Economics & Philosophy (2023), **39**, 351–372
doi:10.1017/S0266267122000189

ARTICLE

The utility of goods or actions? A neurophilosophical assessment of a recent neuroeconomic controversy

Enrico Petracca

School of Economics, Management and Statistics, University of Bologna, Italy
Email: enrico.petracca2@unibo.it

(Received 3 February 2021; revised 7 April 2022; accepted 13 May 2022; first published online 12 July 2022)

Abstract

The paper provides a neurophilosophical assessment of a controversy between two neuroeconomic models that compete to identify the putative object of neural utility: goods or actions. We raise two objections to the common view that sees the 'good-based' model prevailing over the 'action-based' model. First, we suggest extending neuroeconomic model discrimination to all of the models' neurophilosophical assumptions, showing that action-based assumptions are necessary to explain real-world value-based decisions. Second, we show that the good-based model's presumption of introducing a normative neural definition of economic choice would arbitrarily restrict the domain of economic choice and consequently of economics.

Keywords: Utility; economic choice; neuroeconomics; neurophilosophical assessment; model discrimination

1. Introduction

Among the promises of neuroeconomics, defining and measuring utility have been the most enticing but also the most controversial. To the enthusiasm of neuroeconomists, philosophers have replied with vigilant caution (Marchionni and Vromen 2014). Fumagalli (2013, 2019) makes a compelling case that the neural construct of utility has introduced further degrees of freedom rather than clarified and pinned down the constructs of decision and experienced utility. But neuroeconomics has not been without its share of internal controversies over utility. While most neuroeconomists agree that the brain does compute a notion of utility – Glimcher's (2011) 'hard' utility critically called 'true' utility by Fumagalli (2013) – a controversial and debated question has concerned *what* that utility refers to. In comparing two choice options, say one peach and one mango, the brain could encode and compute the subjective desirability either of the goods themselves (one peach; one mango) or of the associated actions (*grasping* one peach; *grasping* one mango). Although neural data consistent with both options were originally found, recordings from the orbitofrontal cortex (OFC) in the prefrontal cortex have in recent years pronounced in favour of the first scenario (Padoa-Schioppa and Conen 2017). The *good-based model* (hereafter, G-model) proposed by Camillo Padoa-Schioppa (Padoa-Schioppa 2011) is thus said to be prevailing over the *action-based model* (hereafter, A-model) originally proposed by Paul Glimcher (Glimcher *et al.* 2005). Moreover, OFC data show that when the brain encodes the utility of goods, or more generally of outcomes, neural signatures are consistent with the cardinal common measure of subjective desirability economists have been looking for since Francis Ysidro Edgeworth (Moscati 2018). This paper argues for reassessing the neuroeconomic controversy on (neuro)philosophical grounds. Not a few philosophical studies of neuroeconomics already adopt, although often implicitly, a neurophilosophical stance, that is, consider neuroscientific arguments as an integral (but of course not exclusive) part of the philosophical inquiry.[1] Here we suggest embracing the neurophilosophical stance more explicitly, as it would allow philosophers of economics to enter the debate and bring neuroeconomists to reconsider what they see otherwise as a piece of progress in their discipline (Camerer 2013). Why should philosophers do that? At least two reasons can be proposed. The first concerns the narrow grounds on which neuroeconomic model discrimination has been pursued, in the specific case of the controversy and in general. While many agree that comparing assumptions is crucial for neuroeconomic discrimination (e.g. Bernheim 2009; McMaster and Novarese 2016), not all relevant assumptions seem to have been considered. We argue that model discrimination in the controversy would yield different results if it included the entire set of what we call, following Herrmann-Pillath (2021), *neurophilosophical assumptions*, i.e. those assumptions that manifest a model's position on issues ranging from brain architecture to congruence with ecological or evolutionary viewpoints. Extending philosophers' gaze to neurophilosophical assumptions would permit them to be aware, for instance, of the G-model's incompatibility with the ecological view of decision-making, a view that, contrary to the G-model, considers action as a ubiquitous constituent of real-world decisions (Cisek and Pastor-Bernier 2014). Being aware of such incompatibility would immediately point to possible severe limits in the G-model's explanatory extent. Neurophilosophical assessment arguably makes discrimination more complex, as it requires delving into neuroscientific details and, even more challengingly, negotiating more extensively between discrimination criteria. Nevertheless, it provides knowledge of philosophically relevant factors – e.g. a model's explanatory extent – that would be unattainable otherwise. Another possible benefit of neurophilosophical assessment is, as we will see, the chance to find out choice models in the neuroscientific panorama that present more desirable neurophilosophical properties.

[1]The philosophical literature usually distinguishes a philosophy of neuroscience, which 'concerns foundational issues within the neurosciences', from neurophilosophy strictly speaking, which 'concerns application of neuroscientific concepts to traditional philosophical questions' (Bickle *et al.* 2019). We follow this distinction throughout the paper.

The second reason for pursuing a neurophilosophical reassessment of the controversy concerns the blunders economists might run into if they took the G-model at its face value. This risk originates in economists' often hasty embrace of neuroeconomics as a reductionist rather than integrative enterprise (e.g. Craver and Alexandrova 2008; Herrmann-Pillath 2016; Marchionni and Vromen 2020). In this context, the prospect of neurally vindicating cardinal utility might induce economists to take the G-model as a reason to disregard the role of action in value determination. This would have consequences for both the object and the boundary of economics. As for the object, following the G-model's lead would imply for economics the need to rethink itself from a science of *goal-directed* to one of *outcome-guided* behaviour (Rangel and Hare 2010; Padoa-Schioppa and Schoenbaum 2015). As for the boundary, neglecting action would rule out entire traditions that see the economic agent as an 'acting man' for whom '[t]he assignment of orders of rank through valuation is done only in acting and through acting' (von Mises 1998 [1949]: 120). However, the reductionist reading of the G-model has mostly been the work of neuroeconomists. Compared with what Ross (2008) has called 'neurocellular economics', the G-model pursues a far more radical form of *normative neural reductionism*. Besides interpreting neural data through economic models (what neurocellular economics is said to do), the G-model turns those data into normative requirements for a neural definition of *economic choice*. In other words, neural data are employed to tell economists the true and discriminatory nature of economic choice, and the verdict is: 'economic choice is *essentially* choice between goods *rather than* choice between actions' (Padoa-Schioppa and Assad 2006: 223, italics added). Reconstructing in detail how the G-model combines normatively different epistemological levels to obtain this (ontological) conclusion will help us identify some limits and risks inherent in this sort of definitional exercise. It can also bring material to address the broader philosophical question of whether normative neural definitions of utility are flawed not merely in fact but also in principle (Fumagalli 2013, 2019).

The paper is organized as follows. Section 2 reconstructs the G-model comparing it with the A-model and discusses on what grounds the G-model is said to be prevailing. Section 3 uncovers the non-evolutionary and non-ecological make-up of the G-model. This is propaedeutic to raise an *ecological* objection to it, which emphasizes the G-model's focus on choice situations that fail to represent value-based decisions in the real world (the latter called in neuroscientific jargon *naturalistic* decisions). This section also introduces an alternative neuroscientific model of choice – the 'distributed consensus' model – which considers both goods and actions as objects of naturalistic value-based choices. Then, section 4 discusses how the G-model, understood by its advocates as a model of 'outcome-guided' behaviour, would not match the traditional understanding of economic choice in terms of 'goal-directed' behaviour. Finally, section 5 critically discusses the way the G-model combines intuition with behavioural and neural data to normatively define economic choice, showing that such a definition ends up restricting too severely the domain of economic choice. In brief, we offer a neurophilosophical reassessment of the neuroeconomic controversy which covers the entire ground from the models' assumptions to their implications for economics. This is presented as an assessment format extendable to other cases of model discrimination in neuroeconomics.

2. Comparing the G-model with the A-model

2.1. From the A-model to the G-model

Historically, the A-model was the first neuroeconomic model of choice. In an early study, Platt and Glimcher (1999) showed that as monkeys chose rewards (drops of juice) through oculomotor movements (saccades), signals recorded from the lateral intraparietal (LIP) area in the parietal cortex correlated with the probability and magnitude of rewards. As area LIP is implicated in eye movement and presents topographic organization – i.e. the response of each neuron is selective to specific portions of the environment where stimuli are located – Platt and Glimcher naturally framed monkeys' choices as being about the direction of the saccade. More briefly, monkeys' choices were framed as choices between actions. In a subsequent study, Dorris and Glimcher (2004) qualified that signals in area LIP were about relative, not absolute, reward values. Many other brain areas in both monkeys and humans, among which dorsolateral prefrontal cortex (dlPFC), anterior cingulate cortex (ACC), dorsal premotor area (PMd) and supplementary motor area (SMA) (for a more exhaustive list, see Padoa-Schioppa 2011: 348), have been shown to encode reward-value signals dependently on the sensory and motor contingencies of choice tasks. In other words, reward-value encoding in these areas would be *non-abstract* insofar as signals depend on features such as the spatial location of the items or the actions required to obtain them.

In another study shortly thereafter, Padoa-Schioppa and Assad (2006) showed that in a different area of monkeys' prefrontal cortex, OFC, reward-value encoding is instead *abstract*, that is, independent of sensorimotor contingencies. The study recorded signals from single OFC neurons while monkeys were presented with variable combinations of juice types and amounts. Three different types of cells were identified: 'offer value' neurons encoding the value of each offered amount of juice, 'chosen value' neurons encoding the value of monkeys' choices independently of the juice type, and 'chosen juice' neurons identifying the type of chosen juice irrespective of the offered amount. These recordings were consistent with a two-stage choice model in which the value of each option is first encoded separately (offer value) and then integrated into a common currency (chosen value). In contrast to reward-value signals in other areas, OFC neurons were specifically found to encode values irrespective of the spatial location and action required to obtain the rewards.[2] For this reason, neural encoding seemed to refer more naturally to goods themselves rather than to actions. The more, unlike previous recordings from OFC (Tremblay and Schultz 1999) and sensorimotor areas (Dorris and Glimcher 2004), encoding was this time found to be about absolute, not relative, reward values. In a series of experiments continuing up to these days, Padoa-Schioppa and colleagues have shown that OFC encoding is range-adapting, as abstract reward-value

[2]To be precise, there was some choice-related sensorimotor encoding in OFC, but it amounted to less than 5% of the total number of neurons (see Padoa-Schioppa and Assad 2006: 223).

representation is flexible to different value ranges (Padoa-Schioppa 2009), and that encoded values are invariant to menu changes, that is, they do not vary when other options vary (Padoa-Schioppa and Assad 2007). This last point is important as it provides the neural underpinning of transitive choices (one pillar of the G-model's normative definition of economic choice, see section 5). In addition to neurophysiological studies, the G-model is also backed up by evidence of OFC lesions that severely disrupt choice behaviour (while lesions in sensorimotor areas do not impair choice to the same extent),[3] and by neurocomputational models proving the sufficiency of OFC encoding for value-based decisions (for further evidence, see, Padoa-Schioppa 2011; Padoa-Schioppa and Conen 2017).

2.2. Open questions about the G-model

Based on the evidence above, the G-model can be assessed from two different points of view. The first line of assessment concerns whether and how well the model explains the phenomena it originally intended to explain. Explanation (or lack thereof) concerns in this case not only the model's capacity to fit experimental data, but also whether good fits alone are sufficient to support the model's claims (and if not why). The second line of assessment goes instead beyond the target phenomena and concerns, for instance, possible extensions to other classes of phenomena, comparison with other models, and the model's implications for economics. As the latter line of inquiry will be pursued later, this section focuses on some remaining open questions – of both experimental and conceptual nature – on how well the G-model explains *simultaneous binary value-based free choices*.[4]

The first concern comes from studies that do find traces of sensorimotor encoding of rewards in OFC (e.g. Abe and Lee 2011), as they would invalidate the very idea of abstract value encoding. A usual response to this concern is that the studies in question identify goods exclusively in terms of their spatial location, so that the sensorimotor traces would likely refer to the goods themselves rather than to spatial features as such (see Padoa-Schioppa and Cai 2011). However, further evidence that OFC encoding might use non-good-based reference frames (Blanchard *et al.* 2015) has induced supporters of the G-model to acknowledge 'the possibility that reference frames are flexible' (Padoa-Schioppa and Conen 2017: 740). If confirmed, the existence of alternative frames for reward-value encoding in OFC would hit the core of the G-model.

Another concern comes from evidence that value signals are generated distributedly in the brain and do not always converge in OFC for integration. In neuroeconomics, goods are defined by those features that contribute to their value, such as quantity or payoff, probability, and costs in terms of effort and

[3]In addition to OFC, another brain area the disruption of which causes dramatic impairment of value-based decision-making is the amygdala (see Padoa-Schioppa and Conen 2017).

[4]Values are understood in the G-model as subjective valuations of rewards in choice tasks. For other constructs of value in neuroeconomics, see O'Doherty (2014). Choice tasks are said to be free when subjects are not instructed about choices and values.

time. Several studies show that these variables are encoded in different areas of the frontal cortex. While OFC has been found to specifically encode payoffs, and the lateral prefrontal cortex (lPFC) probability, the anterior cingulate cortex (ACC) has been found to encode all these variables together as an area of integration (Kennerley *et al.* 2009). When action costs vary, and therefore physical action becomes a differential determinant of choice, values signals generated in other regions have been shown to be integrated in ACC rather than OFC. This gives a sense of how the way action costs are encoded and computed is arguably the most controversial and open question in the G-model (see Padoa-Schioppa 2011: 342; Padoa-Schioppa and Conen 2017: 748; see also Rushworth *et al.* 2012). Recently, Cai and Padoa-Schioppa (2019) have shown that also after varying action costs by changing the amplitude of the saccade required to make the decision, choices are still taken in good space (i.e. in a good-based reference frame) rather than in action space. However, they acknowledge that this finding is dependent on the possibility of experimentally disentangling choices from actions, which is not generally feasible in real-world environments (Cai and Padoa-Schioppa 2019: 9; see also below). All in all, it is still unclear whether it is ACC (and other value-encoding areas) that send sensorimotor signals to OFC to be integrated in good space, or, conversely, it is OFC that sends abstract signals to ACC to be integrated in action space. In particular, it is unclear in which circumstance either case applies. On a more radical tone, Hunt and Hayden (2017) argue that the neural process of value representation is so massively distributed in the brain that the quest for single areas in charge of it would pose a problem of principle (see Hayden and Niv 2021).

Another issue is that not all value-based decisions require OFC encoding. For instance, OFC is not required when choices are between large and small rewards (Schoenbaum *et al.* 2011: 90). Nor does OFC seem to be involved when value-based decisions are taken on the basis of training and are therefore habitual, or when values stem from general affective states (Schoenbaum *et al.* 2011: 90). Moreover, there are decisions about goods that are made on the basis of values that do not concern subjective desirability (e.g. moral values) and as such do not involve OFC encoding (see Padoa-Schioppa and Schoenbaum 2015: 17). This points to the urgent conceptual need of distinguishing between different constructs of value in neuroeconomics (see O'Doherty 2014).

Most of the evidence presented so far involves single-neuron recordings from monkeys' brains. This leads to the last point of concern. Looking for anatomical and functional correspondences between neurophysiological studies on monkeys and imaging studies on humans, we find out that such correspondences are rather rough (Wallis 2012).[5] Human imaging studies locate in the ventromedial prefrontal cortex (vmPFC) – a region more medial than OFC – the area encoding subjective values (e.g. Clithero and Rangel 2014). Although OFC and vmPFC are proximal and in most studies are treated as the same area, they belong to different brain networks and scarcely communicate with each other

[5]For a similar comparative concern, see Glimcher (2011: 358 fn. 3). Notably, while primate OFC encodes signals abstractly, this is not the case with OFC in rats, which only presents sensorimotor encoding (Wallis 2012).

(Padoa-Schioppa 2011).[6] In particular, vmPFC is scarcely connected to motor areas and so unlikely capable of receiving sensorimotor signals necessary to compute action costs. This means that further comparative evidence is required before considering the human vmPFC a region homologous to OFC in monkeys (see Pearson *et al.* 2014).

These concerns notwithstanding, advocates of the G-model continue to accumulate evidence on its ability to explain simultaneous binary value-based free choices. In a recent study, the model's validity has even been extended to sequential choices (Ballesta and Padoa-Schioppa 2019). This has led the main contender, Glimcher, to accept the gist of the G-model (see Grattan and Glimcher 2014), although he seems still unpersuaded that decisions take place uniquely in good space. On the one hand, he envisages common mechanisms guiding decisions about goods and actions (Kable and Glimcher 2009), while, on the other hand, he hypothesizes a hybrid process in which choices switch from good space to action space in the course of the decision (Glimcher 2011; see e.g. Chen and Stuphorn 2015). The persistence of model uncertainty even for simple classes of choice has induced the contending parties to find an agreement on the kind of evidence required for model discrimination (Platt and Padoa-Schioppa 2009; Padoa-Schioppa 2011). An urgent objective is to ascertain whether what follows value encoding – that is, the decision itself – takes place either in good or in action space (Platt and Padoa-Schioppa 2009: 458). But Padoa-Schioppa has identified three specific empirical requirements that any plausible alternative to the G-model should satisfy: '(a) Neural activity must be genuinely motor, (b) neural activity must be modulated by subjective value, and (c) neural activity must not be downstream of the decision' (Padoa-Schioppa 2011: 349). Reviewing possible pieces of evidence challenging the G-model (some of which are discussed in this section), Padoa-Schioppa seems to find them inconclusive, at best (Padoa-Schioppa 2011; Padoa-Schioppa and Conen 2017).

3. Ecological issues with the G-model

As already noticed, any informed assessment of the G-model would be cognizant that the primary phenomenon it aims to explain is simultaneous binary value-based free choices.[7] Two important considerations arise at this point. First, even restricting our assessment to how well the G-model explains this class of choices, substantial differences may still underlie formally similar choices. Even limitedly to paradigmatic good-based choices, the choices made in a restaurant, we might not want to choose between a bottle of Nebbiolo and one of Negramaro just by looking at the carte. We might instead want to grasp the bottles, read their tags, look at them against the light, or even ask for a sample. The choice process may turn out to be fundamentally interactive, considerably relying on action before taking the decision. Such choices would formally remain

[6]Pelletier and Fellows (2019) call ventromedial frontal lobe (VMF) the area encompassing both vmPFC and OFC.

[7]Although G-model studies are restricted to primary reward values, it seems safe to say that the model aspires to extend to wider concepts of value.

simultaneous, binary, value-based, free choices between goods, but the underlying neural signatures might change dramatically due to the performed actions (e.g. Kubanek and Snyder 2015). The second consideration concerns the frequency of the menu type of choice in real life. Decisions about goods or outcomes do not typically involve menus, as choice options often need to undergo a preliminary process of search and identification (Fumagalli 2020). Moreover, most everyday decisions do not involve goods or outcomes at all. Kable and Glimcher (2009: 741) distinguish between 'economics-style tasks' from other tasks in the ecology of real-life decisions in which physical action and time pressure play a decisive role. All in all, these remarks suggest that there might be an issue with what Cisek and Pastor-Bernier (2014) call *ecological validity* with the G-model.

Pursuing ecological validity in neuroeconomics could take different routes. One is to recognize that even within the same class of choice (e.g. simultaneous choices), real-world decisions depend on situational details that may render one choice dramatically different from the other. As even slight situational variations may bring about different neural signatures, we could go in the direction of providing an ad hoc model of choice for virtually each different choice situation. Another path toward ecological validity could instead be conceiving a general model of choice that considers simultaneous binary value-based free choices between goods as a special case. To be really general, such a model would also consider decisions between actions as a special case. So far, the neural model of choice that comes closest to ecological generality is the *distributed consensus model* (DCM) proposed by Paul Cisek (Cisek 2012). DCM sees value-based decisions as ongoing processes taking place in parallel in good and action spaces. In a choice, representations of action planning would run in parallel with representations of subjective desirability, continuously influencing or 'biasing' each other. In the biasing process, conflicts may emerge: one option might be more subjectively desirable but also more motor demanding than the other, leading to a decision stall. In that case, conflict is resolved through additional bias. Eventually, decisions require a 'consensus' between levels of neural representation. As DCM assumes two different (although communicating) neural levels, one for goods and one for actions, when decisions require no action they neurally collapse into pure good-based representation, while when no good is involved they neurally collapse into pure action-based representation. Presented this way, there seems to be no reason not to see DCM as a mere generalization of both the A-model and the G-model. However, as we will see in what follows, this is hardly the case.

Philosophers have emphasized how neurobiological models of choice suffer from underdetermination, as extant evidence is often compatible with different models (see Fumagalli 2013; McMaster and Novarese 2016).[8] Underdetermination is well exemplified by evidence on the interaction between OFC and ACC, an area encoding action costs. As discussed in section 2.2, the G-model hypothesizes that action costs are integrated abstractly (along with other value signals) in OFC following input from ACC, but DCM also hypothesizes an opposite encoding direction (see Kennerley and Walton 2011). How to solve this

[8]Technically, only a theory can be underdetermined by data. In this paper, model and theory are used interchangeably.

conundrum? Cisek (2012: 927) acknowledges that abstract encoding of action costs in OFC is a possibility. Moreover, he also acknowledges that the neural separation assumed by the G-model between decisions and actions, the latter understood as the mere motor implementation of decisions (Cai and Padoa-Schioppa 2014), is a possibility. Why, then, do not he and other neuroscientists of real-world, naturalistic decisions support such hypotheses? Cisek replies to this question with another question which hints at the real reach of the contraposition: 'is such separation desirable from an ecological perspective?' (Cisek 2012: 929).

3.1. Explicating neurophilosophical assumptions

The previous question suggests that discriminating between neural models of choice may not be, as is often held by practitioners of neuroeconomics (see e.g. Padoa-Schioppa 2008; Platt and Padoa-Schioppa 2009: 458–459), an entirely empirical matter. Alongside usual modelling assumptions (Bernheim 2009; McMaster and Novarese 2016), assumptions on the nature of brain and cognition play a crucial role in neuroeconomics. These assumptions contribute to determining what, following Herrmann-Pillath (2021), we call a model's neurophilosophical makeup. One assumption with neurophilosophical status in the G-model is certainly *modularity*. In their first study, Padoa-Schioppa and Assad (2006) emphasize that one advantage of their model is computational efficiency. As they say, '[f]rom a computational perspective, a modular design separating the mental operations of "choosing" and "moving" is more parsimonious' (Padoa-Schioppa and Assad 2006: 225). In another circumstance, Padoa-Schioppa adds that the 'action-based hypothesis violates a principle of modularity because the nervous system could certainly break down the complex operation [choosing and moving] into two separate and simpler operations, [choosing] and [moving]' (Padoa-Schioppa 2007: 247). Furthermore, it is held that 'evolution likely favored modular organizations' (Padoa-Schioppa and Conen 2017: 747).[9] Appealing to the principle of anatomical and functional modularity, the G-model adheres to a view of brain and cognition called 'cognitivism' that has also extensively informed assumptions in economic models (Petracca 2017).

The problem with assuming modularity is that examples of non-modular brain organization are ubiquitous. In area LIP, for instance, the same set of neurons has been found to play different functions, attention-related, choice-related and action-related over the course of the decision (see Cisek and Kalaska 2010). This evidence is compatible with an alternative principle of efficiency in brain organization called 'neural reuse': from the evolutionary viewpoint, it would be more efficient for the brain to recycle extant neural resources than create specialized areas each time a new function is required (Anderson 2014). According to this organization principle, A-model's evidence of value signals in sensorimotor areas is perfectly compatible with neural efficiency, although a different kind of efficiency than that assumed by the G-model. Issues with assuming modularity can also be found in its use as a heuristic principle of task design in neuroeconomics.

[9]Modularity has been construed differently by different authors. Padoa-Schioppa and Assad (2006) mention the views of Herbert Simon and Steven Pinker in support of their argument.

Neuroeconomic experiments routinely try to disentangle choice from motor implementation as neatly as possible, and then take the involvement of different neural areas in either phase as evidence of neural modularity. In other words, *behavioural modularity* is employed to elicit *neural modularity*. But identified this way, findings of neural modularity do not necessarily extend to naturalistic choices in which evaluation and motor phases cannot be behaviourally separated. As Cisek and Pastor-Bernier (2014: 3) claim, 'the temporal distinction between thinking about the choice and then implementing the response, so central to economic theory and laboratory experiments on decisions, simply does not apply to decisions made during interactive behaviour'. In interactive decisions, distinguishing between choices and motor responses is not only extremely complex, and to a certain degree not feasible (either behaviourally or neurally), but could not even be a meaningful goal to pursue. Nonetheless, they are the most common decision situations.

Different neurophilosophical assumptions are behind the way neuroeconomists cope with underdetermined theories and attempt to solve their controversies. Arguably the most persisting puzzle in the controversy we are examining concerns why sensorimotor areas exhibit value signals at all (Glimcher *et al.* 2005). It is instructive to see how different assumptions allow contenders to select different pieces of evidence in support of their views. As disruption of sensorimotor areas does not generally impair the ability to make decisions, the inference drawn by supporters of the G-model is that sensorimotor signals are not genuine value signals (see O'Doherty 2014). They could rather be, they say, attentional signals upstream of decision (Leathers and Olson 2012) or motor-planning signals downstream. On their side, supporters of DCM reply by giving prominence to different sorts of evidence. They emphasize, for instance, that the mechanism of action selection generally starts before a decision is made, so that motor planning cannot be said to be downstream of decisions (see Cisek 2012).

The discussion above suggests that presenting DCM as a mere generalization of the G-model can be misleading. Entirely different neurophilosophical assumptions divide these models, which manifest themselves in different views of brain architecture and experimental design, and lead to emphasize alternative sorts of evidence. If we extend the comparison to the A-model, we see that it does not take sides with either model but shows common aspects with both. While the centrality of action takes the A-model naturally closer to DCM, it instead ends up being closer to the G-model from the conceptual point of view (as they are both two-stage choice models, see Platt and Padoa-Schioppa 2009: 458) and for the kind of experimental approach (the use of non-naturalistic settings). Looking at this situation, discriminating between models would require modellers to take a position on empirical and methodological issues on which no consensus has emerged and is likely to emerge soon. For this reason, the claim 'that the controversy [between the G-model and the A-model] can be addressed based on neuronal measures' (Padoa-Schioppa 2008: 455) seems to be overoptimistic. Also the subscription of a possible compromise recently proposed by Padoa-Schioppa ('[i]f one accepts the DCM framework, it becomes interesting to ascertain under what circumstances exactly decisions are good based, action based, or distributed'; Padoa-Schioppa and Conen 2017: 748) would first require more basic agreement on neurophilosophical assumptions.

4. Choosing as a goal-directed activity

4.1. Learning, actions and goals

At the onset of the controversy, the contending parties agreed that the A-model was part of 'a more general psychological model of behavior', insofar as 'it builds more or less directly on theories of reinforcement learning' (Platt and Padoa-Schioppa 2009: 458). More recently, the success of the G-model has overshadowed this neurophilosophical point (see Herrmann-Pillath 2021). That OFC can encode values 'on-the-fly', that is, also without any prior experience of choice options (Padoa-Schioppa 2011), has led to maintaining that OFC value-encoding does not strictly require learning. Although not strictly required for OFC encoding, learning is however central to understanding the origin of values, and refocusing on learning could help assess neuroeconomic models of choice more comprehensively. For instance, focusing on learning could help understand the broader role of OFC in decision-making, as learning studies demonstrate that OFC also encodes the value of expected outcomes and therefore plays a central role in adaptive behaviour (see Schoenbaum *et al.* 2009). Yet, refocusing on learning would crucially require recognizing the centrality of action and instrumentality in value-based decisions, raising again the issue that values are better attributed to actions rather than to outcomes (Padoa-Schioppa and Schoenbaum 2015).[10] In theories of associative learning, the fundamental unit of analysis is *instrumental action*, a construct that makes behaviour-outcome associations possible (Balleine and Dickinson 1998), and which is therefore essential to what is called 'goal-directed behavior' (Dickinson and Balleine 1994). The present and next sections concentrate on action as a requirement of goal-directed behaviour and discuss whether the G-model framework is neurophilosophically compatible with goal-directedness. As this could not be the case, we suggest that economics, traditionally understood as a science of goal-directed behaviour, would require a neural model that specifically addresses goal-directedness.

Witnessing the centrality of learning at the origins of neuroeconomics, the first comprehensive framework for reward-based decisions built explicitly upon learning constructs. Rangel *et al.* (2008) distinguish between three value systems: a *Pavlovian*, a *habitual* and a *goal-directed* system.[11] As a basic and hard-wired system, the Pavlovian system is deemed to assign values to reactions to simple stimuli: it drives an automatic approach to rewarding stimuli and automatic aversion to aversive stimuli. The Pavlovian system is not generally considered to rely on learning, at least not as much as the second system, the habitual system, which creates associations between behaviours and outcomes through repeated actions.

[10]By focusing on repetition, the learning perspective can be considered to be criticizing the G-model for reasons opposite to those of DCM, which instead focuses on the inherent contingency and time-sensitivity of decision-making.

[11]More exactly, each system is thought of as a set of similar sub-systems.

https://doi.org/10.1017/S0266267122000189 Published online by Cambridge University Press

One question at this point is whether Pavlovian and habitual responses are value-based at all, which is not obvious. To support this hypothesis, Rangel *et al.* (2008: 547–548) insist on the fact that Pavlovian and habitual responses can be overruled in cases when they fail to deliver value to the agent. In the event they are genuine value-based systems, it would also be true that they assign values to actions rather than to goods or outcomes. The very definition of a habitual system by Fehr and Rangel (2011: 22) indeed reads that this system 'makes "choices" over actions but not stimuli'. The way both the Pavlovian and habitual systems work leads to recognizing that there are value-based decisions only apparently about goods but in fact about actions, as decisions concern default actions. In a restaurant, one can pretend every time to look at the menu for desserts and yet habitually (or even conditionally) order a crème caramel.

The discussion turns out to be conceptually more challenging with the goal-directed system proposed by Rangel and colleagues. In this system both goods and actions play a prominent role, but their coexistence is hindered by unresolved conceptual issues. To be aware of these issues, consider the following passage:

> Under ideal conditions, the value that is assigned to an action equals the average reward to which it might lead. We refer to values computed by this system as 'goal values' and to the actions that it controls as 'goal-directed behaviours'. An example of a goal-directed behaviour is the decision of what to eat at a new restaurant. (Rangel *et al.* 2008: 548)

The first impression reading this passage is that it pre-dates the actions vs goods controversy: while the unit of analysis is assumed to be action, the restaurant example points to the paradigmatic good-based situation. However, a more fundamental issue in this passage concerns the meaning of goal-directed behaviour. Rangel and colleagues make clear that in their goal-directed framework values are ultimately dependent on outcomes (Rangel *et al.* 2008: 548). Thus they take goal-directed behaviour to be an equivalent notion of outcome-directed behaviour and as such conceptually compatible with the G-model framework. However, as we will see, the conflation of a 'goal' with an 'outcome' is far from uncontroversial. An illuminating conversation between Padoa-Schioppa and Geoffrey Schoenbaum (a leading learning theorist) touches directly upon this crucial point, so it is worthy of being considered in detail. To Padoa-Schioppa's claim that 'what we call economic choice is closely related to what is often called goal-directed behavior' (Padoa-Schioppa and Schoenbaum 2015: 20), Schoenbaum replies:

> goal-directed behavior, as it is currently defined by learning theorists such as Balleine and Dickinson, is instrumental or specifically based on action-outcome associations. In this view, goal-directed behavior would have to be controlled by an action-based, not a good-based, representation. (Padoa-Schioppa and Schoenbaum 2015: 20)

This objection leads Padoa-Schioppa to reply that the idea of a goal is rather general and as such should not be exclusively related to actions: '[i]t is true', he says, 'that goal-directed behavior is often discussed as based on actions. But from my perspective, the concept of goal-directed behavior could easily generalize to abstract representations that do not depend on specific actions' (20). What is interesting in this dialogue is the compromise subscribed to when Padoa-Schioppa eventually states that '[b]orrowing your language, it is more accurate to say that economic choice is closely related to '*outcome-guided*' (*as distinguished from* '*goal-directed*') *behavior*' (20, italics added). In this way of differentiating things, two sets of qualifications are juxtaposed, one which contrasts an 'outcome' with a 'goal' and the other which contrasts a process being 'guided' with a process being 'directed'. As concerns the first juxtaposition, distinguishing an outcome from a goal goes conceptually in the direction of severing outcomes, as objects of choice, from a wide array of processes that are not merely propaedeutic but necessary to make decisions. For instance, cognitive approaches to learning have long maintained that pursuing a goal requires a complex interplay of cognitive processes such as planning and imagination (Verschure *et al.* 2014). Moreover, when distinguished from a goal, the idea of an outcome seems independent of instrumentality and instrumental action (Dickinson and Balleine 1994). As concerns the second juxtaposition, distinguishing 'guidedness' from 'directedness' seems to return an impoverished sense of agency in the choice process. While guidedness seems to represent agency as somehow constrained by an external lead (guided 'by'), agents' directedness seems rather to point to unconstrained and industrious agency (directed 'at'). Combined, Padoa-Schioppa's qualifications seem intended to carve out outcome-guidedness as a notion alternative to goal-directedness in which action plays a less prominent role, or even no role at all.

4.2. The inescapable ubiquity of action in value-based choices

Outcome-guided behaviour seems sometimes justified by the idea that there are decisions not plausibly having any action as an object. Cisek acknowledges the existence of this class of unmistakably good-based decisions (Cisek 2012). What would these decisions be? Buying a house, for instance. 'When deciding on a house to buy one is presumably not planning potential movements of opening the door, but instead is considering cost, space, commuting distance, and so on' (Cisek 2012: 932). This example hits the mark, for buying a house does not seem to involve any choice about actions. Yet, it can help us highlight a point that is usually overlooked and may lead to confusion. Sometimes the G-model is presented as applying whenever the relevant objects of choice are goods and not actions (as in the house buying example), while when it comes to neural correlates the gist of the model is presented as being the lack of sensorimotor neural signatures upstream of decisions (Padoa-Schioppa 2011). These, however, are two quite different things. When Padoa-Schioppa states that Cisek and other neuroscientists of naturalistic decisions have 'embraced the notion of good-based decisions' (Cai and Padoa-Schioppa 2014: 1140), it seems more likely the case that they have embraced the first rather than the second notion. In real-world environments, motor actions are continuously required even for simple decisions about goods, so a priori expecting a lack of sensorimotor neural signatures would be unwarranted. This section aims to emphasize that action is an inescapable ingredient of any value-based decision. Although not always the object of choice per se, it is otherwise ubiquitous whenever agents make decisions. Action can be so ubiquitous to render conceptually infeasible compressing every aspect of it into the idea of an outcome, as supporters of the G-model would have it. This suggests that a goal-directed framework is likely better equipped, from the conceptual point of view, to account for all the nuances of action in value-based choice.

A suitable place to start this discussion on the ubiquity of action is the role of vision, and particularly of visual fixation, in value-based decisions. In G-model experimental settings, monkeys typically fixate a point on the screen and are then exposed to options between which they choose through eye movements (Padoa-Schioppa 2011). In this setting, oculomotor movements are strictly regulated: choices are free but eyes are not. Krajbich *et al.* (2010) propose an alternative action-based model in which oculomotor movements are also free and the pattern of free fixation predicts choices. This model is interesting as it shows that even seemingly pure decisions about goods display a goal-directed structure dependent on action. The idea of pursuing a goal is represented in the model by a behavioural process in which the chooser accumulates evidence in favour of different options until a decision threshold (the 'goal') is hit for one option, which is then selected in a winner-take-all fashion. This model – called *attentional-drift-diffusion model* (ADDM) – assumes that areas guiding fixation receive inputs from value regions (such as OFC), and then bias decisions in comparator regions such as dorsomedial prefrontal cortex (dmPFC) and intraparietal sulcus (IPS) (Hare *et al.* 2011). This model assumes that visual fixation can capture features of choice options that are not strictly speaking value-based – e.g. the visual saliency of stimuli – but that nonetheless provide a crucial additional contribution to the value-assessment process upstream of decisions. Importantly, in ADDM visual fixation is considered to causally drive choices, as experimental manipulation of fixation patterns has been shown to affect decision-making (Armel *et al.* 2008). However, this line of causality from visual fixation to choice is the object of controversy. As has been remarked by supporters of the G-model, 'causality might well be in the opposite direction, in the sense that subjects in any trial might tend to look longer at the good they are leaning toward' (Padoa-Schioppa and Conen 2017: 748).

ADDM maintains that ocular movements play a causal role upstream, before a decision is made. Action conceived of this way is obviously relevant to the example of house buying. Even if buying a house does not seem to involve any choice about actions, ocular movements can account for the occurrence of framing effects related to the (non-strictly-value-based) perceptual salience of stimuli, which can indeed be crucial to prefer one house to another (as marketing experts know well). However, even if we accepted the G-model framework entirely, there would still be a place for action at the very core of the model: in the notion of good itself. That a good cannot be defined without recurring in some way to action is, we argue, a conceptual issue that should warn neuroeconomists against taking goods and actions as fully contrastable constructs. This is a crucial neurophilosophical point to consider. We have already emphasized that a good is a complex construct, insofar as it may vary in quantity, probability and action costs (Kennerley *et al.* 2009). We have also pointed out that the idea of a good is best expressed in the G-model by the more general idea of an outcome. Yet, one may argue that goods or outcomes are still understood rather narrowly in the G-model, as they are thought of independently of potential actions that goods and outcomes make possible. It is well known that in the visual system there are two systems, the so-called 'what' system which recognizes objects as such, and the so-called 'where and how' system which recognizes the spatial location of objects and the way to interact with them (see Gazzaniga *et al.* 2018: Ch. 6). It has been shown that people with damage to the 'what' system do not see objects in purely perceptual tasks but see them if they are to be used for action and interaction. Conversely, after damage to the 'where and how' system subjects recognize the objects but are unable to use them. As Gazzaniga *et al.* (2018: 263) say, '[w]hen people view pictures of nonliving objects such as tools … a region associated with action planning … is activated. Moreover, this region is also activated when the stimuli are pictures of natural objects that can be grasped and manipulated, such as a rock'. In other words, the value of non-living objects is also conferred by the action possibilities they make available, a notion of value known in psychology as 'affordance' (Tucker and Ellis 1998; de Wit *et al.* 2017). If this shows that the value (not to mention the very perception) of a good depends on the possibility of acting upon it – transcending the purely hedonic view – the concept of a good seems, in other respects, to be too broad in the G-model. A good-based choice can be about drops of juice, bottles of wine, houses, but other less commodity-like options seem equally good-like in the G-model. Consider, for instance, the choice of a friend. As no obvious choice between actions is involved when we choose a friend, friendship should count on the G-model's account as a good or an outcome not dissimilar from buying a house. These broad considerations suggest that a concept at the core of the G-model – the 'G' itself – requires further qualification and that the conceptual distinction between goods and actions requires caution as well.

5. Defining economic choice through the G-model

The debate about what economic choice consists of – goods or actions (or both) – is founded on the presumption that it is possible to define economic constructs through neural data. Although this presumption and its reductionist flaws have been much discussed by philosophers (e.g. Harrison 2008; Fumagalli 2013), there is a particular reason why we bring it up in our discourse. Supporters of the G-model seem to pursue a quite radical reductionist agenda: they do not merely inquire into whether, where, and how a notion of subjective utility is encoded and computed in the brain, but aim to provide a *normative neural definition* of economic choice. This ambition to define an economic choice neurally and normatively may be revealed to be quite problematic if not properly contextualized and epistemologically qualified, and for this reason will be extensively discussed in what follows.

Originally, there were good reasons for attempting to define the specific class of choice object of neuroeconomics. Looking at decision neuroscience before the onset of neuroeconomics, the constructs decision and choice mostly referred to the study of perceptual decisions (see Gold and Shadlen 2007). In perceptual decisions, 'experiments present monkeys with perceptually ambiguous sensory stimuli and ask them to "choose" between two possible reports. In such cases, monkeys are not asked to introspect and decide what they *want* … Instead, monkeys are asked to report what they *perceive*' (Padoa-Schioppa 2007: 233). In the attempt to differentiate their object of study from perceptual decisions, neuroeconomists maintained that '[e]conomic choice can be defined as the behavior observed when individuals make choices based solely on subjective preferences' (Padoa-Schioppa: 2011: 334). In a very specific sense, the label 'economic choice' was used at first to denote the class of non-perceptual, value-based decisions. The situation has become more intellectually ambitious, but also more controversial, when the intended recipients of the definition of economic choice have become economists themselves. Even economists, it has been said, would benefit from a definition of economic choice as '[t]he distinction between "economic" decisions and other types of decisions is not one usually made in economic theory' (Padoa-Schioppa and Schoenbaum 2015: 17 fn. 3). Addressing economists confers a particular intent to the definitional exercise, for the attribute 'economic' is no longer merely instrumental to distinguish one class of choice from another in neuroscience but aims to introduce a definition of what an economic choice is *in economics*.

The G-model definition of economic choice spans different epistemological levels that it is useful to reconstruct. The 'starting point [of the definition] is an appeal to intuition' (Padoa-Schioppa and Schoenbaum 2015: 17), which leads to assuming that economic choice is 'a distinct mental process' (Padoa-Schioppa 2007: 335). The class of reward-based free choices usually called by neutral names (e.g. 'simple choices', see Krajbich *et al.* 2010; Hare *et al.* 2011) becomes, by appealing to intuition, the class of semantically characterized 'economic' choices. Strictly connected to this intuition is the identification of a situational requirement, i.e. the idea that economic choices are the menu choices that one typically makes in a restaurant (the restaurant example is ubiquitous in G-model papers; see e.g. Padoa-Schioppa and Assad 2007; Padoa-Schioppa and Conen 2017). Then, as a central behavioural requirement, it is assumed that '[t]ransitivity is a fundamental trait of economic choice behavior' (Padoa-Schioppa and Assad 2007: 96). Only at this point do these intuitive, situational and behavioural requirements become the basis for the neural definition of economic choice. As Padoa-Schioppa says, '[i]n the long run … appealing to intuition or even to the behavioral criteria described above is somewhat unsatisfactory. In my view, different kinds of decisions should eventually be defined based on the underlying neural mechanisms' (Padoa-Schioppa and Schoenbaum 2015: 17). Based on the abstract way OFC encodes values, the resulting definition reads, as already said, that economic choice 'is essentially choice between goods rather than choice between actions' (Padoa-Schioppa and Assad 2006: 223).

Such a definition of economic choice stretches across four quite heterogeneous conceptual levels: an *intuitive*, a *situational*, a *behavioural* and a *neural* level (in order of appearance). Taken singularly, they have been extensively discussed for their role in defining economic choice. For instance, philosophers of economics have long debated whether failure to satisfy transitivity at the behavioural level would prevent a choice from being deemed economic (see Hands 2015). In other words, philosophers have long discussed whether transitivity is a normative behavioural requirement for economic choice. Beyond considering the normative contribution of each conceptual level to the definition, it would also be necessary for all the levels to integrate one with the other. In this regard, when Padoa-Schioppa says that an '[e]conomic choice can be defined as the behavior observed when individuals make choices based solely on subjective preferences' (Padoa-Schioppa 2011: 334), this is an intuitive definition of economic choice that does not include the behavioural criterion that subjective preferences need also be transitive. From a definitional point of view, it would be crucial to clarify which criteria and levels are necessary and which instead are sufficient for the definition of economic choice. For instance, is subjective desirability merely necessary or also sufficient for a choice to be economic?

To assess the G-model definition of economic choice, another way is to consider subsets of the four conceptual levels. We can consider, for instance, the combination of behavioural and neural levels. In this regard, the G-model employs the gold standard of neuroscience methodology, the so-called 'psychometric-neurometric' approach (see Kable and Glimcher 2007), which looks for isomorphism between behavioural and neural data. The potential issues with this approach are well known and concern, for instance, the assumption of existence of neural representations[12] and the merely correlative nature of the method.[13] However, what interests us here is the G-model's use of this approach for definitional purposes, which can help us shed light on neuroeconomic practices more generally. In neuroeconomics, it is often said that the gist of the method lies in the 'attempt to *describe* the neurobiological substrate of a behavioral choice using a form of *normative* choice theory derived from economics' (Glimcher *et al.* 2009: 8, italics added). As this quote shows, when neuroeconomists consider rational choice theory normative, they mostly refer to the positive use of the theory to make neural predictions. But as Fumagalli's (2013) analysis elucidates and warns against, this may entail a confusion of categories. A possible way to reconcile the normative-descriptive tension inherent in the method of neuroeconomics is by emphasizing, as Padoa-Schioppa does, that the match between behavioural and neural data can ultimately have a definitional role. But the intuitive and situational levels of the G-model definition also give cause for epistemological concern. Definitions based on intuition can lead to misconceptions. To be aware of how it may be, we suggest adopting the 'heuristic' view of scientific discovery (e.g. Nersessian 1992), which, in contrast to other more liberal views, allows to assess discovery normatively (see Simon 1973). As seen, intuition of what economic choice is has strictly intertwined in the G-model with a prototypic choice situation: menu choices in a restaurant.

[12]This is an assumption shared by all parties in the controversy we are examining, so we do not discuss it here. However, it itself can be questioned in neuroeconomics (see Petracca 2020).

[13]In a recent study using electrical stimulation, Ballesta *et al.* (2020) find that OFC is causally related to economic choices.

This intuition has heuristically led to restricting the inquiry to highly artificial and non-sensorimotor-intensive experimental settings. It is usually held that experiments on economic choice are performed in minimal settings for they represent the experimental counterpart of minimal models of economic choice (Grüne-Yanoff 2009), and that the prospect is to progressively extend experiments to more and more complex choice situations. The potential issue here is that, when used heuristically, situational intuition may lead to a form of *confirmation bias*. Restaurant-like situations might not simply be convenient experimental settings but the very situational embodiment of the idea of economic choice. To the point that there would be no need to study economic choice in more complex settings, or, even, that the failure of the G-model to apply to more complex settings would not affect its definition of economic choice.

The distinctive feature of the G-model lies not so much in taking the four conceptual levels together but in understanding them, and their specific content, as the basis of a normative definition of economic choice. However, as seen, what supporters of the G-model aim to do in the long run is even more normatively restrictive, for they aim to define economic choice exclusively through neural data (Padoa-Schioppa and Schoenbaum 2015). The G-model prescribes that to define economic choice, neural signals consistent with expected utility are not enough: neural encoding must also be free of sensorimotor signatures. However, as discussed above, motor-intensive choice situations may render difficult the identification of abstract neural signals. If choosing between a bottle of Nebbiolo and one of Negramaro from a menu were regarded as categorically different from choosing between the same bottles while holding them – as neural traces might be abstract in the former case (satisfying the definition of economic choice) and non-abstract in the latter case (not satisfying the definition of economic choice) – abstract neural encoding would hardly be a reasonable discrimination criterion. The neural requirements posited by the G-model for the definition of economic choice risk imposing categorical distinctions between choices that in real life are not only hardly distinguishable but whose distinction could be unproductive or even misleading. In brief, the G-model's neural requirements might produce arbitrary distinctions between classes of choice.

6. Concluding remarks

The discussion above comes in the wake of philosophers' long efforts to scrutinize experimental findings and frameworks on the neural foundations of utility (e.g. Fumagalli 2010, 2013, 2016, 2019; Vromen 2010; Marchionni and Vromen 2020). The proposed neurophilosophical assessment of the controversy between the G-model and the A-model has contributed to this literature in two ways. The first is a suggestion to extend neuroeconomic model discrimination to the scrutiny of the entire range of neurophilosophical assumptions (Herrmann-Pillath 2021). Similar to economics, neuroeconomic models come with a series of characterizing assumptions: if economists' assumption of perfect knowledge signals neoclassical inclination and consequent resistance to behavioural findings, assuming modularity of valuation and action suggests a non-ecological bent in neuroeconomics (Cisek and Pastor-Bernier 2014). We have argued that both philosophers and economists would benefit from considering these assumptions since they are directly relevant to assessing, among other things, the explanatory status of neuroeconomic models. As shown, non-ecological assumptions render the G-model structurally unfit to explain the entire range of value-based decisions that agents make in real life. But delving into neuroscientific details has also revealed to be useful in other respects, for we have had the chance to identify alternative neural models, such as DCM, conducive to extending the scope of neuroeconomic choice (Petracca 2020).

Our second contribution has concerned the neurophilosophical risks of grounding neuroeconomic models on a reductionist nexus between economics and the G-model. The prospect for economics of neurally vindicating cardinal utility on the one hand, and the promise of the G-model to normatively define economic choice through neural criteria on the other hand, represent two forms of neural reductionism that philosophers have often warned against (e.g. Craver and Alexandrova 2008; Fumagalli 2016; Herrmann-Pillath 2016; Marchionni and Vromen 2020). We have tried to explore the ultimate consequences of such reductionism in order to expose its limits and risks. On the economics side, taking the G-model as a guide for identifying the true object of the discipline – goods instead of actions, outcomes instead of goals – would lead economics to relinquish its traditional image as a science of goal-directedness and embrace that of a science of outcome-guidedness. This would entail leaving entire traditions out of the realm of economics, the Austrian tradition coming immediately to mind as an instance (e.g. von Mises 1998 [1949]). But, more fundamentally, the very image of economics as a science of actors or agents who actively pursue goals in light of instrumental rationality would be put into question. On the neuroeconomic side, positing strict normative neural requirements for the definition of economic choice could take out of the definition's scope choices that by any other criterion we would consider economic in nature. Again, this is a consequence of too strict distinctions both between goods and actions, and between outcomes and goals. As argued above, such distinctions are too radical in experimental practice as well as neurophilosophically disputable. Both sides of reductionism produce in the end excessively restrictive definitions of economic choice and economics. If the ability to 'save the phenomena' is a key criterion for model assessment and discrimination, the G-model currently seems to save neither the realm of economic choice nor that of economics.

Acknowledgements. I thank Roberto Fumagalli for the detailed comments on an earlier version of this paper. Two anonymous reviewers provided helpful and constructive comments for which I am grateful. I also benefited from exchanges with Paul Cisek, Shaun Gallagher, Camillo Padoa-Schioppa and Michael Platt. The usual caveats apply.

References

Abe H. and D. Lee 2011. Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. *Neuron* **70**, 731–741.

Anderson M.L. 2014. *After Phrenology: Neural Reuse and the Interactive Brain*. Cambridge, MA: MIT Press.

Armel K.C., A. Beaumel and A. Rangel 2008. Biasing simple choices by manipulating relative visual attention. *Judgment and Decision Making* **3**, 396–403.

Balleine B.W. and A. Dickinson 1998. Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. *Neuropharmacology* **37**, 407–419.

Ballesta S. and C. Padoa-Schioppa 2019. Economic decisions through circuit inhibition. *Current Biology* **29**, 3814–3824.

Ballesta S., W. Shi, K.E. Conen and C. Padoa-Schioppa 2020. Values encoded in orbitofrontal cortex are causally related to economic choices. *Nature* **588**, 450–453.

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

Bickle J., P. Mandik and A. Landreth 2019. The philosophy of neuroscience. In *The Stanford Encyclopedia of Philosophy* (Fall Edition), ed. E.N. Zalta. https://plato.stanford.edu/archives/fall2019/entries/neuroscience/.

Blanchard T.C., B.Y. Hayden and E.S. Bromberg-Martin 2015. Orbitofrontal cortex uses distinct codes for different choice attributes in decisions motivated by curiosity. *Neuron* **85**, 602–614.

Cai X. and C. Padoa-Schioppa 2014. Contributions of orbitofrontal and lateral prefrontal cortices to economic choice and the good-to-action transformation. *Neuron* **81**, 1140–1151.

Cai X. and C. Padoa-Schioppa 2019. Neuronal evidence for good-based economic decisions under variable action costs. *Nature Communications* **10**, 1–13.

Camerer C.F. 2013. Goals, methods, and progress in neuroeconomics. *Annual Review of Economics* **5**, 425–455.

Chen X. and V. Stuphorn 2015. Sequential selection of economic good and action in medial frontal cortex of macaques during value-based decisions. *Elife* **4**, e09418.

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

Cisek P. and J.F. Kalaska 2010. Neural mechanisms for interacting with a world full of action choices. *Annual Review of Neuroscience* **33**, 269–298.

Cisek P. and A. Pastor-Bernier 2014. On the challenges and mechanisms of embodied decisions. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130479.

Clithero J.A. and A. Rangel 2014. Informatic parcellation of the network involved in the computation of subjective value. *Social Cognitive and Affective Neuroscience* **9**, 1289–1302.

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

de Wit M.M., S. de Vries, J. van der Kamp and R. Withagen 2017. Affordances and neuroscience: steps towards a successful marriage. *Neuroscience & Biobehavioral Reviews* **80**, 622–629.

Dickinson A. and B.W. Balleine 1994. Motivational control of goal-directed action. *Animal Learning & Behavior* **22**, 1–18.

Dorris M.C. and P.W. Glimcher 2004. Activity in posterior parietal cortex is correlated with the relative subjective desirability of action. *Neuron* **44**, 365–378.

Fehr E. and A. Rangel 2011. Neuroeconomic foundations of economic choice: recent advances. *Journal of Economic Perspectives* **25**, 3–30.

Fumagalli R. 2010. The disunity of neuroeconomics: a methodological appraisal. *Journal of Economic Methodology* **17**, 119–131.

Fumagalli R. 2013. The futile search for true utility. *Economics and Philosophy* **29**, 325–347.

Fumagalli R. 2016. Five theses on neuroeconomics. *Journal of Economic Methodology* **23**, 77–96.

Fumagalli R. 2019. (F)utility exposed. *Philosophy of Science* **86**, 955–966.

Fumagalli R. 2020. On the individuation of choice options. *Philosophy of the Social Sciences* **50**, 338–365.

Gazzaniga M.S., R.B. Ivry and G.R. Mangun 2018. *Cognitive Neuroscience: The Biology of the Mind* (6th edn). New York, NY: Norton & Company.

Glimcher P.W. 2011. *Foundations of Neuroeconomic Analysis*. New York, NY: Oxford University Press.

Glimcher P.W., C.F. Camerer, E. Fehr and R.A. Poldrack 2009. Introduction: a brief history of neuroeconomics. In *Neuroeconomics: Decision-Making and the Brain*, eds. P.W. Glimcher, C.F. Camerer, E. Fehr and R.A. Poldrack, 1–12. London: Academic Press.

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

Gold J.I. and M.N. Shadlen 2007. The neural basis of decision making. *Annual Review of Neuroscience* **30**, 535–574.

Grattan L.E. and P.W. Glimcher 2014. Absence of spatial tuning in the orbitofrontal cortex. *PLoS ONE* **9**, e112750.

Grüne-Yanoff T. 2009. Learning from minimal economic models. *Erkenntnis* **70**, 81–99.

Hands D.W. 2015. Normative rational choice theory: past, present and future. http://ssrn.com/abstract=1738671.

Hare, T.A., W. Schultz, C.F. Camerer, J.P. O'Doherty and A. Rangel 2011. Transformation of stimulus value signals into motor commands during simple choice. *Proceedings of the National Academy of Sciences USA* **108**, 18120–18125.

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

Hayden B.Y. and Y. Niv 2021. The case against economic values in the orbitofrontal cortex (or anywhere else in the brain). *Behavioral Neuroscience* **135**, 192–201.

Herrmann-Pillath C. 2016. Constitutive explanations in neuroeconomics: principles and a case study on money. *Journal of Economic Methodology* **23**, 374–395.

Herrmann-Pillath C. 2021. Evolutionary mechanisms of choice: Hayekian perspectives on neurophilosophical foundations of neuroeconomics. *Economics and Philosophy* **37**, 284–303.

Hunt L.T. and B.Y. Hayden 2017. A distributed, hierarchical and recurrent framework for reward-based choice. *Nature Reviews Neuroscience* **18**, 172–182.

Kable J.W. and P.W. Glimcher 2007. The neural correlates of subjective value during intertemporal choice. *Nature Neuroscience* **10**, 1625–1633.

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

Kennerley S.W., A.F. Dahmubed, A.H. Lara and J.D. Wallis 2009. Neurons in the frontal lobe encode the value of multiple decision variables. *Journal of Cognitive Neuroscience* **21**, 1162–1178.

Kennerley S.W. and M.E. Walton 2011. Decision making and reward in frontal cortex: complementary evidence from neurophysiological and neuropsychological studies. *Behavioral Neuroscience* **125**, 297–317.

Krajbich I., C. Armel and A. Rangel 2010. Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience* **13**, 1292.

Kubanek J. and L.H. Snyder 2015. Reward-based decision signals in parietal cortex are partially embodied. *Journal of Neuroscience* **35**, 4869–4881.

Leathers M.L. and C.R. Olson 2012. In monkeys making value-based decisions, LIP neurons encode cue salience and not action value. *Science* **338**, 132–135.

Marchionni C. and J. Vromen, eds. 2014. *Neuroeconomics: Hype or Hope?* London: Routledge.

Marchionni C. and J. Vromen 2020. Neuroeconomics as a case of mechanistic integration. In *Social Neuroeconomics: Mechanistic Integration of the Neurosciences and the Social Sciences*, eds. J. Harbecke and C. Herrmann-Pillath, 13–30. London: Routledge.

McMaster R. and M. Novarese 2016. Neuroeconomics: infeasible and underdetermined. *Journal of Economic Issues* **50**, 963–983.

Moscati I. 2018. *Measuring Utility: From the Marginal Revolution to Behavioral Economics*. New York, NY: Oxford University Press.

Nersessian N. 1992. How do scientists think? Capturing the dynamics of conceptual change in science. In *Cognitive Models of Science*, ed. R. Giere, 3–45. Minneapolis, MN: Minnesota University Press.

O'Doherty J.P. 2014. The problem with value. *Neuroscience & Biobehavioral Reviews* **43**, 259–268.

Padoa-Schioppa C. 2007. Orbitofrontal cortex and the computation of economic value. *Annals of the New York Academy of Sciences* **1121**, 232–253.

Padoa-Schioppa C. 2008. The syllogism of neuro-economics. *Economics and Philosophy* **24**, 449–457.

Padoa-Schioppa C. 2009. Range-adapting representation of economic value in the orbitofrontal cortex. *Journal of Neuroscience* **29**, 14004–14014.

Padoa-Schioppa C. 2011. Neurobiology of economic choice: a good-based model. *Annual Review of Neuroscience* **34**, 333–359.

Padoa-Schioppa C. and J.A. Assad 2006. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441**, 223–226.

Padoa-Schioppa C. and J.A. Assad 2007. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature Neuroscience* **11**, 95–102.

Padoa-Schioppa C. and X. Cai 2011. Orbitofrontal cortex and the computation of subjective value: consolidated concepts and new perspectives. *Annals of the New York Academy of Sciences* **1239**, 130–137.

Padoa-Schioppa C. and K.E. Conen 2017. Orbitofrontal cortex: a neural circuit for economic decisions. *Neuron* **96**, 736–754.

Padoa-Schioppa C. and G. Schoenbaum 2015. Dialogue on economic choice, learning theory, and neuronal representations. *Current Opinion in Behavioral Sciences* **5**, 16–23.

Pearson J.M., K.K. Watson and M.L. Platt 2014. Decision making: the neuroethological turn. *Neuron* **82**, 950–965.

Pelletier G. and L.K. Fellows 2019. A critical role for human ventromedial frontal lobe in value comparison of complex objects based on attribute configuration. *Journal of Neuroscience* **39**, 4124–4132.

Petracca E. 2017. A cognition paradigm clash: Simon, situated cognition and the interpretation of bounded rationality. *Journal of Economic Methodology* **24**, 20–40.

Petracca E. 2020. Neuroeconomics beyond the brain: some externalist notions of choice. *Journal of Economic Methodology* **27**, 275–291.

Platt M. L. and P.W. Glimcher 1999. Neural correlates of decision variables in parietal cortex. *Nature* **400**, 233–238.

Platt M.L. and C. Padoa-Schioppa 2009. Neuronal representations of value. In *Neuroeconomics: Decision Making and the Brain*, eds. P.W. Glimcher, C.F. Camerer, E. Fehr and R.A. Poldrack, 441–462. Amsterdam: Academic Press.

Rangel A. and T. Hare 2010. Neural computations associated with goal-directed choice. *Current Opinion in Neurobiology* **20**, 262–270.

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

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

Rushworth M.F., N. Kolling, J. Sallet and R.B. Mars 2012. Valuation and decision-making in frontal cortex: one or many serial or parallel systems?. *Current Opinion in Neurobiology* **22**, 946–955.

Schoenbaum G., M.R. Roesch, T.A. Stalnaker and Y.K. Takahashi 2009. A new perspective on the role of the orbitofrontal cortex in adaptive behaviour. *Nature Reviews Neuroscience* **10**, 885–892.

Schoenbaum G., Y. Takahashi, T.L. Liu and M.A. McDannald 2011. Does the orbitofrontal cortex signal value?. *Annals of the New York Academy of Sciences* **1239**, 87–99.

Simon H.A. 1973. Does scientific discovery have a logic?. *Philosophy of Science* **40**, 471–480.

Tremblay L. and W. Schultz (1999). Relative reward preference in primate orbitofrontal cortex. *Nature* **398**, 704–708.

Tucker M. and R. Ellis 1998. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance* **24**, 830.

Verschure P.F., C.M. Pennartz and G. Pezzulo 2014. The why, what, where, when and how of goal-directed choice: neuronal and computational principles. *Philosophical Transactions of the Royal Society B: Biological Sciences* **369**, 20130483.

von Mises L. 1998 [1949]. *Human Action: A Treatise on Economics*. Auburn, AL: Ludwig von Mises Institute.

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

Wallis J.D. 2012. Cross-species studies of orbitofrontal cortex and value-based decision-making. *Nature Neuroscience* **15**, 13–19.

Enrico Petracca is a Research Associate at the School of Economics, Management and Statistics of the University of Bologna. His research lies at the intersection between the philosophy of economics and cognitive science, especially ‘embodied’ cognitive science.

Cite this article: Petracca E (2023). The utility of goods or actions? A neurophilosophical assessment of a recent neuroeconomic controversy. *Economics & Philosophy* **39**, 351–372. https://doi.org/10.1017/S0266267122000189