DOI:
10.1007/s10539-016-9546-y

Document Version
Peer reviewed version

On the Neural Enrichment of Economic Models:

Recasting the Challenge

Abstract

In a recent article in this Journal, Fumagalli (2011) argues that economists are provisionally justified in resisting prominent calls to integrate neural variables into economic models of choice. In other articles, various authors engage with Fumagalli's argument and try to substantiate three often-made claims concerning neuroeconomic modelling. First, the benefits derivable from neurally informing some economic models of choice do not involve significant tractability costs. Second, neuroeconomic modelling is best understood within Marr's three-level of analysis framework for information-processing systems. And third, neural findings enable choice modellers to confirm the causal relevance of variables posited by competing economic models, identify causally relevant variables overlooked by existing models, and explain observed behavioural variability better than standard economic models. In this paper, I critically examine these three claims and respond to the related criticisms of Fumagalli's argument. Moreover, I qualify and extend Fumagalli's account of how trade-offs between distinct modelling desiderata hamper neuroeconomists' attempts to improve economic models of choice. I then draw on influential neuroeconomic studies to argue that even the putatively best available neural findings fail to substantiate current calls for a neural enrichment of economic models.

Keywords: Scientific Modelling; Economic Choice; Neuro-biological Integration; Modelling Pluralism; Neuroeconomics.

Word count: 8350

1. Introduction

The proponents of neuroeconomics (henceforth, NE) frequently advocate integrating neural variables into economic models of choice (see e.g. Camerer, 2008, Glimcher, 2010, ch.12-15, and Loewenstein et al., 2008). In their view, this integration enables economists to discriminate between competing models of choice and build more predictive and explanatory models. In a recent article in this Journal, Fumagalli (2011) argues that these calls for a neural enrichment of economic models (henceforth, NEEM) face pressing pragmatic and epistemic challenges. In particular, he articulates a 'refined argument from tractability' to demonstrate that economists are provisionally justified in resisting neuroeconomists' (henceforth, NEs) calls for NEEM. Fumagalli's argument builds on the trade-offs between the modelling desiderata valued by NEs (e.g. fit with the available neural findings) and by other economists (e.g. tractability) respectively. The idea is that these desiderata make dissimilar demands on modellers and that these trade-offs, in turn, hamper NEs' attempts to improve economic models of choice.[1]

In other articles, various authors engage with Fumagalli's argument and try to substantiate three often-made claims concerning NE modelling. First, "for at least some economic model of choice behaviour, the benefits derivable from neurally informing [this] economic model do not involve special tractability costs" (Colombo, 2015, 713; see also Camerer, 2008, and Loewenstein et al., 2008). Second, NE modelling "is best understood within Marr's three-level of analysis framework" for information-processing systems (Colombo, 2015, 713; see also Glimcher, 2003, and Kable and Glimcher, 2009). And third, neural findings enable choice modellers to confirm the causal relevance of variables posited by competing economic models, identify causally relevant variables overlooked by existing models, and explain observed behavioural variability better than standard economic models (Colombo, 2015, 715-7; see also Fehr and Rangel, 2011, Quartz, 2008, and Rustichini, 2009). Let us call these three claims *tractability thesis*, *Marr thesis* and *relevance thesis* respectively.

In this paper, I critically examine these three theses and respond to the related criticisms of Fumagalli's (2011) argument. Moreover, I qualify and extend Fumagalli's account of how trade-offs between distinct modelling desiderata hamper NEs' attempts to improve economic models. I then argue that in spite of recent integrative advances at the interface between NE's parent disciplines, the proponents of NE have failed to substantiate their calls for NEEM. The contents are organized as follows. In *Section 2*, I examine the *tractability thesis* and argue that it fails to address the challenge that Fumagalli's refined argument from tractability poses to the proponents of NEEM. In *Section 3*, I focus on the *Marr thesis* and argue that it stands in tension with important tenets of Marr's (1982) framework and is vulnerable to hitherto unaddressed objections. In *Section 4*, I draw on a useful analytical framework put forward by Bernheim (2009) and Dean (2013) to assess the relevance of NE findings

for the economic modelling of choice. In particular, I inspect influential NE studies of the neural substrates of risk-sensitive choice that have been claimed to inform economic modelling in all the three respects indicated by the *relevance thesis*. I then argue that even these selected NE contributions lack the evidential and explanatory relevance for economic modelling required to substantiate the proffered calls for NEEM. My critical appraisal does not exclude the possibility that neural findings can provide economists with an additional source of evidence to assess existing models (see e.g. Bernheim, 2009, and Kuorikoski and Marchionni, 2016) and develop novel models of choice (see e.g. Dean, 2013, and Krajbich and Dean, 2015). Still, it makes it pressing for the proponents of NEEM to put forward more convincing reasons and evidence to support their calls to incorporate neural variables into economic models of choice.[2]

Before proceeding, three preliminary caveats are in order. First, in this paper I focus on neural - rather than biological - variables since recent progress in neuroscience is especially relevant for assessing the proffered calls for NEEM. However, I shall speak of neural variables broadly to include neuro-biological variables into this category. Moreover, I shall refer in various places to the literature on modelling in neuro-biology, highlighting some parallel concerns regarding the integration of neural and biological variables into economic models (see e.g. Matthewson and Weisberg, 2009, on the existence and significance of trade-offs between distinct modelling desiderata; see also Boone and Piccinini, 2016, on the relevance of tractability considerations for economic and neuro-biological modelling). Second, in the NE literature various authors advocate claims that resemble the *tractability thesis*, the *Marr thesis* and the *relevance thesis*. Below I devote particular attention to Colombo (2015) because this article directly engages with Fumagalli's challenges and offers a commendably clear formulation of these theses. Still, as I illustrate in Sections 2-4, my remarks also bear on other prominent calls for such theses. In this perspective, Colombo's article can be seen as an interesting test case that nicely exemplifies the severity of the evidential and explanatory challenges faced by the proponents of NEEM.

Finally, my aim is not merely to provide a critical response to the proponents of NEEM, but also to advance the ongoing philosophical debate about the prospects of the interdisciplinary modelling of choice (see e.g. Kuorikoski and Ylikoski, 2010, Ross, 2008 and 2011, Vromen, 2007 and 2011, and Weisberg, 2007a and 2007b). More specifically, I aim to advance this debate in three respects of general interest to scientific modellers and philosophers of science, namely: identify some major divergences in the methodological presuppositions and the explanatory goals of choice modellers across distinct decision sciences; clarify these divergences' implications for the relevance of neural findings for the economic modelling of choice; and explicate the most pressing evidential and explanatory challenges that hinder interdisciplinary integration at the interface between economics, psychology and neuro-biology.

2. The Tractability Thesis and Modelling Trade-offs

The *tractability thesis* states that "for at least some economic model of choice, the benefits derivable from neurally informing [this] economic model do not involve special tractability costs", i.e. tractability costs that "are too high and specifically due to neurally informing [such] model" (Colombo, 2015, 713 and 716; see also Camerer, 2008). The notion of tractability has been characterized in several ways by choice modellers. For instance, Fumagalli (2011, 621) takes the number of variables appearing in a model as "an approximate indicator of its tractability" that provides "a convenient rule of thumb for comparing alternative modelling frameworks" (2011, 622; see also Kahneman, 2003). For his part, Colombo contends that complexity theory, which characterizes a model's tractability in terms of "the time a Turing machine needs for [finding a solution to] the model", provides a more 'reliable' indicator of tractability than the indicator used by Fumagalli (2015, 726). I am not persuaded by Colombo's contention regarding these two indicators of tractability (see e.g. Hindriks, 2006, on various differences between economists' and computer scientists' use of 'tractability'; see also Boone and Piccinini, 2016, on the contrast between solvability and tractability in neuro-computational modelling). In particular, I fail to see why exactly the modelling costs involved in NEEM should be assessed in terms of Colombo's complexity theory indicator of tractability as opposed to other indicators of tractability. Still, below I gloss over this definitional concern since my main challenge to the tractability thesis holds irrespective of whether one adopts Fumagalli's or Colombo's indicator of tractability. Let me expand on this challenge.[3]

In articulating his argument, Fumagalli (2011, 627) acknowledges that the workings of the human neural architecture can be occasionally modelled without using many neural variables (see e.g. Schultz et al., 1997, on the dopaminergic underpinnings of basic reward valuation tasks). In his view, constructing descriptively accurate and neurally informed economic models of choice does not always require NEs to represent the neural substrates of choice down to their tiniest details (e.g. size and number of ion channels, number and strength of synaptic connections). However, when it comes to modelling the wide range of decision problems targeted by economists (e.g. choosing between heterogeneous multi-attribute financial portfolios, choosing between alternative careers), "accurately representing the neural substrates of choice […] would *typically* impose significant tractability costs" (Fumagalli, 2011, 628, italics added). Fumagalli bolsters this typicality claim with two often-made observations. First, several neural areas that contribute to decision making "activate in a wide range of decision contexts" (ibid., 628; see also Bernheim, 2009, and Weiskopf, 2016). And second, "even the execution of simple experimental tasks typically engages several areas, with additional variability resulting when it comes to solving decision problems" (Fumagalli, 2011, 628; see also Muldoon and Bassett, 2016, and Vromen, 2010a).

According to Colombo (2015, 722), neither of these observations licenses

the claim that providing descriptively accurate representations of the neural substrates of choice requires one to use many neural variables. In particular, he contends that "there are already descriptively accurate, tractable economic models of choice that incorporate neural [variables]" (ibid., 715). Suppose, for the sake of argument, that this contention is correct. Even so, it remains unclear how such contention is supposed to bear against Fumagalli's typicality claim that, when it comes to modelling the wide range of decision problems targeted by economists, "accurately representing the neural substrates of choice […] would *typically* impose significant tractability costs" (2011, 628, italics added). To undermine this typicality claim, a proponent of NEEM would have to demonstrate not just that NEs have developed *some* descriptively accurate and tractable models of the neural substrates of choice, but also that NEs can provide such models for a *wide range* of decision problems targeted by economists. Regrettably, the proponents of NEEM have hitherto failed to meet this justificatory requirement.[4]

To be sure, Colombo puts forward a detailed case study to vindicate NEs' calls for NEEM with regard to a class of models putatively used in several economic contexts. Moreover, he challenges Fumagalli and "whoever believes that, currently, a neural enrichment of economic models [would typically impose significant tractability costs] on economists [to] provide actual case studies in support of [this belief]" (2015, 716). I shall provide a detailed assessment of Colombo's case study in *Section 4*. For now, I note that it is hard to discern how stringently Colombo's demand for case studies should be interpreted (e.g. how many case studies would be required to vindicate Fumagalli's typicality claim?). Furthermore, it remains difficult to see why exactly support for Fumagalli's typicality claim should consist of particular case studies as opposed to empirical or methodological considerations (e.g. tractability considerations) that purport to hold across case studies. At any rate, focusing on a few selected case studies does not *per se* enable the proponents of NEEM to address the challenge posed by Fumagalli's refined argument from tractability. For this challenge concerns the wide range of decision problems targeted by economists rather than a few selected case studies.[5]

3. The Marr Thesis and the Goals of NE

According to the *Marr thesis*, NE modelling "is best understood within Marr's three-level of analysis framework" for information-processing systems (Colombo, 2015, 713). Marr (1982, 20-29) posits the following tripartition between the computational, algorithmic and physical implementation levels of analysis of information-processing systems (e.g. the visual system). Computational analyses aim to specify what information-processing problems are faced by the examined system in terms of specific input-output mappings. Algorithmic analyses explicate how inputs and outputs are represented, and what algorithms putatively transform the system's inputs into the observed outputs. Finally, physical

implementation analyses aim to identify what physical substrates implement the system's representations and algorithmic operations.[6]

Various leading NEs (e.g. Glimcher, 2003, ch.6-8, and Kable and Glimcher, 2009) take Marr's framework to provide a plausible way to understand NEs' contributions to the interdisciplinary modelling of choice. The idea is that economic models at the computational level target "the kinds of functions that can be optimized by a given type of behaviour", NE models at the algorithmic level identify "the processes by which that behaviour can be carried out", and neuro-biological models at the physical implementation level target "the neural structures and activities that implement a given algorithm" (Colombo, 2015, 729). Suppose, for the sake of argument, that NE modelling *can* be plausibly understood within Marr's framework along these lines (see e.g. Dean, 2013). This falls short of licensing the stronger 'Marr thesis' that NE modelling "is *best* understood" within such framework (Colombo, 2015, 713). Moreover, as I argue below, the 'Marr thesis' stands in profound tension with Marr's own emphasis on the formal independence of distinct levels of analysis. This, in turn, casts doubt on the alleged superiority of a Marrian conceptualization of NE modelling over other conceptualizations.

According to Marr (1982, ch.1; see also Marr et al., 1979), performing informative analyses of information-processing systems requires one to identify precisely what information-processing problems are faced by such systems. As documented by subsequent studies, severe difficulties plague attempts to identify such problems (see e.g. Anderson, 2015, and Warren, 2012, on Marr's attempts to identify the information-processing problems faced by the visual system). This identification task is even more difficult when it comes to extending Marr's analysis of basic visual perception tasks to other cognitive activities (see e.g. Shagrir and Bechtel, 2015, on episodic memory) and the decision problems targeted by economists (see e.g. Harrison, 2008, on choices involving multiple computational goals simultaneously). These identification problems, combined with the fact that many different algorithms can solve specific computational problems and many different neural substrates can implement the investigated algorithms (see e.g. Bernheim, 2009; see also Marr, 1982, ch.1), make it hard to see why exactly NE modelling would be 'best understood' within Marr's framework.

To be fair, NEs may alleviate some of these identification and underdetermination problems by triangulating findings from different levels of analysis (see e.g. Kuorikoski and Marchionni, 2016, and Quartz, 2008, on how neural and algorithmic findings may constrain computational theories). Yet, the point remains that according to Marr "the three levels are only rather loosely related", so that several phenomena "may be [modelled and] explained at only one or two of them" (1982, 25; see also Kaplan, 2011). Indeed, Marr adamantly insists that "from an information-processing point of view [it is] the level of computational theory, which is critically important" (ibid., 27) and that "computational theory [is] independent of the algorithm or

implementation levels" (ibid., 337; see also Craver and Alexandrova, 2008). These remarks do not imply that decision problems are best modelled independently at the computational, algorithmic and physical implementation levels (see e.g. Boone and Piccinini, 2016). Still, they cast doubt on the thesis that NE interdisciplinary modelling is 'best understood' within Marr's framework. In particular, they challenge the proponents of this thesis to specify which tenets of Marr's framework substantiate such a thesis.

These considerations, in turn, have critical implications for the alleged superiority of a Marrian conceptualization of NE research over different conceptualizations. To see this, consider Colombo's claim that "once it is clear that *the relevant* notion of a level used in [NE modelling] is Marr's one", then it becomes evident that "asking whether 'human choice behaviour is more conveniently modelled at the neural - rather than some other - level' is a *misunderstanding* of [NE] methodology" (2015, 729, italics added). This claim invites the following two-fold rejoinder. First, it is highly doubtful that *one single* notion of level is 'the relevant' notion of level in NE modelling *in general*. For different notions of level may be relevant in different modelling contexts, depending on factors such as modellers' pragmatic purposes and the epistemic interests of their target audiences (see e.g. Craver, 2005, and Mäki, 2009 and 2010). And second, the hitherto proffered claims that 'the relevant' notion of level in NE modelling is Marr's one fall short of implying that asking whether human choice behaviour is more conveniently modelled at the neural - rather than some other - level is a *misunderstanding* of NE. For one may consistently use the term 'level' to indicate *both* distinct scientific disciplines *and* different levels of analysis in Marr's sense. In fact, several leading NEs use the term 'level' in both senses when presenting and discussing their findings. For instance, Glimcher advocates a Marrian conceptualization of NE modelling, yet repeatedly speaks of 'levels' with reference to scientific disciplines (see e.g. 2003, ch.6-8, and 2010, ch.2). In particular, he holds that "the goal of [NE] is to produce a single unified model of human decision making that spans the economic, psychological, and neuroscientific levels of analysis" (2010, 4).[7]

These pluralistic remarks about NE *modelling* also hold for prominent characterizations of the *goals* of NE. To see this, consider Colombo's contention that "all characterizations [agree that] the goal of [NE] is an algorithmic description of the human mechanism for choice" (2015, 728). This contention fits prominent characterizations of the goals of NE (see e.g. Glimcher, 2010, ch.8). However, it seemingly overlooks that not all entrenched approaches to NE research aim to provide an algorithmic description of the human mechanism for choice (see e.g. Ross, 2008, on so-called behavioural economics in the scanner; see also Montague, 2007, on two 'natural neuroeconomics', which respectively investigate algorithms and the efficient operation of neural tissue). Furthermore, several authors privilege other levels of analysis over the algorithmic one in their characterizations of NE modelling (see e.g. Rangel et al., 2008, 545, for the claim that "it is the computations that are central to uniting

[distinct] levels"). These observations do not exclude the possibility that various NE studies may be regarded as attempts to identify "indirect ways of ultimately getting at [algorithms]" (Colombo, personal correspondence). Still, they cast doubt on the claim that providing an algorithmic description of the human mechanism for choice is 'the' (or even 'the main') goal of NE. To put it differently, reiterating that "all characterizations" of NE agree that "the goal of [NE] is an algorithmic description of the human mechanism for choice" provides a simplistic portrayal of NE that fails to fit the diverse goals of NE.

4. The Relevance Thesis and the Economic Modelling of Choice

According to the *relevance thesis*, neural findings enable choice modellers to confirm the causal relevance of variables posited by competing economic models, identify causally relevant variables overlooked by existing models, and explain observed behavioural variability better than standard economic models (Colombo, 2015, 715-7; see also Fehr and Rangel, 2011, Quartz, 2008, and Rustichini, 2009, for similar claims). In this section, I assess whether neural findings inform economic modelling in these three respects, focusing on some of the putatively best available NE studies. More specifically, I consider an influential NE study Colombo takes to exemplify "how results from cognitive neuroscience can be used to inform economic models of choice" (2015, 722), namely Niv et al.'s (2012) work on the neural substrates of risk-sensitive choice. This study constitutes an ideal test case for assessing NEs' contributions, since it purportedly "completes a full circle starting from computational theory, through the psychology [and] the neural basis [of choice] and back again to influence and inform the computational theory" (Niv et al., 2012, 561).

4.1 Experimental Setting and Main Findings

Niv et al. scan sixteen human subjects with fMRI to examine the neural substrates of risk-sensitive choice. These subjects earn monetary rewards by choosing between different stimuli. Some of these stimuli are associated with fixed payoffs, whereas the others are associated with a 50% probability of receiving specified payoffs (both the fixed and the risky stimuli are associated with payoffs in the 0¢-40¢ range). Subjects have to learn the payoffs of different stimuli through the experiment. The main aim of Niv et al. is to ascertain whether "neural representations associated with reinforcement learning […] are sensitive only to mean payoffs [or whether] some of the effects of risk on choice may be realized through such learning" (2012, 560).

Niv et al. (2012) first document a close correspondence between a hypothetical, model-derived reward prediction error and fMRI blood oxygenation level-dependent (BOLD) signals in the nucleus accumbens. They then extract from this prediction error signal the learned values of cues that predict rewards of equal mean but different variance, and show that "a close […] coupling exists between the fluctuations of [this neural

measure of] experience-based evaluations of risky options [and observed variations] in behavioral risk aversion" (2012, 551). Finally, Niv et al. assess the relative fit with behavioural and neural data of three temporal difference learning models of risk-sensitive choice. First, we find a standard *TD model*, which focuses on the mean outcome for each stimulus, and does not explicitly take outcome variance into account. Second, there is what Niv et al. call *utility model*, a variant of the standard TD model that incorporates nonlinear subjective utilities for the outcomes that are integrated into TD learning. And third, we have a *risk-sensitive TD (RSTD) model*, where nonlinearity is associated with the learning process itself rather than the evaluation of outcomes.

More formally, in the standard TD model a reward prediction error

$$\delta(t) = r(t) + V(t) - V(t - 1)$$

is computed at each of two consecutive time steps $t_{stimulus}$ and $t_{outcome} = t_{stimulus} + 1$, where V(t) is the predicted value of a given stimulus at time t, and r(t) is the reward at time t. The reward prediction error at $t_{outcome}$ is used to update V(C), the value of the chosen stimulus, according to $V_{new}(C) = V_{old}(C) + \eta \cdot \delta(t_{outcome})$, with η being a learning rate parameter. What Niv et al. call utility model has the same update rule as the standard TD model, but has a different reward prediction error, namely:

$$\delta(t) = U[r(t)] + V(t) - V(t - 1)$$

where U[r(t)] is the subjective utility of the reward at time t, with U(0) = 0, U(20) = 20, and U(40) = $a \cdot 20$. In the RSTD-model, reward prediction errors are defined as in the standard TD model, but there are separate update rules for positive and negative reward prediction errors, namely:

$$V_{new}(C) = V_{old}(C) + \eta^{+} \cdot \delta(t_{outcome}), \text{ if } \delta(t_{outcome}) > 0$$

$$V_{new}(C) = V_{old}(C) + \eta^{-} \cdot \delta(t_{outcome}), \text{ if } \delta(t_{outcome}) < 0$$

such that if $\eta^{+} < \eta^{-}$, the effect of negative prediction errors on learned values is larger than that of positive prediction errors, leading to risk aversion, and vice versa if $\eta^{+} > \eta^{-}$.

All these three models assume that subjects use past experience to estimate the expected payoffs associated with different stimuli and, given a choice, pick between stimuli based on these stimuli's expected values. However, those models yield dissimilar predictions about subjects' behaviour in situations of risk and make different claims about the neural substrates of risk-sensitive choice. More specifically, risk aversion "arises implicitly" in the standard TD model because "risky stimuli are, by definition, associated with outcomes that are larger or smaller than their means", so that "learned predictive values […] fluctuate above and below the mean according to the specific sequence of past experienced rewards" (Niv et al., 2012, 555). In the so-called utility model, instead, risk-sensitive preferences "emerge because the two options that have objectively been designated as having the same mean payoff do not lead to an equal subjective mean reward" (ibid., 556). Again differently, in the

RSTD model risk sensitivity arises because positive and negative prediction errors have asymmetric effects on learning. In the words of Niv et al., "if negative errors have the effect of decreasing [the updated value of the chosen stimulus] more than positive errors increase it, then the learned value will be lower than the mean nominal outcome, leading to risk aversion. Conversely, higher learning rates for positive compared to negative prediction errors will [lead to] overestimation of the values of risky options and thus to risk seeking" (ibid., 556).

Niv et al. (2012, 556-7) compare the 'posterior likelihoods' of each subject's choice data according to each of the three models, and find that the RSTD model fits the behavioural data better than the other two models almost for every subject. They then compare the three models' predictions to neural measurements of the learned values of the stimuli and, again, find that the RSTD model fits the neural data better than the other two models. In their view, these findings indicate that risk sensitivity "is indeed present in prediction error signaling in the nucleus accumbens, with a direct correlation between the risk-averse or risk-seeking choices of [experimental] subjects and the neural prediction error signals" (2012, 561). These findings, in turn, allegedly suggest that risk sensitivity "as in RSTD learning, should be imported into computational models of human choice" (ibid., 561).

4.2 Evaluation

Niv et al.'s findings have been claimed to inform the economic modelling of choice in *all* the three respects indicated by the relevance thesis. More specifically, these findings putatively: (1) "confirm (or disconfirm) the causal relevance of latent, subjective variables or processes posited by competing [economic] models of choice" (Colombo, 2015, 715; see also Niv et al., 2012, 551); (2) "point to causally relevant variables or processes overlooked by existing models" (Colombo, 2015, 715; see also Niv et al., 2012, 551); and (3) "explain [observed] behavioral variability both computationally and neurally, by contrasting three possible explanations for risk-sensitive choice" (Niv et al., 2012, 555; see also Colombo, 2015, 717). Below I draw on a useful analytical framework put forward by Bernheim (2009) and Dean (2013) to examine these three purported contributions in sequence. I shall argue that Niv et al.'s study improves over former NE works on risk-sensitive choice (e.g. Hsu et al., 2009) by clearly specifying in what respects neural findings supposedly inform the economic modelling of choice. Still, even this selected study lacks the evidential and explanatory relevance for economic modelling required to substantiate the proffered calls for NEEM.

4.2.1 Confirming the causal relevance of latent variables

According to Colombo, "the main benefit" of Niv et al.'s study for economic modellers consists in "a kind of *independent test* of competing models of risk-sensitive choice" (2015, 717, italics added). The analytical framework put forward by Bernheim (2009) and Dean (2013) provides a

helpful basis for explicating and assessing this alleged contribution. This framework conceptualizes the economic modelling of choice as an attempt to identify how an individual's choices are causally influenced by a set of conditions that the individual regards as fixed (or at least predetermined). The framework takes economic modellers to assume that some function *f*: $y = f(x, \omega)$ maps a set of observed explanatory variables x (e.g. prices) and a set of unobserved environmental variables ω (e.g. neural activations) on the set of choices y. Standard economic models make no explicit assumptions about the neuro-biological substrates of choice and treat the unobserved environmental variables as noise. For their part, NEs regard standard economic models as a reduced form for the underlying neuro-biological processes. More formally, NEs assume that neural activations z depend on both observed and unobserved variables through a function *Z*: $z = Z(x, \omega)$, with the individual's choices resulting from a function *Y*: $y = Y(z, x, \omega)$. NEs then build on observed neural activations and choices to disclose both function *Z* and function *Y*, thereby helping economic modellers to identify the mapping $f(x, \omega) = Y(Z(x, \omega), x, \omega)$.

Suppose, for the sake of argument, that Niv et al.'s findings enable choice modellers to test competing *neuro-biological* models by confirming the causal relevance of specific neuro-biological variables (e.g. activation patterns of particular neural areas) posited by such models. This does not *per se* imply that these findings enable modellers to test competing *economic* - as opposed to neuro-biological - models. To be sure, what Niv et al. call utility model does incorporate a functional form frequently used by economists, namely nonlinear subjective utilities for outcomes. This, however, by no means implies that this model is plausibly regarded as an economic - as opposed to a neuro-biological or some other kind of - model. For such functional form can be employed to model a wide variety of systems besides those targeted by economists (see e.g. Kacelnik and Bateson, 1996, on animal foraging, and Ross, 2014a, ch.5, on macro-scale physical objects). In this respect, both Colombo and Niv et al. seem to take Niv et al.'s findings to inform economic modelling only because they employ the term 'economic' in an implausibly broad sense.[8]

This point is not definitional hair-splitting, but has critical implications for the alleged relevance of Niv et al.'s findings for the economic modelling of choice. To illustrate this, let me contrast two of the claims Colombo makes concerning the putative relevance of Niv et al.'s findings for economic modellers. In some passages, he claims that these findings suggest that "*algorithmic* models of risk-sensitive […] choice should apply a non-linear transformation to prediction errors" (2015, 732, italics added). In other places, he puts forward the much stronger contention that "*any* descriptively accurate model of human choice should take account of risk sensitivity, as per [Niv et al.'s] risk-sensitive TD-model" (ibid., 722, italics added). The former claim highlights the potential import of Niv et al.'s findings for NE models of the algorithmic underpinnings of choice. The latter contention appears to overstate those findings' evidential and explanatory relevance for economic modellers in at least three major respects.

First, economists have put forward since the 1980s more predictive and sophisticated functional forms than the functional form in Niv et al.'s utility model (see e.g. Quiggin, 1982, on rank-dependent expected utility models, and Yaari, 1987, on dual expected utility models). Moreover, Niv et al. do not give any reason to think that the functional form they favour robustly outperforms economists' best available functional forms. This, in turn, significantly constrains the evidential and explanatory relevance of Niv et al.'s results for economic modellers. That is to say, the proponents of NEEM should examine more predictive and sophisticated functional forms than the functional form in Niv et al.'s utility model if they are to substantiate their calls to incorporate neural variables into economic models (see e.g. Wilcox, 2008 and 2011, on the so-called new structural econometrics, which combines distinct functional forms with stochastic models to estimate structural risk parameters).

Second, severe limitations affect current attempts to rely on temporal difference reinforcement learning models to demonstrate that one functional form provides the uniquely best fit with observed choices across the real-life choice settings targeted by economists. To be sure, temporal difference reinforcement learning models may guide the formulation of empirical hypotheses about behavioral relationships, possibly leading economists to examine novel specifications in choice settings (like the one targeted by Niv et al.) involving the accumulation of experience (see e.g. Caplin and Dean, 2008a). Yet, when it comes to the real-life choice settings targeted by economists, studies focused solely on temporal difference reinforcement learning models lack the potential to demonstrate that one functional form robustly fits observed choices better than other functional forms. For several learning processes besides temporal difference reinforcement learning often influence choices in such choice settings (see e.g. Bernheim, 2009, and Krajbich and Dean, 2015).[9]

And third, inferences about risk attitudes have been shown to vary remarkably depending on whether the latent data generating process is viewed through the lens of one, two or more models (see e.g. Harrison and Rutström, 2008, for detailed illustrations). Moreover, economists have developed various statistical tools to estimate the probability that multiple latent data generating processes generate the observed choices (see e.g. Harrison and Rutström, 2009). Unfortunately, Niv et al.'s comparative estimates of individual models' fit with data do not consider the possibility that multiple latent data generating processes generate the observed choices. This, in turn, constrains both the informativeness of Niv et al.'s comparative estimates and the reliability of Colombo's claim that "any descriptively accurate model of human choice should take account of risk sensitivity, as per [Niv et al.'s] risk-sensitive TD-model" (2015, 722).

A proponent of NEEM may object that Niv et al.'s findings are evidentially and explanatorily relevant for economic modelling on the alleged ground that they target phenomena *modelled by economists*, namely observed choices. Economics has often been characterized as a

'science of choice' during its history (see e.g. Robbins, [1932] 1945, and Weber, [1904] 1949, for famous characterizations of economics as a science that studies human choice behaviour in presence of scarce means having alternative uses). Yet, the mere fact that some findings target phenomena modelled by economists falls short of implying that these findings are evidentially and explanatorily relevant for the economic modelling of choice. To see this, suppose that a particle physicist provided a set of statistically significant correlations between an individual's purchase decisions and the micro-physical bodily movements the individual implements in performing such decisions. The correlations provided by the particle physicist target a phenomenon modelled by economists. This, however, by no means implies that these correlations are evidentially and explanatorily relevant for the economic modelling of choice. In fact, those correlations seem largely orthogonal to economists' evidential and explanatory concerns (see Bernheim, 2009, sec.1, for similar remarks about NEs' correlations between observed choices and endogenous brain activity).

A proponent of NEEM may further object that NEs could in principle *redefine* entrenched criteria for demarcating the set of economic models so as to imply that *any* finding targeting observed choices is *ipso facto* evidentially and explanatorily relevant for economic modelling. This redefinition, however, would trivialize the putative significance of the claim that neural findings are evidentially and explanatorily relevant for economic modelling. For under such redefinition, findings from all sorts of disciplines - ranging from social anthropology to particle physics - would be evidentially and explanatorily relevant for economic modelling. This does not vindicate isolationist characterizations of economics' domain that aprioristically exclude latent neuro-biological variables from such domain (see e.g. Gul and Pesendorfer, 2008). Still, it challenges NEs to provide more nuanced criteria for demarcating the set of economic models and assessing the relevance of neural findings for the economic modelling of choice. As Ross puts it, "imagine a model that features a consumption […] variable on its left-hand side and nothing but neurological variables on its right-hand side. This would strike almost all economists […] as not being an economic model" (2011, 218).

4.2.2 Pointing to new relevant variables

According to Colombo, Niv et al.'s findings "point to *causally relevant* variables or processes overlooked by existing models" (2015, 715, italics added). In his view, these findings help choice modellers to ascertain whether traditional reinforcement learning models (e.g. Sutton and Barto, 1998) "may be extended to take account of [risk sensitivity]", or whether "there might be two separate mechanisms of risk-sensitive choice" (2015, 718), one by which subjects learn the mean values of different options - as per traditional reinforcement learning models - and the other by which subjects learn what variance is associated with reward outcomes. Suppose, for the sake of argument, that Niv et al.'s findings point to causally relevant *neuro-biological* variables or processes overlooked by existing

models of risk-sensitive choice. This result may be highly informative to modellers interested in the neuro-biological substrates of decisions. Yet, it is unclear how exactly such result bears on the merits of *economic* - as opposed to neuro-biological - models of choice. For economic models do not in fact rest on specific presuppositions concerning what neuro-biological variables or processes underlie risk-sensitive choice (see e.g. Fumagalli, 2016b, and Vromen, 2010b). Paraphrasing Dean, economists recognize that neuro-biological processes mediate the influence of the variables they target (e.g. prices) on choices, yet neuro-biological processes "represent neither the environmental [variables] nor the behavioral outcomes [targeted by economists]" (2013, 167).

To be fair, NEs' findings could in principle prompt economists to build their models on neuro-biological presuppositions (see e.g. Bernheim, 2009) and include neuro-biological variables into their models (see e.g. Caplin and Dean, 2015). Nonetheless, severe limitations constrain the suitability of Niv et al.'s findings to prompt such changes in economic modelling. To give one example, Niv et al. investigate the neural substrates of *stimulus-bound* risk-sensitive choices within *time spans* (e.g. up to a few seconds) that are much shorter than those usually targeted by economists. Moreover, they provide no evidence that the algorithms they hypothesize can be plausibly taken to determine risk-sensitive choice when it comes to the *long-term non-stimulus-bound* choices targeted by economists (e.g. think of choices between multi-stage courses of action involving abstract rewards). These limitations affect not just Niv et al.'s study, but also other influential NE studies of the neural substrates of choice (see e.g. Fumagalli, 2016c, and Ross, 2009, for several examples).[10]

More generally, the point remains that identifying and measuring variables that causally influence individuals' choices does not *per se* imply that economists should include these variables into their models. My point is not just that individuals' choices are causally influenced by a panoply of variables that few economic modellers would include into their models (e.g. the amount of solar radiation to which individuals are exposed, how many red blood cells are present in their circulatory systems). Rather, my main concern is that activations in the neural areas targeted by leading NE models of learning are immediate precursors to choices (see e.g. Fumagalli, 2013) and that observing these activations does not significantly improve economists' measurement of the causal relationship between standard exogenous environmental variables and choices (see e.g. Krajbich and Dean, 2015). Furthermore, markets and other information-processing scaffolding innovations can demonstrably lead individuals who drastically differ in their neuro-biological makeup to instantiate observationally equivalent choice patterns (see Ross, 2014b, for illustrations). These considerations do not preclude NEs from identifying increasingly fine-grained associations between observed choice patterns and specific neuro-biological variables (see e.g. Li et al. 2013, and Van den Bos et al., 2014, on some neural measures of structural and functional connectivity). Still, they challenge the proponents of NEEM to support

their calls to incorporate neural variables with more convincing reasons and evidence than the claim that neural findings "point to causally relevant variables or processes overlooked by existing models" (Colombo, 2015, 715). In the words of Bernheim, "the burden of proof is squarely on [the NEs'] shoulders [to provide] a novel economic model derived originally from [NE] research that improves our measurement of the causal relationship between a standard exogenous environmental [variable and] a standard economic choice" (2009, 26).

4.2.3 Explaining economic choices

Niv et al. "set out to *explain* [observed] behavioral variability both computationally and neurally, by contrasting three possible *explanations* for risk-sensitive choice" (2012, 555, italics added). According to Colombo, Niv et al. provide "independent evidence about […] the explanatory power of competing economic models of choice" (2015, 722). In his view, their findings inform economic modelling by "discriminat[ing] the explanatory power of competing economic models of choice" (ibid., 722). Let us assess the cogency of these claims.

Niv et al. do not explicate in what sense they employ the term 'explanation' in their study. This makes it difficult to interpret and assess their claims concerning the explanatory relevance of their findings. For his part, Colombo specifies that for his purposes "to say that a utility function 'explains' some behavioural regularity […] means at least that the function fits data relevant to that behavioural regularity […] reasonably well" (2015, 717). This specification improves over Niv et al.'s vague use of the term 'explanation', but is not very precise either (e.g. what does the expression 'reasonably well' mean? What measure of goodness of fit does this expression presuppose?). In particular, Colombo points to a notion that is more akin to descriptive fit with data than to the accounts of explanation entrenched in NE's parent disciplines (see e.g. Fumagalli, 2014, on the unificationist, mechanistic and interventionist accounts). Moreover, neither Colombo nor Niv et al. give detailed reasons to think that Niv et al.'s findings are explanatory for economists under these entrenched accounts. This, in turn, is problematic since those accounts concur that data fit alone is not sufficient for explanation (see e.g. Craver, 2006, Kaplan, 2011, and Ylikoski and Kuorikoski, 2010).

A proponent of NEEM may object that Niv et al.'s findings are explanatory for *NEs* under various accounts of explanation entrenched in NE's parent disciplines (see e.g. Glimcher, 2010, on the mechanistic account). Suppose that this objection is correct. This would fall short of implying that Niv et al.'s findings are explanatory for *other economists*. For *in primis*, NEs and other economists frequently endorse dissimilar accounts of explanation, which are grounded on different criteria of explanatory relevance (see e.g. Kuorikoski and Ylikoski, 2010, for a detailed review). And second, even those NEs and economists who endorse the same accounts of explanation may give such accounts rather different interpretations. By way of illustration, NEs and other economists

may endorse a mechanistic account of explanation, yet ground such account on distinct notions of mechanism (see e.g. Kuorikoski, 2009, on mechanisms as 'componential causal systems' versus mechanisms as 'abstract forms of interaction') and make dissimilar assumptions as to whether or not providing mechanistic explanations of choices requires one to draw on neural findings (see e.g. Fumagalli, 2014). Due to these differences, the mere fact that NEs are gaining more detailed understanding of the neuro-biological mechanisms underlying choices by no means implies that economists should explain choices in terms of such mechanisms (see e.g. Dietrich and List, 2016). To put it differently, showing that some findings are explanatory for NEs falls short of indicating that such findings are explanatory for other economists.

A proponent of NEEM may further object that "in economics as well as in philosophy of economics, there is little agreement on what precisely constitutes an adequate explanation of [choices]" (Colombo, 2015, 717). In particular, she may point out that some leading economists use 'explanation' as a synonym for descriptive fit with data (see e.g. Friedman, 1953, 8-9). These considerations are not without merit. Yet, once this descriptivist characterization of explanation is adopted, then the claim that Niv et al.'s findings are explanatory for economists loses much of its bite. For several findings that count as explanatory under such descriptivist characterization seem neither informative nor relevant for economists. To see this, consider Niv et al.'s finding that the RSTD model fits the behavioural data better than the basic TD model and what they call 'utility model'. This finding counts as explanatory under the descriptivist characterization of explanation adopted by Colombo, but seems neither informative nor relevant for economists. For the basic TD model "cannot generate" risk-seeking behavior in the experimental setting examined by Niv et al. (2012, 555). Moreover, risk sensitivity was formerly shown to be incompatible with nonlinear utility in the domain of small payoffs targeted by Niv et al. (see e.g. Rabin and Thaler, 2001).

In fact, as Niv et al. (2012, 560) cursorily acknowledge, their findings do not even shed light on whether the asymmetric effect that positive and negative prediction errors putatively have on subjects' predictions is fixed or varies depending on factors such as subjects' amount of training, payoff variations beyond the 0¢-40¢ range, and the degree of risk involved in the examined task. These limitations, in turn, severely constrain the evidential and explanatory relevance of Niv et al.'s findings for the modelling of the real-life decision problems targeted by economists. For all the mentioned factors have been shown to influence experimentally elicited individuals' risk attitudes (see e.g. Harrison and List, 2004, on subjects' amount of training, Andersen et al., 2008, on payoff variations, and Harrison and Rutström, 2008, sec.3, on the degree of risk involved in the examined task; see also Harrison et al., 2015, for an updated review). That is to say, the proponents of NEEM have to meet much more stringent standards of experimental rigor and sophistication if they are to substantiate their calls to incorporate neural variables into economic models of choice.

5. Conclusion

In recent years, promising integrative advances have been made at the interface between economics, psychology and neuro-biology. In light of these advances, it would be implausible to reiterate that most NEs "are in the dark" about how their research "will reshape economics" (Rubinstein, 2008, 486-7) and are "a far cry from [providing] integrated neural and economic models" (Gul and Pesendorfer, 2009, 44). Still, the proffered calls for NEEM face pressing and hitherto unaddressed evidential and explanatory challenges. In this paper, I articulated and defended three such challenges, which respectively target the modelling trade-offs involved in NE models (*tractability thesis*), entrenched conceptualizations of NE modelling (*Marr thesis*), and the relevance of neural findings for the economic modelling of choice (*relevance thesis*). My three challenges do not exclude the possibility that neural findings can provide economists with an additional source of evidence to assess existing models and develop novel models of choice. Still, they make it pressing for the proponents of NEEM to put forward more convincing reasons and evidence to support their calls to incorporate neural variables into economic models of choice.

REFERENCES

Andersen, S., Harrison, G.W., Lau M. and Rutström, E. 2008. Eliciting risk and time preferences. *Econometrica*, 76, 583-618.

Anderson, B.L. 2015. Can Computational Goals Inform Theories of Vision? *Topics in Cognitive Science*, 7, 274-286.

Bechtel, W. and Shagrir, O. 2015. The Non-Redundant Contributions of Marr's Three Levels of Analysis for Explaining Information-Processing Mechanisms. *Topics in Cognitive Science*, 7, 312-322.

Bernheim, B.D. 2009. On the potential of neuroeconomics: a critical (but hopeful) appraisal. *American Economic Journal: Microeconomics*, 1, 1-41.

Bernheim, B.D. and Rangel, A. 2007. Toward Choice-Theoretic Foundations for Behavioral Welfare Economics. *American Economic Review*, 97, 464-470.

Bernheim, B.D. and Rangel, A. 2008. Choice-Theoretic Foundations for Behavioral Welfare Economics. In *The Foundations of Positive and Normative Economics: A Handbook*, ed. Andrew Caplin and Andrew Schotter, 155-192. Oxford University Press.

Bernheim, B.D. and Rangel, A. 2009. Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics. *Quarterly Journal of Economics*, 124, 51-104.

Boone, W. and Piccinini, G. 2016. The Cognitive Neuroscience Revolution. *Synthese*, 193, 1509-1534.

Camerer, C.F. 2008. The Case for Mindful Economics. In Caplin, A. and Schotter, A. Ed. *The Foundations of Positive and Normative Economics. A Handbook*, 43-69. Oxford University Press.

Caplin, A. and Dean, M. 2008a. Axiomatic Neuroeconomics. In *Neuroeconomics: Decision Making and the Brain*, Ch.3. Glimcher, P., Camerer, C., Fehr, E. and Poldrack, R. Eds. Academic Press.

Caplin, A. and Dean, M. 2008b. Dopamine, reward prediction error, and economics. *Quarterly Journal of Economics*, 123, 663-701.

Caplin, A. and Dean, M. 2015. Enhanced Choice Experiments. In *The Method of Modern Experimental Economics*, Ch.4. Frechette, G and Schotter, A. Eds. Oxford University Press.

Caplin, A., Dean, M., Glimcher, P.W. and Rutledge, R.B. 2010. Measuring beliefs and rewards: A neuroeconomic approach. *Quarterly Journal of Economics*, 125, 923-960.

Colombo, M. 2015. For a Few Neurons More… On Tractability and Neurally Informed Economic Models. *British Journal for the Philosophy of Science*, 66, 713-736.

Craver, C. 2005. Beyond Reduction: Mechanisms, Multifield Integration and the Unity of Neuroscience. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 36, 373-395.

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

Craver C. and Alexandrova, A. 2008. No revolution necessary: neural mechanisms for economics. *Economics and Philosophy*, 24, 381-406.

Dean, M. 2013. What Can Neuroeconomics Tell Us About Economic Decisions (and Vice Versa)? In *Comparative Decision Making*, Ch.7. Crowley, P. and Zentall, T. Eds. Oxford University Press.

Dietrich, F. and List, C. 2016. Mentalism versus behaviourism in economics: a philosophy-of-science perspective. *Economics and Philosophy*, 32, 249-281.

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

Friedman, M. 1953. The Methodology of Positive Economics. In *Essays in Positive Economics*. Chicago: Chicago University Press.

Fumagalli, R. 2011. On the neural enrichment of economic models: tractability, trade-offs and multiple levels of description. *Biology and Philosophy*, 26, 617-635.

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

Fumagalli, R. 2014. Neural Findings and Economic Models: Why Brains have Limited Relevance for Economics. *Philosophy of the Social Sciences*, 44, 606-629.

Fumagalli, R. 2016a. Decision Sciences and the New Case for Paternalism: Three Welfare-Related Justificatory Challenges. *Social Choice and Welfare*. Published online 03/06/16.

Fumagalli, R. 2016b. Five Theses on Neuroeconomics. *Journal of Economic Methodology*, 23, 77-96.

Fumagalli, R. 2016c. Choice Models and Realistic Ontologies: Three Challenges to Neuro-Psychological Modellers. *European Journal for Philosophy of Science*, 6, 145-164.

Glimcher, P.W. 2003. *Decisions, uncertainty, and the brain: the science of neuroeconomics*. Bradford.

Glimcher, P.W. 2010. *Foundations of neuroeconomic analysis*. Oxford University Press.

Gul, F. and Pesendorfer, W. 2008. The Case for Mindless Economics. In A. Caplin and A. Schotter, Ed. *The Foundations of Positive and Normative Economics: A Handbook*, 3-42. Oxford University Press.

Gul, F. and Pesendorfer, W. 2009. A comment on Bernheim's appraisal of neuroeconomics. *American Economic Journal: Microeconomics*, 1, 42-47.

Harrison, G.W. 2008. Neuroeconomics: a critical reconsideration. *Economics and Philosophy*, 24, 303-344.

Harrison, G.W., Lau, M. and Rutström, E. 2015. Theory, Experimental Design and Econometrics Are Complementary. In G. Frechette and A. Schotter Eds. *Handbook of Experimental Economic Methodology*, 296-338. Oxford University Press.

Harrison, G.W. and List, J.A. 2004. Field experiments. *Journal of Economic Literature*, 42, 1013-1059.

Harrison, G.W. and Ross, D. 2010. The methodologies of neuroeconomics. *Journal of Economic Methodology*, 17, 185-196.

Harrison, G.W. and Rutström, E. 2008. Risk Aversion in the Laboratory. In Cox, J.C. and Harrison, G.W. (Ed.). *Risk Aversion in Experiments.* Greenwich, CT: JAI Press, 41-196.

Harrison, G.W. and Rutström, E. 2009. Expected utility theory and prospect theory: one wedding and a decent funeral. *Experimental Economics*, 12, 133-158.

Hindriks, F.A. 2006. Tractability assumptions and the Musgrave-Mäki typology. *Journal of Economic Methodology*, 13, 401-423.

Hsu, M., Krajbich, I., Zhao, C. and Camerer, C.F. 2009. Neural Response to Reward Anticipation under Risk Is Nonlinear in Probabilities. *Journal of Neuroscience*, 29, 2231-37.

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

Kacelnik, A. and Bateson, M. 1996. Risky Theories - The Effects of Variance on Foraging Decisions. *American Zoologist*, 36, 402-434.

Kahneman, D. 2003. A psychological perspective on economics. *American Economic Review*, 93, 162-168.

Kaplan, D.M. 2011. Explanation and description in computational neuroscience. *Synthese*, 183, 339-373.

Kitcher, P. 1988. Marr's Computational Theory of Vision. *Philosophy of Science*, 55, 1-24.

Krajbich, I. and Dean, M. 2015. How can Neuroscience Inform Economics? *Current Opinion in Behavioral Sciences*, 4, 51-57.

Kuorikoski, J. 2009. Two Concepts of Mechanism: Componential Causal System and Abstract Form of Interaction. *International Studies in the Philosophy of Science*, 23, 143-160.

Kuorikoski, J. and Marchionni, C. 2016. Evidential Diversity and the Triangulation of Phenomena. *Philosophy of Science*, 83, 227-247.

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

Li, N., Ma, N., Liu, Y., He, X., Sun, D., Fu, X., Zhang, X., Han, S., Zhang, D. 2013. Resting-state functional connectivity predicts impulsivity in economic decision-making. *Journal of Neuroscience*, 33, 4886-4895.

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

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, 107-117.

Marr, D. 1982. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. New York: W.H. Freeman & Company.

Marr, D., Ullman, S. and Poggio, T. 1979. Bandpass Channels, Zero-Crossings and Early Visual Information Processing. *Journal of the Optical Society of America*, 69, 914-916.

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

Montague, P.R. 2007. Neuroeconomics: a view from neuroscience. *Functional Neurology*, 22, 219-234.

Muldoon, S. and Bassett, D. 2016. Network and Multilayer Network Approaches to Understanding Human Brain Dynamics. *Philosophy of Science*, In Press.

Niv, Y., Edlund, J., Dayan, P. and O'Doherty, J. 2012. Neural Prediction Errors Reveal a Risk-Sensitive Reinforcement-Learning Process in the Human Brain. *Journal of Neuroscience*, 32, 551-562.

Quartz, S.R. 2008. From Cognitive Science to Cognitive Neuroscience to Neuroeconomics. *Economics and Philosophy*, 24, 459-471.

Quiggin, J. 1982. A theory of anticipated utility. *Journal of Economic Behavior and Organization*, 3, 323-343.

Rabin, M. and Thaler, R.H. 2001. Risk Aversion. *Journal of Economic Perspectives*, 15, 219-232.

Rangel, A., Camerer, C.F. and Montague, P.R. 2008. A framework for studying the neurobiology of value-based decision making. *Nature Reviews Neuroscience*, 9, 545-556.

Robbins, L. [1932] 1945. *An Essay on the Nature and Significance of Economic Science*, 2nd Rev. Ed. London, Macmillan.

Ross, D. 2008. Two styles of neuroeconomics. *Economics and Philosophy*, 24, 473-483.

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

Ross, D. 2011. Estranged parents and a schizophrenic child: choice in economics, psychology and neuroeconomics *Journal of Economic Methodology*, 18, 217-231.

Ross, D. 2014a. *Philosophy of Economics*. Palgrave Macmillan.

Ross, D. 2014b. Psychological versus Economic Models of Bounded Rationality. *Journal of Economic Methodology*, 2, 411-427.

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, 48-59.

Schultz, W., Dayan, P. and Montague, P.R. 1997. A neural substrate of prediction and reward. *Science*, 275, 1593-1599.

Shagrir, O. 2010. Marr on computational-level theories. *Philosophy of Science*, 77, 477-500.

Shagrir, O. and Bechtel, W. 2015. Marr's Computational Level and Delineating Phenomena. In Kaplan, D.M. Ed. *Integrating psychology and neuroscience*. Oxford University Press.

Sutton, R.S. and Barto, A.G. 1998. *Reinforcement learning: an introduction*. Cambridge, MA: MIT Press.

Van den Bos, W., Rodriguez, C., Schweitzer, J. and McClure, S. 2014. Connectivity strength of dissociable striatal tracts predict individual differences in temporal discounting. *Journal of Neuroscience*, 34, 10298-10310.

Vromen, J. 2007. Neuroeconomics as a natural extension of bioeconomics: The shifting scope of standard economic theory. *Journal of Bioeconomics*, 9, 145-167.

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

Vromen, J. 2010b. On the surprising finding that expected utility is literally computed in the brain, *Journal of Economic Methodology*, 17, 17-36.

Vromen, J. 2011. Neuroeconomics: Two camps gradually converging: What can economics gain from it? *International Review of Economics*, 58, 267-285.

Warren, W. 2012. Does this computational theory solve the right problem? Marr, Gibson, and the goal of vision. *Perception*, 41, 1053-1060.

Weber, M. 1904. Objectivity in Social Science and Social Policy. In *The Methodology of the Social Sciences*. 1949. Ed. and Transl. by Shils, E.A. and Finch, H.A. New York: Free Press.

Weisberg, M. 2007a. Three kinds of idealization. *Journal of Philosophy*, 104, 639-659.

Weisberg, M. 2007b. Who is a Modeler? *British Journal for Philosophy of Science*, 58, 207-233.

Weisberg, M., Okasha, S. and Mäki, U. 2011. Modeling in biology and economics. *Biology and Philosophy*, 26, 613-615.

Weiskopf, D. 2016. Integrative Modeling and the Role of Neural Constraints. *Philosophy of Science*, In Press.

Wilcox, N.T. 2008. Stochastic Models for Binary Discrete Choice Under Risk: a Critical Primer and Econometric Comparison. In J. C. Cox and G. W. Harrison, Eds. *Research in Experimental Economics*, 197-292. Bingley, UK: Emerald.

Wilcox, N.T. 2011. Stochastically More Risk Averse: a Contextual Theory of Stochastic Discrete Choice Under Risk. *Journal of Econometrics*, 162, 87-104.

Yaari, M.E. 1987. The Dual Theory of Choice under Risk. *Econometrica*, 55, 95-116.

Ylikoski, P. and Kuorikoski, J. 2010. Dissecting explanatory power. *Philosophical Studies*, 148, 201-219.

[1] Fumagalli's (2011) refined argument from tractability differs from former critiques of NE, which prevalently target purported methodological flaws in NEs' studies (see e.g. Harrison, 2008, and Harrison and Ross, 2010), putative limitations in the accuracy and reliability of NEs' findings (see e.g. Bernheim, 2009, and Rubinstein, 2008), and the alleged irrelevance of such findings for economic modellers (see e.g. Gul and Pesendorfer, 2008). I shall comment on these critiques and their interrelations in various places throughout the paper.

[2] NE findings might foster the development of new methods for measuring individuals' welfare and evaluating policies' welfare implications on the basis of neural activity. I gloss over these potential normative contributions since my

evaluation focuses on positive economic analyses. For a discussion of NE's potential contributions to normative economic analyses, see e.g. Bernheim and Rangel, 2007, 2008 and 2009, and Fumagalli, 2013 and 2016a.

[3] The term 'tractability' may be used to designate not just a property of *models*, but also a property of the activities of *modelling*, namely "the ease with which modellers can build [a] model or manipulate it" (Colombo, 2015, 727). In this paper, I focus on tractability as a property of models - rather than the activities of modelling - since the controversy about NEEM prevalently concerns the former notion (see e.g. Colombo, 2015, 727, for a similar remark).

[4] This justificatory requirement may be regarded as more or less demanding depending on how one interprets the expression 'wide range'. However, this interpretative concern has limited bearing on Fumagalli's challenge. For on most plausible interpretations of 'wide range', pointing to a few selected descriptively accurate and tractable NE models of the neural substrates of choice falls short of indicating that NEs can provide such models for a wide range of decision problems targeted by economists.

[5] This is not the only respect in which Colombo's defence of the tractability thesis seems to mischaracterize Fumagalli's argument. Two such mischaracterizations are relevant for appraising NEs' calls for NEEM. First, Colombo takes Fumagalli to infer that "since modelling choice behaviour at the neural level would involve too high modelling costs in comparison to models incorporating variables at some other level, economists should refrain from modelling choice behaviour at the neural level" (2015, 715). However, Fumagalli (2011, 627-631) repeatedly emphasizes that assessing NEs' calls for NEEM involves a comparative evaluation of *both* modelling costs *and* modelling benefits. Reconstructing his argument as if it concerned exclusively modelling costs oversimplifies both the structure and the implications of such argument (see Fumagalli, 2011, 627-631, for discussion). And second, *pace* Colombo, Fumagalli nowhere asserts that "providing a descriptively accurate and tractable model of choice *prevents* economists from incorporating variables at the neural level" (2015, 715, italics added). On the contrary, Fumagalli's argument is premised on the pluralistic assumption that economists "may fruitfully combine neural and other disciplines' insights in constructing particular models of choice" (2011, 633). In this respect, Colombo appears to miss the pluralistic spirit of Fumagalli's argument (see Weisberg et al., 2011, 613; see also *Section 3*).

[6] Different positions as to how exactly each level of analysis is to be conceptualized have been advocated (see e.g. Bechtel and Shagrir, 2015, Kitcher, 1988, and Shagrir, 2010). The remarks in the text are sufficiently detailed for the purpose of my evaluation.

[7] A proponent of NE might object that the contrastive character of the question whether 'human choice behaviour is more conveniently modelled at the neural - rather than some other - level' does not fit well NEs' insistence on combining findings from multiple disciplines. However, one may consistently acknowledge that NEs aim to combine findings from multiple disciplines, yet argue that NEs' attempts to implement such combination face pragmatic and epistemic challenges (see e.g. Fumagalli, 2011 and 2013). In this respect, no misunderstanding of NE methodology seems inherent in the question whether 'human choice behaviour is more conveniently modelled at the neural - rather than some other - level'.

[8] Niv et al. might rebut that what they call 'utility model' is plausibly regarded as an economic model on the alleged ground that such model "is the standard explanation for risk sensitivity from economics" (Niv et al., 2012, 555). I address this rebuttal in point *4.2.3* below.

[9] This does not exclude that one may capture nonparametrically the testable implications of choice models that contain latent variables. In fact, leading economists have already applied rigorous axiomatic approaches to test reward prediction error (RPE) models of the same class as those compared by Niv et al. For instance, Caplin and Dean (2008b) use an axiomatic approach to test RPE models with neuro-biological data and identify three axiomatic conditions that characterize the entire class of RPE models in a simple, nonparametric way. These axiomatic conditions yield a minimal requirement for RPE models in the sense that if neural activity is to satisfy any one of the class of RPE models, then such activity must also satisfy those axiomatic conditions (Caplin et al., 2010).

[10] A proponent of NEEM might conjecture that the "heterogeneity of risk attitudes at extended timescales might […] be partly explained by people's varying levels of short-timescale […] risk sensitivity" (Colombo, 2015, 726). This conjecture highlights the potential of neural findings to shed light on the neural substrates of intertemporal behavioural patterns. However, due to the limitations constraining the evidential and explanatory relevance of current neural findings for the modelling of the real-life decision problems targeted by economists (see point *4.2.3* below), such conjecture does not presently support the claim that economists should build their models of choice on neuro-biological presuppositions and include neuro-biological variables into their models.