decisions. In particular, she might contend that neural findings are more explanatory than behavioural and psychological findings, since they enable modellers to answer more w-questions regarding agents' decisions than these other findings. Regrettably, NEs have not substantiated these assertions. Moreover, some claims to the contrary have been put forward in the literature. For instance, as persuasively illustrated by Kuorikoski and Ylikoski (2010), identified correspondences between neural activations and observed decisions do not enable NEs to answer a wide range of w-questions outside the investigated experimental settings. Furthermore, choice models can be frequently made more explanatory in the interventionist sense by abstracting away from the neural substrates of decisions. Let me explicate this point.

NEs can identify relatively fine-grained correspondences between decisions and neural activations by demonstrating how decisions vary under experimentally regimented interventions on anatomically delimited neural areas. However, choice modellers can frequently answer a wider range of w-questions about agents' decisions by targeting variations in psychological (e.g. motivations) rather than neural variables. Moreover, information concerning agents' environmental constraints and incentives usually allows choice modellers to make generalizations that are invariant under a broader set of interventions than generalizations based on neural findings. To give one example, consider recent studies of the neuro-biological substrates of pathological gambling (see e.g. Holst et al., 2010, and Potenza, 2008).

The mere fact that pathological gamblers exhibit similar addictive propensities does not imply that they all share some common neuro-biological syndrome that explains their behavioral similarities (Ross et al., 2008, ch.1). On the contrary, information regarding those agents' environmental constraints (e.g. social sanctions) and incentives (e.g. visual cues triggering addictive responses) typically enables one to account for their behavioural similarities across a wider range of situations (see e.g. Robinson and Berridge, 2003, on the influence that exposure to cues can have on addicts' behaviour). This does not preclude NEs from identifying informative generalizations concerning the neuro-biological substrates of pathological gambling (see e.g. Ross, 2009, on how combining behavioural and neural findings allows researchers to answer a growing range of w-questions about agents' addictive propensities). Yet, these substrates seem exceedingly heterogeneous across individuals and choice settings for generalizations based on neural findings to be invariant under a broader set of interventions than generalizations based on other disciplines' findings.

Conclusion

NEs' calls to use neural findings in constructing and evaluating economic models often rest on the claim that such findings enable economists to improve their models of choice. In this article, I examined three respects in which NEs take neural findings to do so. The idea is that these findings enable economists to: increase the descriptive accuracy of their models (*Section 1*); provide more accurate, robust and reliable predictions of choices (*Section 2*); and build more explanatory models of choice (*Section 3*). As we have seen, economists can improve some models of choice with regard to specific desiderata by using neural findings. At the same time, there are at least two reasons to doubt NEs' calls to employ such findings in constructing and evaluating economic models.

The first reason, explored in Fumagalli (2011), is that due to the trade-offs between the desiderata respectively valued by NEs and other economists, showing that neural findings help economists to satisfy specific desiderata falls short of implying that NEs enable economists to build better models of choice. The second reason, discussed in the present paper, is that NEs have hitherto failed to significantly improve economic models even with regard to individual desiderata. Moreover, they gloss over methodological and evidential concerns that will continue to hinder their attempts to build models that supersede those developed by other economists. These concerns do not license unqualified scepticism regarding the prospects of NE. Still, taken together, they cast serious doubts on NEs' calls to employ neural findings in constructing and evaluating economic models.

References

Bechtel, W. 2002. Aligning multiple research techniques in cognitive neuroscience: Why is it important? *Philosophy of Science*, 69, S48-S58.

Bechtel, W. 2008. *Mental mechanisms: Philosophical perspectives on cognitive neuroscience*. Routledge.

Bechtel, W. and Richardson, R.C. 1993. *Discovering complexity: Decomposition and localization as scientific research strategies*. Princeton University Press.

Bernheim, B.D. 2009. On the Potential of Neuroeconomics: A Critical (but Hopeful) Appraisal. *American Economic Journal: Microeconomics*, 1(2), 1-41.

Berns, G.S. and Moore, S.E. 2012. A neural predictor of cultural popularity. *Journal of Consumer Psychology*, 22, 154–160.

Brocas, I. and Carrillo, J.D. 2010. Neuroeconomic theory: Using neuroscience to understand the bounds of rationality. Available at: http://www.voxeu.org/index.php?q=node/4758, 18 March 2010.

Camerer, C.F. 2007. Neuroeconomics: Using Neuroscience to Make Economic Predictions. *The Economic Journal*, 117(519), C26–C42.

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

Camerer, C.F. 2008b. The Case for Mindful Economics. In Caplin, A. and Schotter, A. Eds. *The foundations of positive and normative economics*. New York: Oxford University Press, 43-69.

Camerer, C.F., Loewenstein, G., Prelec, D. 2005. Neuroeconomics: how neuroscience can inform Economics. *Journal of Economic Literature*, 43(1), 9-64.

Cartwright, N. 2009. If no capacities then no credible worlds. But can models reveal capacities? *Erkenntnis*, 70 (1), 45-58.

Cisek, P. 2012. Making decisions through a distributed consensus. *Current Opinion in Neurobiology*, 22, 927–936.

Craver, C.F. 2006. When mechanistic models explain. *Synthese*, 153, 355-376.

Craver, C.F. and Alexandrova, A. 2008. No Revolution Necessary: Neural Mechanisms for Economics. *Economics and Philosophy*, 24, 381-406.

De Quervain, D.J., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A. and Fehr, E. 2004. The neural basis of altruistic punishment. *Science*, 305, 1254–1258.

Dietrich, F. and List, C. 2011. A model of non-informational preference change. *Journal of Theoretical Politics*, 23 (2), 145-164.

Fehr, E. and Rangel, A. 2011. Neuroeconomic Foundations of Economic Choice - Recent Advances. *Journal of Economic Perspectives*, 25(4), 3–30.

Fellows, L.K. and Farah, M.J. 2007. The role of ventromedial prefrontal cortex in decision making: judgment under uncertainty or judgment per se? *Cerebral Cortex*, 17, 2669–74.

Fernandes, H.L. and Kording, K.P. 2010. In praise of 'false' models and rich data. *Journal of Motor Behavior*, 42 (6), 343-9.

Friedman, M. 1974. Explanation and Scientific Understanding. *Journal of Philosophy*, 71, 5-19.

Fumagalli, R. 2010. The Disunity of Neuroeconomics: a Methodological Appraisal. *Journal of Economic Methodology*, 17 (2), 119-131.

Fumagalli, R. 2011. On the Neural Enrichment of Economic Models: Tractability, Trade-offs and Multiple Levels of Descriptions. *Biology and Philosophy*, 26, 617-635.

Fumagalli, R. 2013a. The Futile Search for True Utility. *Economics and Philosophy*, 29, 325-347.

Fumagalli, R. 2013b. Choice Models and Neuro-Psychological Ontologies: Three Challenges to Realisticness. Unpublished Manuscript.

Gibbard, A. and Varian, H.R. 1978. Economic Models. *The Journal of Philosophy*, 75 (11), 664-677.

Glennan, S.S. 2005. Modeling mechanisms. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36(2), p.443–464.

Glimcher, P.W. 2010. *Foundations of Neuroeconomic Analysis*. Oxford University Press.

Glimcher, P.W., Dorris, M.C. and Bayer, H.M. 2005. Physiological utility theory and the Neuroeconomics of choice. *Games and Economic Behavior*, 52, 213–256.

Gode, D.K. and Sunder, S. 1993. Allocative efficiency of markets with zero intelligence traders: Market as a partial substitute for individual rationality. *Journal of Political Economy*, 101 (1), 119-37.

Goeree, J.K. and Holt, C.A. 2001. Ten Little Treasures of Game Theory and Ten Intuitive Contradictions. *American Economic Review*, 91(5), 1402-1422.

Guala, F. 2005. *The methodology of experimental economics*. Cambridge University Press.

Gul, F. and Pesendorfer, W. 2008. The case for mindless economics. In Caplin, A. and Schotter, A. Eds. *The foundations of positive and normative economics*. Oxford University Press, 1-40.

Harrison, G. 2008. Neuroeconomics: A Critical Reconsideration. *Economics and Philosophy*, 24, 303-344.

Hausman, D.M. 2008. Mindless or Mindful Economics: a Methodological Evaluation. In *The Foundations of Positive and Normative Economics: A Handbook*, Ed. Caplin, A. and Schotter, A. Oxford University Press, Ch.6, 125-152.

Hirshleifer, D.A. and Shumway, T. 2003. Good day sunshine: stock returns and the weather. *Journal of Finance*, 58, 1009-32.

Hitchcock, C. and Sober, E. 2004. Prediction versus Accommodation and the Risk of Overfitting. *British Journal for the Philosophy of Science*, 55, 1-34.

Holst, R.J., Brink, W., Veltman, D.J. and Goudriaan, A.E. 2010. Why gamblers fail to win: a review of cognitive and neuroimaging findings in pathological gambling. *Neurosci Biobehav Rev*, 34, 87-107.

Holt, C.A. and Laury, S.K. 2002. Risk Aversion and Incentive Effects. *American Economic Review*, 92(5), 1644-1655.

Kable, J.W. and Glimcher, P.W. 2009. The neurobiology of decision: consensus and controversy. *Neuron*, 63, 733-745.

Kitcher, P. 1981. Explanatory Unification. *Philosophy of Science*, 48 (4), 507-531.

Kitcher, P. 1989. Explanatory Unification and the Causal Structure of the World. *In Kitcher, P. and Salmon, W. Eds. Minnesota Studies in the Philosophy of Science, 13*, Univ. Minnesota Press, 410-505.

Knutson, B., Rick, G.S., Wimmer, E., Prelec, D. and Loewenstein, G. 2007. Neural Predictors of Purchases. *Neuron*, 53, 147-56.

Kosfeld, M., Heinrichs, M., Zak, P.J., Fischbacher, U. and Fehr, E. 2005. Oxytocin Increases Trust in Humans. *Nature*, 435, 673-6.

Kuorikoski, J., Lehtinen, A. and Marchionni, C. 2010. Economic modelling as robustness analysis. *British Journal for the Philosophy of Science*, 61(3), 541-567.

Kuorikoski, J., Lehtinen, A. and Marchionni, C. 2012. Robustness analysis disclaimer: please read the manual before use! *Biology and Philosophy*, 27, 891-902.

Kuorikoski, J. and Ylikoski, P. 2010. Explanatory relevance across disciplinary boundaries: the case of neuroeconomics. *Journal of Economic Methodology*, 17 (2), 219-228.

Lehtinen, A. and Kuorikoski, J. 2007. Unrealistic assumptions in rational choice theory. *Philosophy of the Social Sciences*, 37, 115-38.

Levy, D.J. and Glimcher, P.W. 2012. The root of all value: A neural common currency for choice. *Current Opinion in Neurobiology*, 22, 1-12.

Loewenstein, G., Rick, S. and Cohen, J.D. 2008. Neuroeconomics. *Annual Review of Psychology*, 59, 647-672.

Logothetis, N.K. 2008. What we can do and what we cannot do with fMRI. *Nature*, 453, 869-878.

Loomes, G. and Sugden, R. 1982. Regret Theory: an Alternative Theory of Rational Choice under Uncertainty. *The Economic Journal*, 92 (368), 805-824.

Machamer P., Darden, L. and Craver, C. 2000. Thinking about mechanisms. *Philosophy of Science*, 67, 1-25.

Machina, M. 1987. Choice Under Uncertainty: Problems Solved and Unsolved. *The Journal of Economic Perspectives*, 1(1), 121-154.

Mäki, U. 2009. MISSing the World. Models as Isolations and Credible Surrogate Systems. *Erkenntnis*, 70, 29–43.

Mäki, U. 2010. When economics meets neuroscience: hype and hope. *Journal of Economic Methodology*, 17(2), 107-117.

Matthewson, J. and Weisberg, M. 2009. The structure of tradeoffs in model building. *Synthese*, 170, 169-190.

McCabe, K. 2003. Neuroeconomics. *Encyclopedia of Cognitive Science*, 294-8. Macmillan, New York.

McCabe, K. 2008. Neuroeconomics and the Economic Sciences. *Economics and Philosophy*, 24, 345-368.

Mitchell, S.D. 2002. Integrative Pluralism. *Biology and Philosophy*, 17, 55-70.

Odenbaugh, J. and Alexandrova, A. 2011. Buyer beware: robustness analyses in economics and biology. *Biology and Philosophy*, 26, 757-771.

Padoa-Schioppa, C. 2011. Neurobiology of Economic Choice: A Good-Based Model. *Annual Review of Neuroscience*, 34, 333-359.

Padoa-Schioppa, C. and Assad, J.A. 2008. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nat. Neurosci.*, 11, 95–102.

Pessiglione, M., Seymour, B., Flandin, G., Dolan, R.J. and Frith, C.D. 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*, 442, 1042–1045.

Potenza, M.N. 2008. Review. The neurobiology of pathological gambling and drug addiction: an overview and new findings. *Philos Trans R Soc Lond B Biol Sci*, 363, 3181-3189.

Robinson, T. and Berridge, K. 2003. Addiction. *Annual Reviews of Psychology*, 54, 25-53.

Ross, D. 2008. Two Styles of Neuroeconomics. *Economics and Philosophy*, 24, 473-483.

Ross, D. 2009. Integrating the Dynamics of Multiscale Economic Agency. In *The Oxford Handbook of Philosophy of Economics*, Ed. Kincaid, H. and Ross, D. Oxford University Press, 245-279.

Ross, D., Sharp, C., Vuchinich, R. and Spurrett, D. 2008. *Midbrain Mutiny: The Picoeconomics and Neuroeconomics of Disordered Gambling*. MIT Press.

Rubinstein, A. 2008. Comments on Neuroeconomics. *Economics and Philosophy*, 24, 485-494.

Rustichini, A. 2009. Is There a Method of Neuroeconomics? *American Economic Journal: Microeconomics*, 1(2), 48-59.

Sanfey, A.G., Rilling, J.K., Aronson, J.A., Nystrom, L.E. and Cohen, J.D. 2003. The neural basis of economic decision-making in the ultimatum game. *Science*, 300 (5626), 1755-8.

Schotter, A. 2008. What's so informative about choice? In *The Foundations of Positive and Normative Economics: A Handbook*, Caplin, A. and Schotter, A. Eds. Oxford University Press, Ch.3, 70-94.

Simon, H.A. 1955. A Behavioral Model of Rational Choice. *The Quarterly Journal of Economics*, 69(1), 99-118.

Starmer, C. 2000. Developments in nonexpected utility theory: the hunt for a descriptive theory of choice under risk. *Journal of Economic Literature*, 38, 332-382.

Sugden, R. 2000. Credible worlds: the status of theoretical models in economics. *Journal of Economic Methodology*, 7(1), 1-31.

Sugden, R. 2009. Credible Worlds, Capacities and Mechanisms. *Erkenntnis*, 70, 3–27.

Vromen, J. 2010. Where economics and neuroscience might meet. *Journal of Economic Methodology*, 17(2), 171-183.

Weisberg, M. 2007. Three Kinds of Idealization. *The Journal of Philosophy*, 104 (12), 639-659.

Wilcox, N. 2008. Against Simplicity and Cognitive Individualism. *Economics and Philosophy*, 24, 523-532.

Woodward, J. 1997. Explanation, Invariance, and Intervention. *Philosophy of Science*, 64, S26-S41.

Woodward, J. 2003. *Making things happen*. New York: Oxford University Press.

Woodward, J. 2006. Some varieties of robustness. *Journal of Economic Methodology*, 13(2), 219–240.

Wout, M., Kahn, R., Sanfey, A. and Aleman, A. 2005. Repetitive transcranial magnetic stimulation over the right dorsolateral prefrontal cortex affects strategic decision-making. *Neuroreport*, 16 (16), 1849-52.

Zak, P. 2004. Neuroeconomics. In *Philosophical Transactions of the Royal Society Biology*, 359, 1737-48.

Zak, P. 2008. The brains behind economics. *Journal of Economic Methodology*, 15 (3), 301-312.