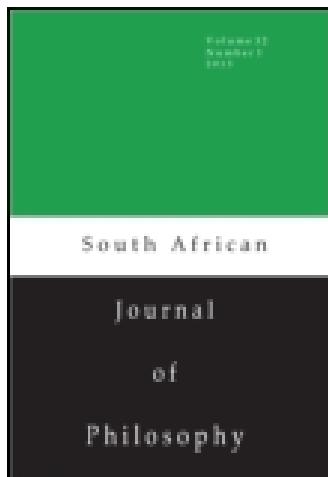


This article was downloaded by: [UNIVERSITY OF KWAZULU-NATAL]

On: 09 July 2014, At: 02:00

Publisher: Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## South African Journal of Philosophy

Publication details, including instructions for authors and subscription information:

<http://www.tandfonline.com/loi/rsph20>

### Philosophers should be interested in 'common currency' claims in the cognitive and behavioural sciences

David Spurrett<sup>a</sup>

<sup>a</sup> Philosophy, University of KwaZulu-Natal, Durban, South Africa

Published online: 04 Jul 2014.

To cite this article: David Spurrett (2014) Philosophers should be interested in 'common currency' claims in the cognitive and behavioural sciences, *South African Journal of Philosophy*, 33:2, 211-221

To link to this article: <http://dx.doi.org/10.1080/02580136.2014.923694>

PLEASE SCROLL DOWN FOR ARTICLE

Taylor & Francis makes every effort to ensure the accuracy of all the information (the "Content") contained in the publications on our platform. However, Taylor & Francis, our agents, and our licensors make no representations or warranties whatsoever as to the accuracy, completeness, or suitability for any purpose of the Content. Any opinions and views expressed in this publication are the opinions and views of the authors, and are not the views of or endorsed by Taylor & Francis. The accuracy of the Content should not be relied upon and should be independently verified with primary sources of information. Taylor and Francis shall not be liable for any losses, actions, claims, proceedings, demands, costs, expenses, damages, and other liabilities whatsoever or howsoever caused arising directly or indirectly in connection with, in relation to or arising out of the use of the Content.

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden. Terms & Conditions of access and use can be found at <http://www.tandfonline.com/page/terms-and-conditions>

# Philosophers should be interested in ‘common currency’ claims in the cognitive and behavioural sciences<sup>1</sup>

David Spurrett

Philosophy, University of KwaZulu-Natal, Durban, South Africa  
spurrett@ukzn.ac.za

A recurring claim in a number of behavioural, cognitive and neuro-scientific literatures is that there is, or must be, a unidimensional ‘common currency’ in which the values of different available options are represented. There is striking variety in the quantities or properties that have been proposed as determinants of the ordering in motivational strength. Among those seriously suggested are pain and pleasure, biological fitness, reward and reinforcement, and utility among economists, who have regimented the notion of utility in a variety of ways, some of them incompatible. This topic deserves philosophical attention for at least the following reasons. (1) Repeated invocation of the ‘common currency’ idiom isn’t merely terminological coincidence because most of the claims are *competing* explanations either of manifest pattern in choices, or of order in the processes producing choice. (2) We can’t suppose that the different currency claims within each area are compatible, because there are significant obstacles to identifying pairs of members of either the ‘pattern’ or ‘process’ group. (3) There are, finally, seriously opposed positions about the relationships (generally, and in specific cases including that of humans) between the pattern facts and the process facts. There are philosophical positions both favouring and opposing a common currency. But direct consideration of the abstract relationships between claims about common currencies across scientific settings, and the arguments for and against these claims, is relatively rare. There is, though, much of philosophical interest to be found here.

## Introduction

A recurring claim in a number of behavioural, cognitive and neuro-scientific literatures is that there is in fact, or must be, a unidimensional ‘common currency’ in which the values (actual, or expected) of different available options are represented. These currency metaphors partly succeed older, less overtly economic, yet similarly quantitative ways of speaking of decision-making as a kind of ‘weighing’, or of options attracting with varying ‘force’. What these images share is doing approximate justice to the pre-theoretical notion that the motivational import – to the choosing agent – of two options may be equal, or unequal, and when unequal that the relative difference between them can differ in magnitude.<sup>2</sup> When *all* options are taken to stand in these relations, whatever metaphors are invoked, we are on the way to a *common currency thesis*.

Theses about common currencies get fleshed out in a variety of ways as the pre-theoretical notion is regimented in specific scientific and philosophical contexts. Different kinds of argument and evidence are appropriate to defending the more specific formulations. Here are two fairly typical and widely cited examples:

In natural settings, the goals competing for behavior are complex, multidimensional objects and outcomes. Yet, for orderly choice to be possible, the utility of all competing

1 I’m grateful to David Papineau and Blaize Kaye for discussion of some of the issues in this paper, and to an anonymous reviewer for some very constructive and useful suggestions.

2 Some have an opposed hunch that there is no fact of the matter about the relative ranking of at least some options, perhaps because they are incommensurable (e.g. Williams 1981; Raz 1986).

resources must be represented on a single, common dimension (Shizgal & Conover 1996: pp. 37–38).

Any attempt to understand behavior in terms of the evolutionary advantage that it might confer has to find a “common currency” for comparing the costs and benefits of various alternative courses of action (McNamara and Houston 1986: p. 358).

These arguments are at least superficially similar enough that they might be taken as variants of a single argument, or as complementary arguments for the same claim. Both seem to demand a *single* currency, encouraging viewing them as complementary. Indeed, the prospect that they, or combinations of similar seeming arguments, might collectively justify multiple and fundamentally distinct currencies (as opposed to superficially distinct ones that turned out to be equivalent) is *prima facie* perplexing.

One claim I defend here is that this superficial appearance – of complementary arguments for one claim – is misleading. Significantly different claims about common currencies can and should be recognised. The remark by Shizgal and Conover occurs in discussion of experiments demonstrating that behaviour allocation in rats choosing between natural (gustatory)<sup>3</sup> rewards and brain stimulation reward is sensitive to opportunity cost in the unchosen reward, and that combination rewards with components in both modalities are (approximately) valued as *sums* of their components. The ‘common currency’ in question is reward magnitude for the individual rat. And Shizgal and Conover claim that this kind of pattern in allocation – the sensitivity to opportunity cost and disposition to value combination rewards as sums of their components that they call ‘orderly choice’ – is empirically observable and warrants an inference to a psychologically real common currency.

McNamara and Houston, on the other hand, are describing the project of behavioural ecology. They observe that this presupposes that behaviours have determinate fitness consequences, and seeks to determine those consequences and quantify their relative contributions to fitness. One of their central claims is, furthermore, that the main successes of behavioural ecology up to that time had focused on restricted currencies, such as net rate of energy intake in optimal foraging models (e.g. Pyke et al. 1977), and made relatively little headway with interpreting more comprehensive behavioural repertoires and life histories from the perspective of fitness. If such a comprehensive mapping could be found, though, it would express the relative values of all behaviours in an evolutionary common currency.

The two currency theses (Shizgal and Conover’s on the one hand, and McNamara and Houston’s on the other) are clearly not equivalent. Among the differences, one is centrally concerned with *fitness*, the other with something closer to *subjective (expected) utility*. Only one of them seems concerned with the choice-making process. Perhaps the repeated occurrence of ‘common currency’ talk is simply an uninteresting coincidence of terminology: researchers talking about distinct phenomena happen to use a superficially similar idiom. Perhaps, though, there is a tangle worth trying to unravel.

I favour the latter view, and in this brief paper offer a compressed survey of the terrain in which common currency theses occur, along with a preliminary defence of the claims that there is a tangle at all, and that the tangle deserves philosophical attention. Identifying all of the interestingly distinct claims about common currencies, and assessing the bodies of evidence relevant to each, would require more than a single paper. My aims here are more modest. In what follows I introduce a classification of common currency theses (‘Common currency claims distinguished’), and explain why some of the thus classified theses should be understood as competitors (‘Competitors, not complements’). Not all currency theses conflict, but where combinations of them are possibly complementary, we find additional disputes and disagreements (‘Inferences from ultimate to proximal currency theses’), and this set of issues is philosophically interesting (‘This is philosophically interesting’).

<sup>3</sup> The ‘natural’ rewards were modified in various ways in order to manipulate their similarities to and differences from brain stimulation reward (BSR). Shizgal and Conover make further inferences about the neural representation of value on the basis of how BSR and gustatory reinforcement respond differently to these manipulations.

### Common currency claims distinguished

Distinct claims about common currencies or value scales occur in various scientific settings. Generically and perforce loosely, because different more precise formulations pull in partly incompatible directions, a common currency is a unidimensional quantity that different options ‘have’ in varying amounts. A common currency is a representation of what an agent maximises, or, more weakly, it is some value ordering with which the agent’s behaviour is consistent.

The two main ways of clarifying currency claims are, first, to specify the characteristics of the value scale (whether ordinal, interval, etc.) and, second, to describe groupings *within* the range of currency theses and the types of explanations in which they feature. The latter task commands priority because the specific regimentation of the notion of scale that is appropriate depends on what purportedly is being measured or described and how. Here I focus mostly on classification, making only passing remarks about scale.<sup>4</sup>

Claims about common currencies are offered as explanations of one or both of two putative facts:

- the currency represents the fundamental principle in some pattern in the choices made by some agent, or
- the currency is a psychologically real characteristic of the processes by which choice is produced.

Claims about pattern and process can be combined, or made separately, as I explain below. The two are importantly distinct. To mark the distinction I follow, but slightly adapt, an established distinction, and refer to ‘ultimate’ and ‘proximal’ currencies. An *ultimate* currency is a construct in a descriptive or explanatory theory of the behaviour of some agent. A *proximal* currency, on the other hand, is supposed to play some role in the processes by which options are selected by an agent.

Any *single* ranking of options on a scale such that behaviour can be described as – perhaps approximately – consistent with that ordering counts as an instance of what I will call an ‘ultimate’ common currency. An ultimate currency relates values to options, or to what selecting those options achieves or perhaps has the function of achieving. It is easy to imagine *possible* instantiations of an ultimate currency (‘all of Jim’s actions are efficiently ordered to contribute to the greater glory of the Flying Spaghetti Monster’) but among the scientifically interesting forms of consistency, two families stand out. One of these relate to *fitness*, and the other to some of other form of *utility*.

I therefore say that an ultimate currency can be *evolutionary* or *subjective*. Here I’m knowingly, albeit slightly, departing from standard usage, insofar as the usual way of distinguishing proximal from ultimate has the latter taken to be synonymous with fitness promoting. The standard sense is partly preserved here, because what I call an evolutionary ultimate currency attributes values that are a function of contribution to fitness. A guiding presumption of behavioural ecology is that behavioural dispositions make contributions to fitness, and that to the extent that the dispositions have a heritable basis, selection will tend to drive them towards making (constrained) optimal contributions. The statement by McNamara and Houston briefly discussed above is an exemplary assertion about an evolutionary ultimate currency.

There are compelling reasons for thinking that natural selection will operate on at least some behavioural tendencies, and so provide a strong general justification for the project of behavioural ecology. Nonetheless, the question of whether the behaviour allocation of the individuals in any particular species does indeed tend to optimise fitness (or would have in historical selective environments) is an empirical one. Behavioural ecologists have studied many species and types of behaviour and achieved striking successes in restricted domains such as foraging and mate selection. These successes have often relied on focusing on a simplified and more empirically tractable currency such as net rate of calorie intake (in the case of foraging), or the health of the selected mate (in mate selection). These proxy currencies plausibly contribute to fitness. Even so, success in relating behaviour patterns to the proxies falls short of establishing relationships between *all* behavioural dispositions and overall fitness in any species. Two of the most serious shortfalls are in the area of relative allocation between significantly different modalities (such as

<sup>4</sup> For a classic discussion of the formal properties of some notions of preference see Luce and Suppes (1965), and for more general remarks on scales and measurement see Suppes and Zinnes (1963).

calorie intake versus pursuit of mating opportunities versus predator avoidance) and variation in allocation over extended periods of time, including the full life-history.

A subjective ultimate currency, in contrast, attributes values that are a function of the revealed preferences or inferred utilities of the individual agent, without requiring any relationship to fitness. The paradigmatic sciences of subjective ultimate currencies are microeconomics (with variously regimented notions of utility functions revealed through consumption) and behavioural psychology (where strength of reinforcement is defined in terms of effect on patterns of behaviour allocation). Evolutionary ultimate currencies, then, have stronger empirical conditions than subjective ones, because the latter require ‘mere’ consistency in behaviour, whereas the former require consistency in contributing to *fitness*.

The distinction between evolutionary and subjective ultimate currencies, as I’ve sketched it here, conceals considerable technical detail. There are competing fitness concepts, debates over the level at which selection operates, and differing positions over the correct individuation principles for genes, genomes, species and other relevant categories. There are also competing utility concepts, offering different explanations of the same empirical data. (For example, some make randomness a feature of the utility representation itself, while others hypothesise randomness in the form of ‘trembling hands’ in the process of expressing the preferences.) Surveying this varied terrain is beyond the scope of a brief paper, but one further complication concerning two different ways of understanding ‘consistency’ must be noted.

Economists tend to favour a strict notion of consistency because they recognise that predictable inconsistency makes agents vulnerable to systematic exploitation. Perhaps the best-known *specific* version of this worry is the argument that an agent with cyclical preferences<sup>5</sup> could be used as a ‘money pump’ (see Ross 2005: ch. 5). Such an agent would freely pay for a series of trades that eventually left it with no effective money or stock. This convinces most economists that viable agency requires acyclicity (‘transitivity’), among other criteria for consistency. Behavioural psychologists, on the other hand, seek phenomenological fits of functions to empirical data, and have recognised patterns in behaviour objectionable to microeconomists. The clearest example of this is their accepting that the generalised Matching Law applies to delayed rewards (Chung and Herrnstein 1967). This implies that rewards are valued in inverse proportion to delay (i.e. by a hyperbolic function), and the relative desirability of incentives at different times can change simply with the passage of time. Given appropriate repeated choices, cyclical preferences follow.<sup>6</sup> That should mean vulnerability to money pumping, and non-behavioural economists have favoured delay discounting according to exponential functions largely for this reason (again, see Ross 2005: ch. 5). The hyperbolic delay discounter is inconsistent insofar as she *temporarily* prefers smaller rewards that are immanently available. But to the behavioural psychologist her choices are all *consistently* reward seeking, in the sense that once some empirical parameters have been determined, relative rates of behaviour are predictable.

The other main role for currency claims is to explain the *processes* by which options are selected. Such theses assert the existence of a ‘proximal’ common currency. Here are some examples. A realist about desires who holds that for any pair of desires there is a fact of the matter about whether they are of equal strength or one is stronger, is committed to a common currency thesis. Shizgal and Conover’s inference regarding a ‘single, common dimension’ is a scientifically motivated claim about the cognitive requirements of producing ‘orderly choice’. A conventional chess-playing programme generating a tree of possible game-states, then attaching values to them on the basis of some algorithm, in order to select a best move (among the options explored in the available time) implements a common currency. Finally, the leading current scientific research programme focused on a proximal common currency is neuroeconomics, which seeks to determine how utilities are represented in brains, and how these representations are processed in choosing and learning (e.g. Levy & Glimcher 2012).

A claim asserting a *proximal* common currency, then, is not *any* assertion about mechanisms producing behaviour, even consistent behaviour. A proximal common currency is, rather, a

<sup>5</sup> In the sense of preferring bundle *a* to *b*, bundle *b* to *c*, and *also* bundle *c* to *a*.

<sup>6</sup> Nobody has done more than George Ainslie when it comes to thinking seriously about what it would mean for humans if we value rewards approximately in inverse proportion to their delay (see Ainslie 1992, 2001).

structured and integrated set of states that represents values, in (at least) the weak sense that it carries information about a mapping between actions and/or states of the world and values in an ultimate currency, and this information is consumed by the cognitive processes selecting actions.

Not everyone who thinks that behaviour is consistent, and that there is a mechanical process explaining behaviour selection, is committed to a proximal common currency. This is because not all views about how behaviour is caused involve representations, including value representations. One reason is respect for ‘Lloyd Morgan’s canon’:

In no case may we interpret an action as the outcome of the exercise of a higher psychical faculty, if it can be interpreted as the outcome of one which stands lower on the psychological scale (Morgan 1894: p. 53).

Giant sea slugs (*Pleurobranchaea*), for example, are carnivorous and typically eat any animal matter they run into, ‘including other sea-slugs and their eggs’ (Manning & Dawkins 1998: p. 226). They do not, however, eat their own eggs during egg laying. This disposition is obviously fitness enhancing: creatures that routinely consume their own offspring leave fewer viable descendants. The mechanism that stops sea slugs from eating their own eggs, though, while it could be regarded as in some very broad sense ‘cognitive’ is not one in which values are represented (e.g. Godfrey-Smith 2002).<sup>7</sup> When sea slugs lay eggs, they release a hormone that inhibits movement of the mouth (Davis et al. 1977). This simple override exemplifies the ‘subsumption’ relationship between control layers championed by Rodney Brooks (e.g. 1991), who famously maintained that ‘intelligent’ behaviour could be achieved in the absence of representation.

There are three generic and complementary forms of argument in favour of a proximal currency. The first is an inference to best explanation where an internal currency representation is hypothesised on the basis of observed order in behaviour (e.g. Shizgal & Conover 1996). The second considers abstract features of a control system, and argues that control bottlenecks make it more likely that some unified value representation is playing a role in determining how the degrees of freedom in the system are allocated. This line of thinking, going back at least to Sherrington, is sometimes called the ‘final common path’ argument (e.g. McFarland & Sibley 1975). Finally, there are cases where it is claimed that a proximal currency has effectively been observed in action, through study of the behaviour control system at work (e.g. in neuroeconomics). All three are, furthermore, contested.

Scientifically interesting talk about common currencies can almost entirely be classified into arguments in favour of or against one kind of currency (proximal, evolutionary ultimate or subjective ultimate), and in favour of or against specific inferences from ultimate to proximal currencies. In the following section I argue briefly that the currency theses and related arguments in current science warrant philosophical examination.

### **Competitors, not complements**

I discuss attempted inferences *between* distinct currency thesis types, especially from ultimate to proximal, in the next section. Here I consider relationships between ultimate (pattern) and proximal (process) common currency theses. The number of at least superficially distinct common currency theses is striking. As we’ve seen, behavioural ecologists seek to relate patterns of behaviour to their contribution to *fitness*. Early utilitarians, and some contemporary theorists about pain and analgesia claim that *pain* and *pleasure* provide a common scale (e.g. Bentham 1789; Cabanac 1971; Leknes & Tracey 2010). Behaviourist psychologists refer to *reward* or *reinforcement*, while contemporary economists are more likely to advance a currency thesis about *utility*. In addition, some of the central concepts that characterise the scales have been theoretically elaborated in varied ways. This is most striking in the case of utility, both within economics, including behavioural economics, and in neighbouring areas such as decision theory.

This mere proliferation is not *intrinsically* interesting. Perhaps the various theses are complementary. In the case of ultimate currencies, there is, *prima facie*, a prospect of pluralistic harmony, because such

<sup>7</sup> I follow Godfrey-Smith’s suggested policy regarding understanding ‘cognitive’ inclusively to embrace the processes controlling behaviour, even in systems (e.g. some plants and fungi) that lack central nervous systems.

theses are claims about pattern, and more than one kind of pattern can be manifest in the same data. It is, though, an observation familiar to the point of banality that subjective preferences – for example, for sex with contraceptives – don't always coincide with what is fitness promoting. More generally, and not only in humans, what is fitness promoting is not always motivating and what is motivating is not always fitness promoting. The fact that animal subjects in behavioural experiments would work for non-nutritive sweeteners, for example, was recognised decades ago as an obstacle to identifying reward with evolutionary interest (see e.g. Rachlin 1991: ch. 3).

It is not unusual, in fact, to find biologists expressing scepticism about the plausibility of a currency thesis outside biology, even while being optimistic within it. Maynard-Smith, for example, suggested that ‘it has turned out that game theory [...] is more readily applied to biology than to the field of economic behavior for which it was originally designed’. And part of his justification for this is that he finds utility to be ‘a somewhat artificial and uncomfortable concept’, whereas in biology ‘Darwinian fitness provides a natural and genuinely one-dimensional scale’ (Maynard Smith 1982, quoted in Glimcher 2002: p. 323).

Many economists and behaviourists dispute this, on the grounds that the utility concept is in much better shape than Maynard-Smith allows. My point is not to endorse either view, but to observe that the lack of consensus suggests work for applied philosophy of science. One impediment to taking sides is, in any event, that economists disagree with each other. There are competing formulations of utility and disputes over which do best justice to the evidence, or best suit what theoretical purposes. I noted above that economists’ concern with consistency leads them to favour exponential delay discount functions, while behavioural psychologists are more inclined towards hyperbola-like functions. This is just one instance of a wider pattern. To give one more example, defenders of prospect theory (e.g. Kahneman & Tversky 1979) claim that their model, built to account for phenomena including apparent violations of the independence axiom, and asymmetries of risk sensitivity in gains and losses, does better justice to data about real human choices. Sceptics, though, express frustration that the sheer number of free parameters in prospect theory undermines the empirical value of such fits (see Glimcher 2011: ch. 5).

A further incompatibility between the currency claims of behavioural ecologists and psychologists arises because they seek to explain somewhat different things, even while both calling it ‘behaviour’. We’ve noted the giant sea slug’s apparent restraint over eating its own eggs. My point there was to explain why behavioural ecologists are often carefully agnostic about proximal common currencies. Similar cases can also be used to make a different point. For example, a textbook case in behavioural ecology is clutch sizes in oviparous species (on birds, see e.g. Lack 1966). There’s substantial evidence that clutch sizes in many bird species are close to what would maximise lifetime reproductive success of parent birds, whose situation involves trade-offs between current and future clutches, and between members of individual clutches, given that larger (better fed) young generally do better.

Some of the determinants of clutch size plausibly respond only to phylogenetic rewards, where successful offspring are the payoffs in repeated games with genotypes as the players or strategies. There is little reason to think, though, that clutch size is modifiable by reward or punishment directed at the individual bird, any more than one might bribe a giant sea slug into eating its own eggs. But responsiveness to reinforcement is what would make it behaviour for a behavioural psychologist or economist.

What about proximal common currency theses? It seems that multiple *distinct* proximal theses raise the prospect of over-determination. That is, if the mechanisms of choice include more than one structured and integrated set of states representing values and involved in causing behaviour, then we’d have more explanation than we needed (‘I did it because it had the most expected utility, and it was pleasurable and I preferred it...’). This worry could be dissolved if it turned out that the different proximal currency theses expressed substantially the same claim, so that what some kind of agent wanted (or desired) was also what it liked (or gave it pleasure) *and* what promoted its fitness.

The prospects for such an outcome, though, are slim. Among other reasons, behavioural economists have drawn attention to various examples of apparently motivated behaviour involving considerable pain, for example mountaineering (e.g. Loewenstein 1999). Brain scientists studying

the learning, enjoying and choosing brain now mostly maintain that ‘wanting’ and ‘liking’ (including enjoying and suffering) are neurally and functionally dissociated and that only the former promises to provide a perspective from which behaviour can be understood as consistent with some value representation (e.g. Berridge 2004). These arguments conflict directly with the claim defended by some researchers on hedonic experience that pain and pleasure provide a common currency for behaviour selection (e.g. Leknes and Tracey 2010). Whatever consensus eventually emerges, it seems clear that no more than one of the current proximal common currency theses can be correct.

### **Inferences from ultimate to proximal currency theses**

Within categories – ultimate and proximal – different common currency theses, then, are competitors, but there are obvious ways that one thesis of each type could be complementary. If the order in behaviour warrants attributing an ultimate common currency, then this might justify hypothesising a proximal currency involved in producing the pattern. Conversely, if we had grounds to suppose that some agent’s cognitive processes implemented a proximal common currency, we might expect its behaviour to exhibit pattern consistent with an ultimate currency. Versions of both inferences have been defended, and contested.

The range of options for relations between proximal and ultimate currency theses is approximately analogous with those regarding the status of folk psychological kinds in the philosophy of cognitive science. (Analogous might not be the best term, because folk psychology *includes* desires and beliefs. But most of the debates over folk psychology focused on epistemic states – like belief, perception, and memory – to the relative neglect of motivation states, like desires.) The major options are realism (beliefs are scientifically respectable), eliminativism (cognitive processes include nothing sufficiently like beliefs for belief talk to pass scientific muster) and attributionism, for example Dennett’s intentional stance (the conditions for belief attribution exclusively concern pattern in behaviour). Approximately corresponding to this, we find the following positions regarding proximal common currencies.

First, there are realists committed to the existence of cognitive states standing for the motivational strength of different courses of action in humans, and at least some other agents. Examples include typical realists such as Fodor (1983), some scientists specialising in hedonic experience (e.g. Cabanac 1971) and neuroeconomists identifying preferences with brain states. A paradigmatic example of an inference from pattern to proximal representation is, of course, the syllogism stated by Shizgal & Conover (1996) quoted above, linking ‘orderly choice’ with the requirement of ‘representation’ on a common scale.

Second, are eliminativist positions, denying – to varying degrees – the reality or necessity of cognitive states corresponding to desires. Brooks (1991), for example, argues against the need for representations of any kind to produce ‘intelligence’. For him the world is ‘its own best representation’. Brooks directs most of his fire against representations in the sense of models of the external environment, but his arguments clearly imply rejection of representations of values. Brooks’ work remains an inspiration to some philosophers. Clark (e.g. 1997: ch. 9) claims that pattern in human economic choices depends heavily on highly scaffolded choice environments. Sterelny (2003), on the other hand, suggests that the cognitive implementation of preferences in humans is ‘incomplete’, and that pattern in the behaviour of most non-humans can be explained without hypothesising preference like states as parts of their cognitive architectures.<sup>8</sup>

Finally, we find attributionists, emphasising the view that the work of those who trade in ultimate common currencies is independent from proximal considerations. The exemplary contemporary attributionist about folk psychology is Dennett, and scientific attributionists share behaviourist inspiration. Microeconomists mostly regiment their notion of utility so that it makes no psychological or hedonic commitments, in favour of specifying different degrees of consistency that can be empirically manifest in behaviour (e.g. Samuelson 1938; see Ross 2005). Behaviourist psychologists are similarly suspicious of – or hostile to – claims about hedonic experience, and favour more empirically tractable notions such as reinforcement and reward (Thorndike 1927).

8 See Spurrett (unpublished).

The task of assessing the strengths and weaknesses of realism, eliminativism and attributionism in the case of the more epistemic aspects of folk psychology (believing, perceiving, remembering, inferring...) attracted wide and deep philosophical activity over several decades. Over this period considerably less attention was paid to the motivational aspects of folk psychology (wanting, desiring and choosing). In the final section of this brief survey I attempt to say a little more about why these topics deserve more attention.

### This is philosophically interesting

Suppose that I've convinced a reader that there is a tangle of variously competing and (possibly) complementary theses about common currencies across a number of cognitive and behavioural sciences. Why think that this tangle is of any *philosophical* significance? Among the reasons I could offer, let me single out three.

First, there's an image of philosophers associated with Locke as a kind of conceptual janitor for empirical science (Locke's image, in the Epistle to the Reader of *An Essay Concerning Human Understanding*, was of an 'under-labourer'). Such conceptual work is more clearly indicated, perhaps, in cases where there is conflict within and between empirical sciences. To illustrate this, consider the different situations of behaviourism in psychology, and economics. Behaviourist psychologists thought of themselves, to state the obvious, as being in the psychology business, although driven by a distinctive vision of what it meant to do that business in a methodologically serious way. Their rejection of introspection was methodological, motivated by the consideration that scientific evidence should be inter-subjectively available. Consistently with that commitment, some more recent behaviourists have argued that reports of subjective experience can be a kind of data (e.g. Dennett's [1991] 'heterophenomenology'). Furthermore, advances in measuring devices have made previously unobservable brain processes amenable to empirical study, and so allowed different kinds of data to pass muster by behaviourist lights. Neuroeconomics now promises to provide an empirical basis for a theory of reinforcement or reward that explains observed pattern in what behaviour can be reinforced, and to what degree. While some individual behaviourists might reject specific theoretical suggestions in this area, the lack of *some* theory commanding wide acceptance has long been recognised.

Given the noted inspiration behaviourism provided for contemporary formulations of the concept of utility in economics, one might expect similar enthusiasm for neuroeconomics among working economists. Instead, one finds lively and sometimes intensely polarised debate. This should not be especially surprising. Unlike behaviourists, who had a distinctive view about how psychology should be done, economists (who used to view their discipline as at least closely linked to psychology – e.g. Jevons 1871) view their discipline as having made progress by getting out of the psychology business entirely. While some behavioural economists have more recently sought to move in the opposite direction, they by no means dominate the profession.

This helps explain some of the response to an enthusiastic early manifesto for neuroeconomics (Camerer et al. 2005). Here, and in related publications around the time, it was claimed that a new and importantly psychological research programme held out the hope of providing new 'foundations' for microeconomics. While the experimental techniques would have amazed Mill, Bentham, Jevons and their colleagues, the generic view of the relationship between economics and psychology probably would not. But 2005 is not the eighteenth century, and one of the most spirited responses to neuroeconomics was called 'The case for mindless economics', and argued forcefully (whether or not correctly) that economics and psychology were separated by a 'logical' gulf, that they addressed 'different questions', used 'different abstractions' and relied on 'different types of empirical evidence', with the consequence that neuroscience could have no bearing on economics whatsoever (Gul & Pesendorfer 2005). This dispute is, among other things, a special case of wider and older debates about the status of folk-psychology and the prospects for reductionism. Economists – as scientists of something related to a folk notion of desire, even if utility is drained of most of the psychological content of the folk notion – are divided, *inter alia*, into those emphasising the autonomy of their enterprise, and those welcoming closer links with psychology and

neuroscience. A striking feature of much of these debates is repeated invocation of the notion of a common currency (e.g. Glimcher 2011).

Second, the specific issues relating to the status of folk-psychological explanation that arise in the case of motivation, and regarding the credibility of specific common currency theses, are not simply repetitions or generalisations of arguments about beliefs or representations. Were that the case, there would still be philosophical work to do, because applying whatever lessons emerged from the earlier debates would depend partly on details that science has revealed in the domain of preference and decision. Brooks' general case against representation, for example, is widely recognised as failing in the case of belief except for very simple control systems, and may also fail in the case of preference, but perhaps not for the same reasons.

But there are also important ways in which things are *different* here. The 'old' debate over folk psychology was, as noted, disproportionately focused on epistemic states such as belief, memory and perception, and so unsurprisingly much of the traffic with philosophy was in areas relating to epistemology and language. But those aren't the most obvious or promising sources of exchange in the case of desires. Instead, in fact, the fruitful exchange is more likely to be in areas engaged with thinking about practical reasoning and decision, and thinking about *value*.

Third, and finally, philosophers already defend, assume or deny common currency theses or positions implying such commitments. A common opening move among decision theorists, for example, is to assume that values – whatever their 'content' – can be represented with the real numbers (e.g. Briggs 2010). More generally, what one might call 'empiricist' views of motivation suppose that there is a single marketplace of desires competing on the basis of their strength, however that strength is understood. On the other hand more 'rationalist' views favour the view that in at least some cases motivation is not simply a matter of relative strength of desires, but is rather rule-based, or otherwise based on reason in *opposition* to desire.

Whether or not directly defending rationalism in this very loose sense, a number of lines of philosophical thinking oppose, directly or by implication, the thought that all options are valued on a single scale. Among the examples of this are: arguments that some options or values might be incommensurable, in the sense that there is no fact of the matter about which is more valuable, or whether they are equal (e.g. Raz 1986; Williams 1981); positions maintaining that some values are ranked 'lexically' in the sense that any amount of some – no matter how small – is worth more than any amount of others, no matter how large (e.g. Rawls 1971). In addition, in some recent empirical literature we find arguments that some values are 'protected' or 'sacred' and somehow isolated from trade-offs with others (e.g. Baron & Spranca 1997; Atran & Ginges 2012).

A number of long-standing philosophical questions about desire, decision, preference and practical reasoning, then, (and this survey is woefully short of comprehensive) are partly questions about the degree to which motivational systems are, or can be, unified, and about the kinds of ranking and hierarchy systems of value or motivational strength can or should have. Common currency theses – whether affirmed or denied – provide a useful level of abstraction that permits comparison of otherwise mutually isolated theses and theories, and a substantial and tangled scientific terrain is debating aspects of these very questions.

## References

- Ainslie, G. 1992. *Picoeconomics: the Strategic Interaction of Successive Motivational States within the Person*. Cambridge: Cambridge University Press.
- Ainslie, G. 2001. *Breakdown of Will*. New York: Cambridge University Press.
- Atran, S., Ginges, J. 2012. 'Religious and sacred imperatives in human conflict', *Science* 336(855), 855–857.
- Baron, J., Spranca, M. 1997. 'Protected values', *Organizational Behavior and Human Decision Processes* 70(1), 1–16.
- Bentham, J. 1789. *Introduction to the Principles of Morals and Legislation* (1823 edn). Oxford: Clarendon Press.
- Berridge, K.C. 2004. 'Motivation concepts in behavioral neuroscience', *Physiology and Behavior* 81(2), 179–209.

- Briggs, R. 2010. 'Decision-theoretic paradoxes as voting paradoxes', *Philosophical Review* 119(1), 1–30.
- Brooks, R.A. 1991. 'Intelligence without representation', *Artificial Intelligence* 47(1–3), 139–159.
- Cabanac, M. 1971. 'Physiological role of pleasure', *Science* 173(4002), 1103–1107.
- Camerer, C.F., Loewenstein, G., Prelec, D. 2005. 'Neuroeconomics: how neuroscience can inform economics', *Journal of Economic Literature* 34(1), 9–64.
- Chung, S., Herrnstein, R.J. 1967. 'Choice and delay of reinforcement', *Journal of the Experimental Analysis of Behavior* 10(1), 67–74.
- Clark, A. 1997. *Being There*. Cambridge, Massachusetts: MIT Press.
- Davis, W.J., Mpitsos, G.J., Pinneo, J.M., Ram, J.L. 1977. 'Modification of the behavioral hierarchy of *Pleurobranchaea*. I. Satiation and feeding motivation', *Journal of Computational Physiology* 117(1), 99–125.
- Dennett, D. 1991. *Consciousness Explained*. Boston: Little, Brown.
- Fodor, J.A. 1983. *The Modularity of Mind*. Cambridge, Massachusetts: MIT Press.
- Glimcher, P. 2002. 'Decisions, decisions, decisions: choosing a biological science of choice', *Neuron* 36(2), 323–332.
- Glimcher, P. 2011. *Foundations of Neuroeconomic Analysis*. Oxford: Oxford University Press.
- Godfrey-Smith, P. 2002. 'Environmental complexity and the evolution of cognition', in Sternberg, R., Kaufman, J. (eds), *The Evolution of Intelligence*. Mahwah: Lawrence Erlbaum, pp. 233–249.
- Gul, F., Pesendorfer, W. 2008. 'The case for mindless economics', in Caplin, A., Schotter, A. (eds), *Foundations of Positive and Normative Economics*. New York: Oxford University Press, pp. 3–42.
- Jevons, W.S. 1871. *The Theory of Political Economy*. London: Macmillan & Co.
- Kahneman, D., Tversky, A. 1979. 'Prospect theory: an analysis of decision under risk', *Econometrica* 47(2), 263–291.
- Lack, D. 1966. *Population Studies of Birds*. Oxford: Oxford University Press.
- Leknes, S., Tracey, I. 2010. 'Pain and pleasure: masters of mankind', in Kringselbach, M.L., Berridge, K.C. (eds), *Pleasures of the Brain*. Oxford: Oxford University Press, pp. 320–336.
- Levy, D.J., Glimcher, P.W. 2012. 'The root of all value: a neural common currency for choice', *Current Opinion in Neurobiology* 22(6), 1027–1038.
- Loewenstein, G. 1999. 'Because it is there: the challenge of mountaineering ... for utility theory', *Kyklos* 52(3), 315–343.
- Luce, R.D., Suppes, P. 1965. 'Preference, utility and subjective probability', in Luce, R.D., Bush, R.R., Galanter, E.H. (eds), *Handbook of Mathematical Psychology*, vol 3. New York: Wiley, pp. 249–410.
- Manning, A., Dawkins, M.S. 1998. *An Introduction to Animal Behaviour* (3rd edn). Cambridge: Cambridge University Press.
- Maynard-Smith, J. 1982. *Evolution and the Theory of Games*. Cambridge: Cambridge University Press.
- McFarland, D.J., Sibley, R.M. 1971. 'The behavioural final common path', *Philosophical Transactions of the Royal Society of London B: Biological Sciences* 270(907), pp. 265–293.
- McNamara, J.M., Houston, A.I. 1986. 'The common currency for behavioral decisions', *American Naturalist* 127(3), 358–378.
- Morgan, C.L. 1894. *An Introduction to Comparative Psychology*. London: W. Scott.
- Pyke, G.H., Pulliam, H.R., Charnov, E.L. 1977. 'Optimal foraging: a selective review of theory and tests', *Quarterly Review of Biology* 52(2), 137–154.
- Rachlin, H. 1991. *Introduction to Modern Behaviorism* (3rd edn). New York: Freeman.
- Rawls, J. 1971. *A Theory of Justice*. Cambridge: Harvard University Press.
- Raz, J. 1986. *The Morality of Freedom*. Oxford: Clarendon Press.
- Ross, D. 2005. *Economic Theory and Cognitive Science, vol. 1: Microexplanation*. Cambridge, Massachusetts: MIT Press.
- Samuelson, P.A. 1938. 'A note on the pure theory of consumer's behaviour', *Economica, New Series* 5(17), 61–71.

- Shizgal, P., Conover, K. 1996. ‘On the neural computation of utility’, *Current Directions in Psychological Science* 5(2), 37–43.
- Spurrett, D. Unpublished. The Descent of Preference Reconsidered.
- Sterelny, K. 2003. *Thought in a Hostile World*. Oxford: Blackwell.
- Suppes, P., Zinnes, J.L. 1963. ‘Basic measurement theory’, in Luce, R.D., Bush, R.R., Galanter, E. (eds), *Handbook of Mathematical Psychology*, vol. 1. New York: Wiley, pp. 1–76.
- Thorndike, E.L. 1927. ‘The law of effect’, *American Journal of Psychology* 39(1–4), 212–222.
- Williams, B. 1981. *Moral Luck*. Cambridge: Cambridge University Press.