(F)utility Exposed

Roberto Fumagalli*†

In recent years, several authors have called to ground descriptive and normative decision theory on neuropsychological measures of utility. In this article, I combine insights from the best available neuropsychological findings, leading philosophical conceptions of welfare, and contemporary decision theory to rebut these prominent calls. I argue for two claims of general interest to philosophers, choice modelers and policy makers. First, severe conceptual, epistemic, and evidential problems plague ongoing attempts to develop accurate and reliable neuropsychological measures of utility. Second, even if these problems are solved, neuropsychological measures of utility lack the potential to directly inform welfare analyses and policy evaluations.

1. Introduction. Over the last 2 decades, neuropsychological findings have enabled choice modelers to make significant advances in predicting and explaining individuals' choices (e.g., Camerer 2008; Glimcher 2009). Building on these advances, several authors have called to ground decision theory on neuropsychological measures of utility (e.g., Dolan and Kahneman 2008; Glimcher 2010). The idea is that while standard decision theory (e.g., Savage 1954) takes rational agents to behave *as if* they maximize expected utility, recent neuropsychological findings demonstrate that utility is literally computed by individuals' brains and can be accurately and reliably measured in *psychophysical* terms (e.g., Kahneman, Wakker, and Sarin 1997; Kable and Glimcher 2009). These findings, in turn, are claimed to prompt revolutionary changes in both descriptive and normative decision theory. Descriptively, they allegedly provide "a neurobiological foundation for standard [decision]

* To contact the author, please write to: Department of Political Economy, Bush House—North East Wing, King's College London, Aldwych, WC2B 4BG, London; e-mail: roberto.fumagalli@kcl.ac.uk.

† I received helpful feedback from participants at the 26th Biennial Meeting of the Philosophy of Science Association (Seattle), the 92nd Joint Session of the Aristotelian Society and Mind Association (Oxford), the 2018 Annual Conference of the British Society for Philosophy of Science (Oxford), and the conference Neuroethics: Re-mapping the Field (Milan).

Philosophy of Science, 86 (December 2019) pp. 955–966. 0031-8248/2019/8605-0012$10.00
Copyright 2019 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/705454 Published online by Cambridge University Press

theory and a tool for measuring preferences neurobiologically" (Levy and Glimcher 2012, 1027). Normatively, they purportedly undermine decision theory's "tight connection between choice and welfare" and enable researchers to identify a neuropsychological notion of "*true utility*, which encapsulates well-being" (Bernheim and Rangel 2009, 51).[1]

Calls to ground decision theory on neuropsychological measures of utility have gained increasing popularity among choice modelers (e.g., Dolan and White 2007; O'Doherty 2014) and support dissimilar evaluations of policies' welfare implications than standard decision theory (e.g., Hausman 2010; Angner 2011). In this article, I combine insights from the best available neuropsychological findings, leading philosophical conceptions of welfare, and contemporary decision theory to rebut such prominent calls. I argue for two claims of general interest to philosophers, choice modelers, and policy makers. First, severe conceptual, epistemic, and evidential problems plague ongoing attempts to develop accurate and reliable neuropsychological measures of utility. And second, even if these problems are solved, neuropsychological measures of utility lack the potential to directly inform welfare analyses and policy evaluations. My point is not merely that different authors speak of 'utility' in dissimilar senses or that current neuropsychological measures are insufficiently developed. Rather, my thesis is that the search for true utility rests on ill-founded empirical and methodological presuppositions and that researchers should abandon this search altogether.

The article is organized as follows. Section 2 contrasts the notion of *decision utility* that figures prominently in standard decision theory and the notions of *true utility* put forward in the recent neuropsychological literature. Section 3 examines the issue whether the available neuropsychological findings enable researchers to develop accurate and reliable neuropsychological measures of utility. Section 4 critically assesses the potential of neuropsychological *measures* of utility to inform *welfare analyses* and *policy evaluations*. In doing so, it identifies and rebuts various objections that the proponents of true utility may put forward to vindicate their reliance on neuropsychological measures of utility.

2. Decision Utility versus True Utility. The notion of *decision utility* (DU) refers to a mathematical representation of agents' preferences over choice options to be inferred from agents' choices between such options (e.g., Broome 1991). DU figures prominently in the representation theorems at

1. I speak of 'neuropsychology' broadly to encompass several studies targeting individuals' hedonic experiences and neural areas' activation patterns. Also, I use the terms 'well-being' and 'welfare' interchangeably to indicate what makes a life good for the individual living such a life (e.g., Griffin 1986).

https://doi.org/10.1086/705454 Published online by Cambridge University Press

the core of standard decision theory (e.g., Savage 1954). The idea is that if an agent's preferences satisfy specific axiomatic requirements (e.g., transitivity, completeness), then the agent necessarily behaves *as if* she maximizes expected utility; that is, of any two options, the one preferred is assigned higher expected utility. In this context, utility is regarded as derivative from choices and is ascribed no sensory or psychological properties (e.g., Okasha 2016). That is to say, "when [the relevant] axioms are obeyed . . . utility is not 'a feeling', or 'happiness', or 'a hedonic impulse' [but is just a mathematical index of] 'choice' " (Glimcher 2009, 505).

The notion of *true utility* (TU) refers to a measurable psychophysical magnitude rather than a mathematical construct such as DU (e.g., Fumagalli 2013). In the recent neuropsychological literature, various notions of TU have been proposed. The notions of experienced utility and neural utility are especially influential in this context. These two notions refer to empirical facts about individuals' neuropsychology rather than to a mathematical index of agents' choices. More specifically, the notion of *experienced utility* (EU) refers to a psychophysical magnitude that putatively reflects individuals' hedonic experiences and causally influences individuals' choices (e.g., Kahneman 2000). EU figures prominently both in psychological studies of individuals' hedonic experiences and in some philosophers' interpretations of decision theory (e.g., Dietrich and List 2016). In fact, EU is often claimed to resemble early utilitarian philosophers' notion of utility (e.g., Kahneman et al. 1997, on Bentham). The notion of *neural utility* (NU) refers to the activation patterns of specific neural areas and the rewards' subjective values allegedly computed and integrated by such areas (e.g., Glimcher 2009, 507; also 2010, chaps. 12–15). NU figures prominently in the specialized neuropsychological literature. In this literature, there is growing consensus that activations in the ventromedial prefrontal cortex/orbitofrontal cortex (vmPFC/OFC) track rewards' subjective values across multiple reward types (e.g., Levy and Glimcher 2012). The idea is that these subjective values "are exactly the kind of value representation posited by most decision theories . . . analogous to what economists call 'utilities' " (Kable and Glimcher 2009, 734).

3. True Utility: Descriptive Evaluation. In this section, I examine the issue whether the available neuropsychological findings enable researchers to develop accurate and reliable neuropsychological measures of utility. I argue that ongoing attempts to develop such measures face at least four major challenges: the challenge from *measurement divergences*, the challenge from *indeterminate referents*, the challenge from *unclear unit of measurement*, and the challenge from *conceptual barriers*. Some of these challenges (e.g., measurement divergences) affect also DU measures. Other challenges, instead,

https://doi.org/10.1086/705454 Published online by Cambridge University Press

specifically target TU (rather than DU) measures and license selective skepticism regarding TU measures.[2]

3.1. Challenge from Measurement Divergences. The challenge from measurement divergences builds on the profound divergences between measurements that purportedly target the same notions of TU to question the prospects of identifying "a single, unifying concept that motivates all human choices and registers all relevant feelings and experiences" (Kahneman and Krueger 2006, 4). This challenge targets both EU and NU measurements.

EU measurements can vary remarkably depending on what methods one uses to measure EU (e.g., Fumagalli [2013], on divergences between physiological indices and individuals' reports of their own hedonic experiences). In fact, significant divergences arise even if one focuses on specific subtypes of EU such as *instant utility*, which targets momentary evaluations of ongoing hedonic experiences, *remembered utility*, which targets retrospective evaluations of past hedonic experiences, or *anticipated utility*, which targets predictions concerning future hedonic experiences (e.g., Dolan and Kahneman 2008). These divergences arise not just from putative limitations of current neuropsychological measurements but also from specific properties of the notions of EU supposedly targeted by such measurements. For instance, the act of remembering past experiences typically involves active reconstruction (rather than passive recall) of these experiences. As a result, remembered utility measurements frequently fail to accurately reflect how pleasant past events were while being experienced (e.g., Kahneman [2000], on so-called duration neglect). These concerns exacerbate when one considers that EU measurements can vary remarkably because of factors (e.g., temporary weather conditions) that are not directly related to the valence/phenomenal qualities of individuals' hedonic experiences.

Regarding NU, at least four features of the human neural architecture can generate significant divergences between different NU measurements: (1) rewards' subjective values depend not only on rewards' *objective attributes* but also on interpersonally and intertemporally variable *reference points*, and different measures diverge as to how these reference points are affected by individuals' past, current, and anticipated rewards (e.g., Glimcher 2010); (2) to encode a potentially infinite range of reward values in a finite range of activations, the neural substrates of reward valuation do not represent rewards' *absolute* values but rely on *normalization* processes that hinder the use of specific areas' neural activations as accurate and reliable value-related measures across individuals and measurement contexts (e.g., Louie, Khaw,

2. I speak of measurement broadly to include distinct approaches to measurement (e.g., Alexandrova and Haybron 2016; Moscati 2018). I mention these approaches in passing since my evaluation does not directly rest on those approaches' merits and interrelations.

https://doi.org/10.1086/705454 Published online by Cambridge University Press

and Glimcher 2013); (3) different NU measurements rest on dissimilar assumptions concerning neural substrates' *information processing* (e.g., efficiency of information coding) and neural activations' degree of *dependence* on the available rewards and stimuli (e.g., Padoa-Schioppa [2011], on various differences between reinforcement learning models and other models of reward valuation); (4) the brain does not *invariably* evaluate the impact of all factors that influence individuals' choices in terms of a direct computation of rewards' subjective values (e.g., Ross 2008), so "there are many behaviors that can be described as a decision but are not [directly based on rewards'] subjective values" (Padoa-Schioppa and Schoenbaum 2015, 17).

3.2. Challenge from Indeterminate Referents. The proponents of TU often leave it indeterminate whether TU tracks only psychophysical *feelings* (*hedonic* interpretations of TU) or also individuals' cognitive/evaluative *attitudes* regarding such feelings (*attitudinal* interpretations of TU). In fact, even the same authors frequently fail to specify which of these interpretations of TU they adopt (e.g., Angner 2011, on Kahneman 2000). This lack of specificity complicates attempts to provide accurate and reliable neuropsychological measures of utility in at least three respects: (1) neuropsychological measurements *significantly vary* depending on whether one adopts hedonic or attitudinal interpretations of TU (e.g., Crisp [2006], on hedonic and attitudinal EU; Berridge and O'Doherty [2014], on hedonic and attitudinal NU); (2) authors who agree on whether to adopt hedonic (or attitudinal) interpretations of TU frequently disagree regarding *what set* of feelings (or attitudes) are tracked by TU (e.g., Loewenstein and Ubel 2008); (3) attempts to determine the referents of TU pose the following *dilemma* to the proponents of TU. On the one hand, hedonic interpretations of TU fail to track many of the referents that researchers associate with TU (e.g., Heathwood [2006], on physiological indices' failure to accurately track the influence of individuals' desires on the valence/phenomenal qualities of their own hedonic experiences). On the other hand, attitudinal interpretations of TU face severe measurability concerns. To illustrate these concerns, consider measurements of attitudinal EU.

There are several reasons why measurements of attitudinal EU may fail to accurately track the valence/phenomenal qualities of individuals' hedonic experiences: (1) the valence/phenomenal qualities of some hedonic experiences are hard to *assess* for individuals (e.g., Haybron [2005], on various moods), and individuals' *abilities* to assess such valence/phenomenal qualities vary across individuals and situations in ways that are difficult to monitor (e.g., Alexandrova and Haybron 2016); (2) the *scales* individuals use to evaluate their own hedonic experiences vary across time and situations to a degree that hampers both the intrapersonal and the interpersonal comparability of individuals' evaluations (e.g., Loewenstein and Ubel [2008], on

https://doi.org/10.1086/705454 Published online by Cambridge University Press

individuals' tendency to normalize hedonic evaluations of their own health conditions based on implicit standards of comparison); (3) individuals' tendency to *adapt* to affects and circumstances can alter their hedonic evaluations in ways that elude both their own awareness and physiological indices of EU (e.g., Van der Deijl 2017); (4) *social norms and pressures* can demonstrably modify individuals' hedonic evaluations in the absence of experimentally measurable variations in their hedonic experiences (e.g., Dolan and White 2007); (5) individuals' hedonic evaluations can be significantly affected by their own *expectations* concerning how they ought (or are likely) to feel in specific circumstances (e.g., Haybron 2007).

3.3. Challenge from Unclear Unit of Measurement. The proponents of TU frequently maintain that neuropsychological measures of utility enable researchers to "objectively compare mental state between individuals" (Glimcher 2010, 425) and provide "policymakers [with] a standard unit of measurement for comparisons of well-being across domains" (Dolan and White 2007, 71). However, it remains unclear what unit of measurement is supposed to ground these comparisons. Moreover, attempts to identify such a unit of measurement face daunting difficulties in case of both EU and NU.

EU is often claimed to track a common quality of pleasantness or desirability allegedly shared by individuals' hedonic experiences (e.g., Kahneman and Krueger 2006). The qualitative differences between distinct hedonic experiences do not exclude the possibility that these experiences may vary in magnitude along some common scale. Still, the proponents of EU have hitherto failed to agree as to *what* this putative common scale consists in. In fact, several authors doubt the prospects of identifying a common scale for measuring the pleasantness or desirability of individuals' hedonic experiences (e.g., Crisp 2006). A proponent of EU may respond that even if researchers are unable to identify such a common scale, EU can be taken to track all experiences toward which individuals have some positive attitude. However, to address the challenge from unclear units of measurement, this response must be supplemented with a clear specification of *which* positive attitudes are tracked by EU and how exactly these attitudes are to be *measured*. In the absence of this specification, lumping all experiences toward which individuals have some positive attitude under the label of EU obscures (rather than solves) the measurability problems posed by the diversity (and potential incommensurability) of individuals' hedonic experiences.

As to NU, several authors hold that for individuals to be able to choose between complex (e.g., multiattribute) alternatives, their neural architecture must compute and compare rewards' subjective values on a common neural scale (e.g., Landreth and Bickle 2008). In particular, leading researchers concur that anatomically delimited activations in vmPFC/OFC "encode the subjective values of different types of rewards on a neural common scale"

https://doi.org/10.1086/705454 Published online by Cambridge University Press

(Levy and Glimcher 2012, 1027). There are at least four reasons to doubt these identifications of a common neural scale: (1) *several areas* besides vmPFC/OFC are involved in computing and comparing value-related signals across choice situations (e.g., Rushworth et al. 2012), which makes it dubious that the indicated neural populations "contain all the neurons we require to extract [subjective values] for any object" (Glimcher 2009, 509); (2) activations in vmPFC/OFC contribute to *multiple functions* (e.g., Hunt et al. 2012), which complicates researchers' attempts to establish that such activations encode rewards' subjective values rather than contribute to some other function (e.g., O'Doherty 2014); (3) most proffered correspondences between individuals' choices and the neural substrates of reward valuation target *stimulus-bound* choices rather than the non-stimulus-bound choices individuals face in nonlaboratory settings (e.g., Ross 2011), and researchers lack convincing evidence that such correspondences generally hold also for non-stimulus-bound choices (e.g., Fumagalli 2016b); (4) many correspondences between individuals' choices and the neural substrates of reward valuation target relatively *coarse* characterizations of these substrates and break down when more fine-grained characterizations of such substrates are targeted (e.g., Padoa-Schioppa and Schoenbaum [2015], on the dissimilar functional roles of distinct subregions of OFC). As a result, such correspondences are more plausibly regarded as by-products of researchers' taxonomic practice rather than independent empirical evidence of a common neural scale.

3.4. Challenge from Conceptual Barriers. In the specialized philosophical literature, various conceptual differences between DU and EU have been identified (e.g., Fumagalli 2013; Okasha 2016). Here I identify and explicate three major conceptual differences between DU and NU: (1) DU represents *ordinal* relations between choice options rather than unique *cardinally* comparable magnitudes. For this reason, a neuron cannot be plausibly taken to represent a choice option's DU even if this neuron's firing rate is linearly proportional to the subjective value of such an option (e.g., Glimcher 2009, 506); (2) while DU is a mathematical index of agents' choices, neural measures of rewards' subjective values are empirically *dissociable* from choices and can *causally* influence choices (e.g., Ross 2011); (3) there is ongoing debate as to *how many* kinds of value-related *signals* are computed in the human neural architecture and how exactly these signals are integrated into overall measures of value (e.g., Padoa-Schioppa [2011], on action-related values and values computed in the space of goods irrespective of sensorimotor contingencies; Padoa-Schioppa and Schoenbaum [2015], on values learned and retrieved from memory and values computed at the time of choice). These conceptual differences do not prevent researchers from identifying statistically significant correlations between individuals' choices and specific areas' activation patterns (e.g., Kuorikoski and Marchionni 2016).

https://doi.org/10.1086/705454 Published online by Cambridge University Press

However, they cast serious doubt on leading authors' claim that rewards' subjective values "are exactly the kind of value representation posited by most decision theories . . . analogous to what economists call 'utilities'" (Kable and Glimcher 2009, 734).

4. True Utility: Normative Evaluation. In this section, I critically assess the potential of neuropsychological measures of utility to inform welfare analyses and policy evaluations. In doing so, I identify and rebut various objections that the proponents of TU may put forward to vindicate their reliance on neuropsychological measures of utility. Decision theorists commonly take individuals to be well off to the extent that their actual, informed, or ideal *preferences* are satisfied (e.g., Hausman and McPherson 2009). An individual's preferences count as satisfied when the state of affairs targeted by these preferences obtains. Knowing that one's preferences are satisfied in this sense may give one a feeling of satisfaction. Yet, preference satisfaction does not have to involve any feeling of satisfaction and may even obtain when one is unaware that her preferences are satisfied (e.g., Hausman 2010). For their part, the proponents of TU regard welfare as measurable on interpersonally comparable scales and rely on neuropsychological measures of utility as the relevant *indexes* of welfare (e.g., Kahneman and Krueger 2006). The idea is that neuropsychological measures of utility provide accurate and reliable indexes of welfare that enable policy makers to establish both "when people . . . choose what is best for them [and] what good policies follow" (Camerer 2008, 416; Bernheim and Rangel 2009, 85).

I already discussed in section 3 some of the limitations that plague current neuropsychological measures of utility. Below I assume, for the sake of argument, that researchers are able to overcome these limitations. I then argue that even if such limitations are overcome, informing welfare analyses and policy evaluations requires researchers to solve at least five major problems: (1) provide plausible and informative criteria to identify what welfare-relevant operations (if any) are tracked by neuropsychological measures of utility (*identification problem*); (2) make sure that no welfare-relevant operations are neglected by these measures (*exclusion problem*); (3) make sure that no welfare-irrelevant operations are tracked by such measures (*purification problem*); (4) determine what weights to ascribe to the welfare-relevant operations purportedly tracked by neuropsychological measures (*determination problem*); (5) aggregate the weighted measures of welfare-relevant operations thus determined into summative welfare indexes for the targeted individuals/population subsets (*aggregation problem*; e.g., Bernheim [2009], for a similar categorization).

A proponent of TU may object that analogous problems affect also choice-based indexes of welfare. However, the proponents of TU have hitherto failed to demonstrate that neuropsychological indexes solve such

https://doi.org/10.1086/705454 Published online by Cambridge University Press

problems better than choice-based indexes (e.g., Fumagalli 2013). Moreover, each of those problems poses formidable challenges to the proponents of TU. More specifically: (1) solving the identification problem requires researchers to make controversial *normative/axiological presuppositions* concerning the nature and the referents of welfare (e.g., Is welfare related more closely to EU or NU? Is welfare tracked more reliably by momentary, retrospective, or anticipated evaluations of individuals' hedonic experiences?); (2) solving the exclusion problem is complicated by the fact that even basic welfare-relevant operations (e.g., anticipating or reminiscing specific experiences) typically engage *several* neuropsychological processes besides those targeted by neuropsychological measures of utility (e.g., Muldoon and Bassett 2016); (3) solving the purification problem is complicated by the fact that the neuropsychological processes that purportedly contribute to welfare-relevant operations *also* contribute to several *non*-welfare-relevant operations, which facilitates spurious ascriptions of welfare relevance to observed neuropsychological activations (e.g., Kable and Levy 2015); (4) solving the determination problem requires researchers to establish what *normative/axiological significance* should be ascribed to neuropsychological measurements, which "are explicitly positive in nature" (Glimcher 2010, 412); (5) solving the aggregation problem is complicated by the wide *diversity* (and potential incommensurability) of welfare-relevant goods/experiences (e.g., Hausman 2012) and by the limited *interpersonal* comparability of neuropsychological measures (sec. 3).

A proponent of TU may grant that problems 1–5 challenge attempts to provide *perfectly* accurate and reliable neuropsychological indexes of welfare. However, she may object that neuropsychological indexes of welfare are *sufficiently* accurate and reliable to evaluate most policies' welfare implications (e.g., Kahneman 2000). There are at least three reasons to doubt this objection. First, neuropsychological indexes track a *relatively narrow subset* of the goods/experiences that on most conceptions of welfare count as welfare enhancing (e.g., Alexandrova [2018], on freedom and achievement). Second, the goods/experiences that on most conceptions of welfare count as welfare enhancing frequently contribute to welfare *irrespective* of what impact they happen to have on neuropsychological indexes (e.g., Hausman and McPherson 2009). And third, neuropsychological indexes track several goods/experiences that on most conceptions of welfare do *not* count as welfare enhancing (e.g., pleasures based on mistaken beliefs) and may even *detract* from welfare (e.g., pleasures derived from immoral activities).

A proponent of TU may object that neuropsychological indexes inform the evaluation of policies' welfare implications because they reliably track *fundamental constituents* of welfare (e.g., Dolan and Kahneman [2008], for statistically significant correlations between neuropsychological indexes and various indicators of health). However, it is dubious that neuropsychological

https://doi.org/10.1086/705454 Published online by Cambridge University Press

indexes *reliably* track many fundamental constituents of welfare (e.g., dopaminergic activations targeted by such indexes can be induced by addictive substances that demonstrably hamper individuals' health). Moreover, some goods/experiences may be fundamental constituents of welfare yet fall short of providing reliable (or even approximate) indicators of welfare. This, in turn, constrains neuropsychological indexes' potential to inform the evaluation of policies' welfare implications even under the assumption that these indexes reliably track fundamental constituents of welfare. In this respect, one may well claim that different indexes are complementary and that "evaluations of welfare will inevitably have to be informed by a combination of [choice-based and neuropsychological indexes] patched together in a fashion that depends on the specific context" (Loewenstein and Ubel 2008, 1795). Still, this claim does not provide informative insights regarding policies' welfare implications unless one supplements it with detailed indications regarding what indexes are considered, by means of what criteria one should select the relevant indexes, how these indexes are 'patched together', how one should resolve conflicts between different indexes, and how 'the specific context' determines the normative/axiological significance of the goods/experiences tracked by such indexes.

A proponent of TU may object that neuropsychological indexes enable researchers to evaluate policies' welfare implications by categorizing individuals into distinct behavioral/neuropsychological *types* (e.g., Kable and Levy [2015], for statistically significant correlations between individuals' risk propensities and morphometric indicators such as gray matter volume). Even so, these categorizations do not directly inform researchers' evaluations of policies' welfare implications. For showing that a policy is welfare enhancing for specific behavioral/neuropsychological types of individuals requires researchers to establish systematic correspondences between these behavioral/neuropsychological types and the policy's welfare implications for individuals of those types. And a given policy may have dissimilar welfare implications for individuals of the same behavioral/neuropsychological types (e.g., Fumagalli 2016a). Therefore, researchers' hypothesized ability to categorize individuals into distinct behavioral/neuropsychological types falls short of indicating that researchers can accurately evaluate policies' welfare implications for individuals of these types.

A proponent of TU may further object that neuropsychological indexes enable researchers to *resolve disagreements* concerning policies' welfare implications without having to make normative/axiological assumptions (e.g., Bernheim and Rangel [2009, 85], for the claim that neuropsychological indexes can "in principle [resolve] any serious disagreement" as to whether specific policies enhance individuals' welfare). Still, justifiably inferring that policies that enhance the value of specific neuropsychological indexes also enhance individuals' welfare requires researchers to specify how such

https://doi.org/10.1086/705454 Published online by Cambridge University Press

indexes map on individuals' welfare. In fact, even this specification does not per se enable researchers to establish "what good policies follow" from neuropsychological indexes (Camerer 2008, 416). For demonstrating that a particular policy is justified requires one to show not just that this policy enhances individuals' welfare but also that such a policy does not involve morally unacceptable violations of normatively significant values such as autonomy or consent (e.g., Fumagalli 2016a). In this respect, it is telling that distinct proponents of TU disagree regarding what policies (if any) are supported by the neuropsychological indexes they advocate (e.g., Kahneman and Sugden 2005).

5. Conclusion. The critique put forward in this article points to two major reasons to think that the search for the Holy Grail of 'true utility' should be abandoned. First, severe conceptual, epistemic, and evidential problems plague ongoing attempts to develop accurate and reliable neuropsychological measures of utility. And second, even if these problems are solved, neuropsychological measures of utility lack the potential to directly inform welfare analyses and policy evaluations. These two reasons do not prevent researchers from identifying systematic correspondences between individuals' choices and psychophysical magnitudes amenable to neuropsychological measurements. Yet, if the present critique is correct, such correspondences provide neither 'a neurobiological foundation for standard [decision] theory' nor a neuropsychological notion of 'true utility, which encapsulates well-being'.

REFERENCES

Alexandrova, A. 2018. "Can the Science of Well-Being Be Objective?" *British Journal for the Philosophy of Science* 69:421–45.
Alexandrova, A., and D. Haybron. 2016. "Is Construct Validity Valid?" *Philosophy of Science* 83:1098–109.
Angner, E. 2011. "Are Subjective Measures of Well-Being 'Direct'?" *Australasian Journal of Philosophy* 89:115–30.
Bernheim, D. 2009. "On the Potential of Neuroeconomics." *American Economic Journal* 1:1–41.
Bernheim, D., and A. Rangel. 2009. "Beyond Revealed Preference." *Quarterly Journal of Economics* 124:51–104.
Berridge, K., and J. O'Doherty. 2014. "From Expected Utility to Decision Utility." In *Neuroeconomics: Decision Making and the Brain*, 2nd ed., ed. Paul W. Glimcher and Ernst Fehr, 335–54. Amsterdam: Elsevier.
Broome, J. 1991. "Utility." *Economics and Philosophy* 7:1–12.
Camerer, C. 2008. "Neuroeconomics." *Neuron* 60:416–19.
Crisp, R. 2006. "Hedonism Reconsidered." *Philosophy and Phenomenological Research* 73:619–45.
Dietrich, F., and C. List. 2016. "Mentalism versus Behaviourism in Economics." *Economics and Philosophy* 32:249–81.
Dolan, P., and D. Kahneman. 2008. "Interpretations of Utility and Their Implications for the Valuation of Health." *Economic Journal* 118:215–34.
Dolan, P., and M. White. 2007. "How Can Measures of Subjective Well-Being Be Used to Inform Public Policy?" *Perspectives on Psychological Science* 2:71–85.

https://doi.org/10.1086/705454 Published online by Cambridge University Press

Fumagalli, R. 2013. "The Futile Search for True Utility." *Economics and Philosophy* 29:325–47.
———. 2016a. "Decision Sciences and the New Case for Paternalism." *Social Choice and Welfare* 47:459–80.
———. 2016b. "Five Theses on Neuroeconomics." *Journal of Economic Methodology* 23:77–96.
Glimcher, P. 2009. "Choice." In *Neuroeconomics: Decision Making and the Brain*, 2nd ed., ed. Paul W. Glimcher and Ernst Fehr, 503–21. Amsterdam: Elsevier.
———. 2010. *Foundations of Neuroeconomic Analysis*. Oxford: Oxford University Press.
Griffin, J. 1986. *Well-Being: Its Measure and Importance*. Oxford: Clarendon.
Hausman, D. 2010. "Hedonism and Welfare Economics." *Economics and Philosophy* 26:321–44.
———. 2012. *Preference, Value, Choice, and Welfare*. Cambridge: Cambridge University Press.
Hausman, D., and M. McPherson. 2009. "Preference Satisfaction and Welfare Economics." *Economics and Philosophy* 25:1–25.
Haybron, D. 2005. "On Being Happy or Unhappy." *Philosophy and Phenomenological Research* 71:287–317.
———. 2007. "Do We Know How Happy We Are?" *Nous* 41:394–428.
Heathwood, C. 2006. "Desire Satisfactionism and Hedonism." *Philosophical Studies* 128:539–63.
Hunt, L., N. Kolling, A. Soltani, M. Woolrich, M. Rushworth, and T. Behrens. 2012. "Mechanisms Underlying Cortical Activity during Value-Guided Choice." *Nature Neuroscience* 15:470–76.
Kable, J., and P. Glimcher. 2009. "The Neurobiology of Decision." *Neuron* 63:733–45.
Kable, J., and I. Levy. 2015. "Neural Markers of Individual Differences in Decision-Making." *Current Opinion in Behavioral Sciences* 5:100–107.
Kahneman, D. 2000. "Experienced Utility and Objective Happiness." In *Choices, Values, and Frames*, ed. Daniel Kahneman and Amos Tversky, chap. 37. Cambridge: Cambridge University Press.
Kahneman, D., and A. Krueger. 2006. "Developments in the Measurement of Subjective Well-being." *Journal of Economic Perspectives* 20:3–24.
Kahneman, D., and R. Sugden. 2005. "Experienced Utility as a Standard of Policy Evaluation." *Environmental and Resource Economics* 32:161–81.
Kahneman, D., P. Wakker, and R. Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics* 112:375–406.
Kuorikoski, J., and C. Marchionni. 2016. "Evidential Diversity and the Triangulation of Phenomena." *Philosophy of Science* 83:227–47.
Landreth, A., and J. Bickle. 2008. "Neuroeconomics, Neurophysiology and the Common Currency Hypothesis." *Economics and Philosophy* 24:419–29.
Levy, D., and P. Glimcher. 2012. "The Root of All Value." *Current Opinion in Neurobiology* 22:1027–38.
Loewenstein, G., and P. Ubel. 2008. "Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy." *Journal of Public Economics* 92:1795–810.
Louie, K., M. Khaw, and P. Glimcher. 2013. "Normalization Is a General Neural Mechanism for Context-Dependent Decision Making." *Proceedings of the National Academy of Sciences of the USA* 110:6139–44.
Moscati, I. 2018. *Measuring Utility*. Oxford: Oxford University Press.
Muldoon, S., and D. Bassett. 2016. "Network and Multilayer Network Approaches to Understanding Human Brain Dynamics." *Philosophy of Science* 83:710–20.
O'Doherty, J. 2014. "The Problem with Value." *Neuroscience and Biobehavioral Reviews* 43:259–68.
Okasha, S. 2016. "On the Interpretation of Decision Theory." *Economics and Philosophy* 32:1–25.
Padoa-Schioppa, C. 2011. "Neurobiology of Economic Choice." *Annual Review of Neuroscience* 34:333–59.
Padoa-Schioppa, C., and G. Schoenbaum. 2015. "Dialogue on Economic Choice, Learning Theory, and Neuronal Representations." *Current Opinion in Behavioral Sciences* 5:16–23.
Ross, D. 2008. "Two Styles of Neuroeconomics." *Economics and Philosophy* 24:473–83.
———. 2011. "Estranged Parents and a Schizophrenic Child." *Journal of Economic Methodology* 18:217–31.
Rushworth, M., N. Kolling, J. Sallet, and R. Mars. 2012. "Valuation and Decision-Making in Frontal Cortex." *Current Opinion in Neurobiology* 22:946–55.
Savage, L. 1954. *The Foundations of Statistics*. Hoboken, NJ: Wiley.
Van der Deijl, W. 2017. "Which Problem of Adaptation?" *Utilitas* 29:474–92.

https://doi.org/10.1086/705454 Published online by Cambridge University Press