

**Economists' Odd Stand on the Positive-Normative Distinction:
A Behavioral Economics View**

John B. Davis

Marquette University and University of Amsterdam

Abstract: This chapter examines economists' indefensible attachment to the positive-normative distinction, and suggests a behavioral economics explanation of their behavior on the subject. It reviews the origins of the distinction in Hume's guillotine and logical positivism, and shows how they form the basis for Robbins' understanding of value neutrality. It connects philosophers' rejection of logical positivism to their rejection of the positive-normative distinction, explains and modifies Putnam's view of fact-value entanglement, and identifies four main ethical value judgments that contemporary economists employ. The behavioral explanation of economists' denial of these value judgments emphasizes loss aversion and economists' social identity as economists.

Keywords: value neutrality, Robbins, Putnam, fact-value entanglement, loss aversion

JEL codes: A13, B41, D03

1. *Economists' Odd Stand on the Positive-Normative Distinction*

There has been much change in recent economics ranging from theoretical innovation to new methods of research (Davis, 2008), indeed so much so that there seems to be less and less consensus today over how economics ought to be defined in terms of scope and essential content. I suggest, however, that one important unchanging aspect of economics remains especially central to its identity as a distinct discipline, if not to its definition: economists' position on the positive-normative distinction. Here we see little disagreement among economists and little change in what economists think today from what they thought from when Lionel Robbins asserted that interpersonal comparisons of utility implicitly involved ethical claims and should be excluded from economics. Further, their view on the subject distinguishes them from most other social scientists who also differentiate the positive and the normative but generally do so in a far less severe way than economists. Why has this part of the identity of economics so strongly resisted change when so much else in economics has undergone change? This question is even more perplexing in light of the fact that economists' position on the positive-normative distinction was rejected a half century ago by those who are far more expert on the subject, namely philosophers, who saw the positive-normative distinction as a naïve vestige of prewar logical positivism, and who like other social scientists have since worked with a far more nuanced view of the positive and the normative. Economists' view of the positive-normative distinction is thus an odd one: their position is basically indefensible but they remain steadfast in their conviction that it is correct in a time when they have changed their views of many other matters in economics. Thus amidst all this change, the positive-normative distinction is arguably now even more important to the identity of economics. Such a state of affairs calls out for more attention and investigation, despite the fact that as an official non-issue in economics few seem to see much reason to address it.

This chapter examines economists' odd stand on the positive-normative distinction. It first reviews its origins in logical positivism and the now well-established reasons that the standard view is indefensible to show the exceptional status of the issue and its special importance to economics. But it is obviously not enough to review this history and these reasons since they have had little influence for more than a half century, and so are apparently insufficient to explain why economists hold so steadfastly to their formulaic view of the positive and the normative. Thus what the chapter goes on to do is set out how economists' view of the positive-normative distinction might be explained by developing a behavioral economics account of economists' behavior on the subject. That is, just as behavioral economics is used to explain the choices of economics agents, so it can be used to explain the choices of economists as economic agents. Of course there are different ways this explanation could be developed because the various psychological heuristics which psychologists and behavioral economists have investigated are quite versatile in their application and interpretation. However, to provide what would count as a sort of mainstream behavioral economics explanation I rely on the most well demonstrated heuristics and widely employed types of arguments, namely,

the basic idea of reference dependence, and one of the main claims that prospect theory makes about human psychology, specifically, that people are loss-averse, or that a “salient characteristic of attitudes to changes in welfare is that losses loom larger than gains” (Kahneman and Tversky, 1979, p. 279). At the same time, the behavioral economics explanation of economists’ behavior offered here is also modified from what prevails in many such explanations to reflect the disciplinary dimension of economists’ behavior. Economists know they are members of the economics profession, and they socially identify with it. This needs to be part of what is involved in explaining how they interpret their choices. Behavioral economics explanations often ignore social identity effects because social identity is not involved in experiments that focus primarily on how psychological heuristics operate in general. But when we investigate decision makers in an explicit professional or disciplinary context, social identity should not be ignored. Thus in the discussion here the goal is to explain the behavior of a representative type of decision-maker: economists who see themselves as economists. This characterization clearly may not fit all individual economists.

Section 2 traces the standard view of the positive-normative distinction back to positions that the logical positivists developed in the 1930s regarding facts and values, which drew on David Hume’s earlier empiricist thinking and claims about the difference between ‘is’ and ‘ought’ statements. Lionel Robbins adopted the logical positivists’ view in his famous definition and characterization of economics, employed it in his paradigm critique of interpersonal utility comparisons, and put in place the value neutrality view of the positive-normative distinction which most economists continue to hold. Section 3 reviews why philosophers’ postwar rejection of logical positivism was also a rejection of the fact/value dichotomy and Hume’s guillotine. The purpose here to show that economists’ understanding of the positive-normative distinction lacks reasonable foundations. At the same time, the main criticisms of logical positivism do not directly get at economists’ understanding of value neutrality. Section 4 thus looks at one of the many ways in which philosophers have subsequently rethought the relationship between facts and values, namely Hilary Putnam’s fact-value ‘entanglement’ view. Putnam’s view is modified slightly in order to directly address Robbins’ particular understanding of value neutrality as specific to ethical values. Section 5 then turns back to the economics profession, and argues that economists embrace four ethical value judgments, despite their espousal of value neutrality. To explain this odd stance, section 6 offers a behavioral economics explanation of economists’ attachment to value neutrality. Section 7 concludes with brief comments on the identity of economics.

2. *Logical positivism and the fact-value dichotomy*

Logical positivism, as derived from the interwar Vienna Circle (Hands, 2001, pp. 72ff), had as its central principle a criterion of meaningfulness called the verifiability principle. The logical positivists classified all

scientific statements or propositions either as analytic, which were tautological or true by definition (and thus trivially meaningful), or as synthetic *a posteriori*, which were meaningful in that they could be shown to be empirically true or false. They distinguished synthetic *a posteriori* propositions from synthetic *a priori* ones, which also made claims about the world but were not subject to empirical verification. Immanuel Kant had argued that synthetic *a priori* propositions such as Newton's laws were universally true and thus meaningful. However, the logical positivists believed that synthetic *a priori* propositions could never be proved empirically true or false, and accordingly argued that allowing them to be regarded as scientific knowledge opened the door to a wide range of arbitrary claims about the world. It is worth noting that this view was developed at a time in Europe when racist and nationalist ideas were regularly claimed to be scientific, and that members of the Vienna Circle were among those personally exposed to violence and abuse in the name of such views. In any event, the verifiability principle said that scientific statements needed to be subject to empirical investigation determined to be seen as scientifically meaningful, and all other statements were deemed 'metaphysical.' Its basis lay in positivist philosophies of science and especially in the empiricist thinking of David Hume.

Hume had made a distinction similar to the logical positivists' analytic-synthetic distinction two hundred years earlier when he distinguished matters of fact and relations of ideas (Hume, 1739; 1978). The fact-value dichotomy was a consequence of this distinction. Hume characterized statements that express matters of fact as 'is' statements, and complained that "systems of morality" commonly switched from 'is' statements to 'ought' or 'ought not' statements without ever explaining this "new relation" (Hume, 1739; 1978, pp. 469-470). Thus Hume's Law, or the Humean guillotine as the fact-value dichotomy came to be called, is generally associated with the idea that one cannot infer statements about values from statements about facts. But behind this idea lies an even stronger claim, namely, that there are no matters of fact about values. For example, it can never be a fact that murder is morally wrong. Indeed if 'ought' statements did refer to facts, then there would be a way of moving to them from 'is' statements by linking sets of facts. So Hume's Law depended not so much on the different linguistic characters of 'is' and 'ought' statements as upon his view that values and value judgments could never be matters of fact, a view he defended by differentiating the origins of facts and values. Facts had their basis in the world in that they are ideas that 'resemble' the world. In contrast, values are ideas which Hume believed had their basis in our minds and derived from our 'passions' and 'sentiments.' This was not a very good argument, since there could be facts about our 'passions' and 'sentiments' just as there are facts about the world. But the association of facts with the world and values with the mind stuck, and from this it was subsequently inferred that facts were objective and could ground science, while values were subjective and had no place in science.

Thus when the logical positivists adopted Hume's Law and his empiricism, they also adopted his view that values were inherently subjective. Their primary goal had been to demarcate science from non-science

(Boumans and Davis, 2010, pp. 9ff). Thus it was not necessary to go on to characterize ethics, theology, literature, etc. which were clearly not science as metaphysical, nor did it was necessary to see these domains as subjective in a pejorative sense. But early followers of logical positivism, A.J. Ayer (1936) and Charles L. Stevenson (1944), settled upon precisely this meaning in a companion doctrine termed emotivism. Following Hume, Ayer characterized value judgments and moral statements as expressions of our feelings and attitudes, or as mere ‘emotings.’ Stevenson was more subtle, and allowed that value judgments and ethical language could have both descriptive and emotive meaning, but believed the latter was nonetheless essential and independent of the former. The implication for both was that values and ethics occupied a domain in which only persuasion and expression mattered, because they lacked objective basis. This meant that there was no way to say objectively that there was right or wrong or good or bad and that language expressing such judgments was simply a matter of one imposing one’s feelings on others.

This is basically the view that Robbins adopted in his famous definition of economics and critique of interpersonal utility comparisons as being motivated by value judgments (1932; 1938; Davis, 1990, 2005). Indeed, he went further than Ayer and Stevenson to treat ethical disagreements as not just expressions of feelings and attitudes, but as irresolvable contests for power. Arguing that contests for power were always over either means or ends, and associating economics with the investigation of means and ethics with debate over values, Robbins asserted that science alone was able to resolve such conflicts.

If we disagree about ends it is a case of thy blood or mine – or live and let live according to the importance of the difference, or the relative strength of our opponents. But if we disagree about means, then scientific analysis can often help us resolve our differences. If we disagree about the morality of the taking of interest (and we understand what we are talking about), then there is no room for argument (Robbins, 1932, p. 134).

The problem with interpersonal utility comparisons for Robbins, then, was that they were irretrievably subjective because they were based on people’s feelings of utility, so that saying one person got more utility from some amount of consumption than another was simply crediting that person with more power over consumption (Hands, 2012pp. 222ff). That person’s subjective end, in effect, was counted more strongly than others’ ends, and so to keep economics scientific and free of power contests interpersonal utility comparisons had to be excluded. Interpersonal utility comparisons thus became the paradigm example of how values damaged economics as a science.

Robbins lived through the same period as the logical positivists, and was no less concerned than they that science might become a battleground for contesting ideologies. Thus he also thought it imperative to remove from science what he believed to be the ground on which such contests might be played out, namely, competing values. The expression of this commitment was the idea that science is value-neutral. Of course

the recommendation that science be value-neutral is a value judgment, so it might be argued that Robbins' position was self-refuting. But it seems fair to say that his view of the matter was more nuanced. He presumably recognized that economists' investigation of means entailed making value judgments about how one ought to carry out one's investigations. These kinds of value judgments, however, are what Mark Blaug termed methodological value judgments (Blaug, 1998), and while scientists can disagree about how science ought to be practiced, these values were generally not those that generate power conflicts, or at least not the kind associated with clashing social ideologies. Thus the values Robbins really wanted excluded from economics were those associated with ethics. Consequently, a corollary to his value-neutrality view is that methodological values employed in economics hold no implications for ethical values. The former are essentially benign, and merely part of the everyday practice of science. It is important to keep this meaning of value neutrality in mind as we proceed, because saying there are methodological values in economics in order to refute Robbins' view targets something he accepted and found unproblematic, and because the truly fundamental issue behind the idea of value neutrality is that ethical values such as what is morally good or right cannot be a part of a scientific economics. This latter stronger view of value neutrality, however, is precisely what philosophers have rejected.

3. *Philosophers' postwar rejection of logical positivism and the fact-value dichotomy*

The principal sign-post in postwar philosophy's turn against logical positivism is W. V. Quine's attack on the analytic-synthetic distinction (Quine, 1953). Quine's basic point was that there are many propositions or statements in science which cannot be easily characterized as either analytic or synthetic *a posteriori*. Partly he took this position because he thought the logical positivists had mischaracterized analytic statements – the language of mathematics and logic – which as true by definition and not as matters of fact were supposed to be independent of matters of fact or what can be shown to be empirically true about the world. Quine, who was influenced by American pragmatist philosophers, believed that mathematics and logic possessed many conventional elements. Thus he thought it was wrong to say analytic statements were trivially true. How they were otherwise to be understood is a large philosophical question, particularly in regard to how matters of fact entered into the language of mathematics and logic, which need not be addressed here. What is important is that Quine's doubts about the logical positivists' view of analytic statements created doubts about the whole analytic-synthetic distinction, and this then stimulated philosophers' reflection on the logical positivists' understanding of synthetic *a posteriori* statements.

The basic idea that synthetic *a posteriori* statements were true as a matter of fact seems straightforward and uncontroversial. The logical positivists regarded them as observation sentences, meaning that if one said 'the sky is blue,' one could verify that statement by looking to see if the sky was blue. This meant that each

observation sentence was individually evaluated according to its correspondence to the world, and one carried out the evaluation of a set of scientific claims sentence-by-sentence in sequential inspections of the world. By the late 1930s, however, this simple model had undergone significant revision. Observation sentences could not be evaluated individually, because not all the language used to express them corresponded to observational features of the world. For example, in physics terms such as ‘electron’ and ‘charge’ were not definitions (thus not derived from analytic statements) nor were they observable, and accordingly it was conceded by the logical positivists that they needed to be “taken as primitive” (Carnap, 1938). However, admitting theoretical or abstract terms in this way was still thought to be consistent with the logical positivists’ overall view if whole collections of observation sentences taken together could be said to produce successful predictions about the world. Synthetic *a posteriori* statements, that is, were still essentially matters of fact, and the analytic-synthetic distinction still seemed to hold.

But the admission of theoretical or abstract terms and acceptance of the idea that individual observation statements could not be evaluated individually opened the door to an entirely different view of matters of fact which Quine’s doubts about the analytic-synthetic distinction and thoughts about analytic statements now fostered. Thus if analytic statements contained conventional elements that were important in determining the nature of mathematics and logic, it seemed fair to say that scientific theories contained conventional elements that were also important in determining what counted as a matter of fact. But this implied that synthetic *a posteriori* statements were no longer pure observation statements. Whereas the logical positivists had thought observation occurred through inspection of the world, the view was now that scientific conventions determined how that inspection was to be carried out. The concept of ‘inspection’ was far too simplistic a representation of how scientists organized their empirical research. This meant that what counted as a fact reflected how a collection of sentences was assembled and investigated by scientists to produce an acceptable scientific theory. Scientific theories thus displaced scientific statements as the principle unit of understanding, and facts then provided evidence for theories according to how evidence was to be interpreted in a given theory. Thus the rejection of the analytic-synthetic distinction ultimately changed philosophers’ understanding of science to being about theories rather than statements, and produced a more subtle view of what facts are. This raises the issue of what the consequences of this were for the analytic-synthetic distinction companion fact-value dichotomy.

Corresponding, then, to the idea that there is no sharp line between analytic and synthetic statements we find the idea that there is also no sharp line between facts and values – despite the clear differences that exist between ‘is’ statements and ‘ought’ statements. This followed directly from the changed status of facts in post-logical positivist philosophy and the recognition that conventional aspects of scientific theories often depend on value judgments about how theories should be constructed, where this can include value judgments about the significance of theoretical or abstract terms and about how evidence is to bear on

theories' predictive capacities. At the very least, then, that values underlie scientific practice and play a role in determining what counts as fact means that the fact-value dichotomy cannot be sustained. Note how this discredits Hume's argument about the linguistic difference between 'is' statements and 'ought' statements. Hume also reasoned in terms of individual sentences when he used this distinction to infer a metaphysical divide between facts and values. But when we think in terms of scientific theories rather than individual sentences, it turns out that these linguistic differences tell us very little about facts and values. Rather, the linguistic difference between 'is' statements and 'ought' statements apparently conceals rather than exhibits how values and facts are closely connected in scientific theories. Thus it cannot be argued that language alone tells us what is objective and subjective in science. For that matter, it is no longer clear what it means to talk about what is objective and subjective in science. There surely is something subjective in scientists relying on conventions and values to determine what counts as scientific knowledge. But is scientific knowledge then subjective or objective precisely because it depends on established conventions regarding scientific practice? I leave these important questions aside, and return to the main issue here: Robbins and value neutrality.

To be clear, the argument above applies specifically to what Blaug referred to as methodological value judgments, not to the sort of value judgments associated with ethics. So economists could still reject the analytic-synthetic distinction and the fact-value dichotomy while maintaining a value neutrality that specifically identifies ethical values as the types of values that have no place in economics. This is the view that I believe is enshrined in the textbooks. It separates positive economics seen to be free of ethical commitments from normative economics which employs a value-neutral positive economics in ethically-guided policy decision making. But philosophers question even this view of the fact-(ethical)value dichotomy. The following section presents Hilary Putnam's arguments on this score.

4. Putnam's 'entanglement' view of facts and values

Putnam makes an argument like Quine's, though not with respect to the classification of scientific statements but rather with respect to the classification of concepts in language. Whereas Quine said many statements are neither purely analytic nor purely synthetic *a posteriori*, Putnam, addressing the associated fact-value dichotomy, says that many concepts we employ in ordinary language and science are neither purely descriptive nor purely normative but are somehow descriptive *and* normative. Further, Quine was critical of the logical positivists' intention to reform science by imposing a philosophical regime upon it using the analytic-synthetic distinction to demarcate science from non-science. Similarly, Putnam is critical of the logical positivists' intention to "*rationaly reconstruct* our language ... to make a 'sharp and clear' distinction between value terms and descriptive terms" (Putnam, 2002, p. 24.). His basic position is that science

develops within the confines of ordinary language, and replacing its natural form of expression with a regimented formal logical language, as for example was promoted in the logical empiricist program developed by Rudolph Carnap (1928), is likely to distort the meanings of the concepts we employ. In particular, Putnam argues, ordinary language concepts commonly ‘entangle’ facts and values, where the definition of ‘entanglement’ is that we often cannot understand these concepts, and say what the fact of the matter is on a given subject unless we recognize the value judgments we make in using those concepts.

Consider, then, the logical positivists’ concept of an observation term, such as would appear in observation sentences. On their view, observation terms are purely descriptive concepts free of values which have their meaning solely in virtue of how they correspond to some feature of the world. Putnam asks us to consider whether this view fits the case of an historian who seeks to describe a Roman emperor, and uses ‘cruel’ as observation term (Putnam, 2002, pp. 24ff). It is clearly not the case that the term ‘cruel’ simply corresponds to some feature of the world in the way that the term ‘blue’ can be said to correspond to color of the sky. It also not the case that ‘cruel’ is an abstract theoretical term like ‘electron’ that refers to nothing in the world, since the historians believe some Roman emperors actually had the characteristic of being cruel. Perhaps one might associate the property of being cruel with certain neurological states of human beings in order to still be able to treat ‘cruel’ as a value-free observation term. But Putnam believes that this is not what most historians are thinking of when they explain cruelty. Cruel Roman emperors may well have been in certain brains states when they were cruel, but the historian’s explanation is pitched at the level of psychological descriptions. The conclusion we need to draw, he then argues, is that historians are indeed interested in describing the world, but that these descriptions often depend on value judgments we make about such things as cruelty. ‘Cruel’ is indeed an observation term, but an ‘entangled’ one that requires value judgments, and thus not the sort of observation term envisioned by the logical positivists. Moreover, we fail to accurately describe the behavior of cruel Roman emperors if we omit these value judgments. That is, thus our descriptions are objective only if they are normative as well as descriptive.

To be sure, the logical positivists were not much concerned with historical descriptions, and instead focused on natural science terms with physics as their model. Thus they had little reason to be concerned about value-entangled terms such as ‘cruel,’ because natural science does not employ psychological descriptions where they tend to arise. Relatedly, the natural sciences generally employ highly specialized vocabularies which lack the nuances of ordinary language. Thus it could be said in their defense that the issue was less a matter of reconstructing ordinary language than of using technical scientific language precisely. Note, however, what this implies about economics. Putnam could be correct on the whole about the ‘entanglement’ of facts and values in ordinary language, but his view could still be set aside by economists in the spirit of logical positivism were they to model economics on the natural sciences. Consequently the issue would then be the nature of the vocabulary in economics.

Economic methodologists, in fact, have debated this issue in terms of whether economics relies on ordinary language concepts – what has been referred to as ‘commonsensibles’ – in connection with debate over whether economics is realist (Guala, 2012, Hands, 2012a). Is the concept of a preference, for example, an ordinary language term or commonsensible with potential value dimensions subject to Putnam’s strictures, or is it an abstract theoretical term like electron suitable only to formal language analysis? What the truth of the matter is (if there is one) may be less important than whether economists elect to treat concepts in economics in a technical way on the model of the natural sciences. Then, even if the analytic-synthetic distinction and the fact-value dichotomy are flawed doctrines as philosophers have argued, economists might proceed as if they were *de facto* acceptable principles by practicing economics in such a way as to exclude the interpretation of economic concepts most likely to give rise to value-entangled commonsensibles thinking. Note that many of those concepts that might fall into the ordinary language commonsensibles category – what are referred to as folk psychology concepts such as concern agent beliefs, preferences, etc. – are like the value-entangled descriptive terms historians use. Thus the strategy economists could adopt to preserve the fact-value distinction is to insist that these are formal concepts, and that folk psychology has no place in economics. This view has recently been given especially strong statement as a response to the emergence of behavioral economics as a science-based folk psychology approach (Gul and Pesendorfer, 2008). What, then, is the relevance of Putnam’s entanglement view to economics?

First, suppose that in economics technically defined terms function as substitutes for ordinary language terms, do not operate *per se* like ordinary language terms, and are thus ostensibly immune to the entanglement critique. If Putnam is right that science develops within the confines of ordinary language, then it would still be reasonable to argue that technical terms’ ordinary language associations are somehow preserved despite their redefinition. That is, the substitution of technical meanings for ordinary meanings must derive the former from the latter, or else the technical term is not a derivation but an entirely new concept. Then it could be argued that the fact-value entanglement which concepts have in ordinary language would be somehow implicit in technical concepts as well. Of course one could resist this view, and say that technical concepts such as the axiomatically defined preference concept are entirely new concepts and are not derived from ordinary language concepts. But then it would be difficult to say that their subject matter was still people’s decision behavior. What would an axiomatically defined preference concept be about if it were not a formal representation of what people ordinarily understand (if less rigorously) by the meaning of the term ‘preference?’ Most economists likely thus believe that the technical concepts in economics are derived commonsensible concepts.

Second, suppose that since technical concepts are not ostensibly value-entangled, that the effect of their derivation from ordinary language is to produce fact-value entanglement in an indirect rather than a direct way, specifically at the level of the theories in which they are employed rather than at the concept level. Then

the technically defined axiomatic preference concept could still be said to be value-free itself, but the economic theory in which it is employed could nonetheless be shown to exhibit the fact-value entanglement Putnam ascribes to concepts. This modification of Putnam's view consequently moves the analysis from how individual terms are evaluated to how the theories in which they appear are evaluated. If his definition of fact-value entanglement at the level of the concept is that we cannot understand what the concept describes unless we understand the value judgments we make in using it, then the modified Putnam definition of fact-value entanglement would be that we cannot understand the explanations these theories provide unless we understand the value judgments we make in these theories.

Of course there are good reasons to be skeptical of this reformulated view. One could agree that technical vocabularies are derived from ordinary language, yet also argue against Putnam that their formal expression specifically suppresses or neutralizes the value associations of the latter. This was indeed essentially the rationale the logical positivists offered for their logical empiricism research program (Hands, 2001, pp. 82ff). At the same time, it could be replied to this that post-positivist philosophy offers broad grounds for a reformulated Putnam entanglement view in the idea that theory construction involves many conventional elements and methodological values, so that theories, even when formulated in specialized vocabularies clearly contain value judgments. But as methodological value judgments are not ethical value judgments, it remains an open question whether this reformulation of Putnam's view successfully challenges the standard value neutrality conception of economics in the stronger sense that says positive economics is free of normative ethical judgments. The following section accordingly puts this question to the test by asking whether standard economic theory that employs a technical vocabulary nonetheless presupposes ethical value judgments.

5. *Four ethical value judgments of standard economic theory*

For the sake of the argument, then, let us assume that the concept of a preference in economics is not an ordinary language commonsensible and is a technical term that is ostensibly free of value associations. Further, it seems fair to say that in standard economic theory the concept of a preference is foundational to Walrasian equilibrium explanations of markets that are characterized as Pareto-efficient or Pareto-optimal, where that concept describes a particular relationship between individuals' preferences. Specifically, the first fundamental theorem of welfare economics (known as the Arrow-Debreu theorem) states that any Walrasian equilibrium, defined in terms of individual preference maximization, is Pareto-efficient (Mas-Colell *et al.*, 1995, chapter 16). If the standard value regarding neutrality view holds, then this explanation should be value-neutral, and no ethical value judgments are made (though methodological value judgments are allowed). Is the first fundamental theorem of welfare economics indeed free of ethical value judgments?

The rationale for the Pareto concept was established by Robbins when he argued that cardinal interpersonal utility comparisons rested on value judgments, and should therefore be excluded from economics. Interpersonal utility comparisons had been previously employed to argue that low income individuals gained more utility from a given additional income than high income individuals lost were their income reduced by the same amount, so that transferring income from high to low income individuals increases total utility. Robbins' critique of interpersonal utility comparisons was that they were inevitably subjective, so that it could not be reliably established that low income individuals indeed gained more utility from additional income than high income individuals lost from reduced income. As a result of his arguments, most economists ultimately rejected interpersonal utility comparisons and gave up the cardinal utility concept, making the Pareto concept the sole welfare evaluation concept in standard economics. What the Pareto concept consequently does is rule out total utility-increasing income transfers between individuals by saying that no individual's preference satisfaction can carry any more weight than any other individual's. A state of affairs is accordingly Pareto-efficient when no individual can acquire greater preference satisfaction except at the expense of someone else. Note, then, that there are four important ethical judgments that this reasoning involves.

First, that Pareto judgments require that all individuals' preferences have the same weight in terms of preference satisfaction irrespective of their other characteristics is an ethical value judgment associated with saying how different individuals' well-being counts in judgments of well-being. Interpersonal utility comparisons presuppose a different weighting principle, namely, that low income individuals' utility satisfaction counts more than high income individuals' utility satisfaction. Pareto judgments weight all individuals' preference satisfaction equally. Each weighting conception has ethical meaning, but they conflict, and choosing one standard over the other makes an ethical value judgment. Robbins' complaint that interpersonal utility comparisons are subjective is a separate issue regarding their feasibility, not their ethical content. Further, while the ordinal meaning of preference underlying the Pareto concept is a technically defined, clearly the concept is still derived from the commonsensible preference idea, since Pareto judgments concern how real world preferences should be addressed in economic policy choices. Thus when the technical preference concept is used in a theory to explain efficiency, it makes an 'entangled' ethical value judgment.

Second and relatedly, since Pareto judgments necessarily count all individuals equally in terms of preference satisfaction, distributional issues regarding the distribution of preference satisfaction are completely set aside. Distributional issues arise in ethics in connection with the concepts of fairness and justice. Preference satisfaction, however, is a well-being concept in ethics, where the focus is on what it is that makes people well off. It is a teleological ethical concept framed in terms of ends, whereas fairness and justice concern what is right, a deontological type of ethical concern. Thus when one employs Pareto judgments, one is effectively saying that distributional issues are not an ethical priority, or more strongly, that well-being is the only value

priority. This by itself is an ethical value judgment. But the Pareto concept also makes an implicit ethical value judgment about what is fair or just. In cases where it is Pareto efficient to increase the preference satisfaction of one very well off individual though all other individuals are not well off the implicit ethical judgment made is that this is fair or just. Thus Pareto reasoning is not only 'entangled' with the ethical issue regarding whose well-being counts, but also with ethical issues regarding the relation of teleological to deontological ethics.

Third, Pareto efficiency judgments ignore the content of preferences, and this also constitutes an ethical value judgment. Some preferences, of course, are morally objectionable, such as a preference to cause others harm. To say, then, that such preferences should not be disallowed, which is the effect of saying the content of preferences is irrelevant to Pareto judgments, is to make the ethical value judgment that the freedom to satisfy one's preferences is more important than whatever morally undesirable consequences might result from people freely pursuing their preferences. It is true that the idea of being free to act on one's preferences can be understood descriptively without introducing normative meaning. But when taken in the context of Pareto judgments, free pursuit of one's preferences implicitly recommends one set of ethical values over others. Thus though the Pareto concept is a technical term in economics, the theory in which it is employed entangles its meaning with one particular normative meaning of freedom.

Fourth, the Pareto principle provides a specific welfare conception of well-being, that is, well-being explained in terms of preference satisfaction. But there are many other welfare concepts of well-being which are put aside, and this constitutes an ethical value judgment. Within the utility welfare framework itself, cardinal utility is a measure of well-being that can be explained in terms of desire satisfaction. Another conception of welfare well-being is subjective happiness, where this can be understood in terms of individuals' reports of perceived happiness. However, one can also speak of well-being using concepts not framed in terms of welfare at all. An example of the latter is the idea of a just society, as when one says one's well-being is greater in a just society. As noted above, the Pareto principle says nothing about justice, so it is not only selective in its welfare basis for well-being, but also excludes non-welfarist well-being ideas. Multiple ethical value judgments on multiple levels are thus made when one uses the Pareto concept, despite the fact that economists generally regard the Pareto concept as a value-neutral, descriptive technical term.

Thus, if we take the concept of preference and Pareto efficiency judgments as foundational to standard economics theory, the reformulated Putnam fact-value entanglement view accurately describes that theory. As a technically defined idea, the preference concept may be ostensibly value-free, but the role it plays in explaining equilibrium outcomes as Pareto-efficient makes numerous ethical value judgments. Economists consequently promote one ethical vision of the world while claiming that economics is a positive value-neutral subject, and extol the positive-normative distinction while systematically violating it. How are we to

understand this odd combination of beliefs? One might reply that economists pay little attention to philosophers and arguments such as the one above, and are simply unaware of how untenable their thinking is on facts and values. Their professional practice is highly empirical in emphasizing data and evidence, and perhaps they would revise their views on this subject were they more exposed to these issues. But this seems an inadequate response since while it is understandable that philosophers' arguments might be remote for economists, the ways in which ethical judgments operate in Pareto thinking are not any more difficult to understand than many other complex conceptual issues in economics, while economists are surely aware that Pareto-based policy recommendations are ordinarily contested from other ethical perspectives. Thus the section that follows offers a behavioral economic analysis of economists' steadfast attachment to the positive-normative distinction.

6. *A behavioral economics interpretation of economists' value neutrality view*

Behavioral economics investigates the psychological dimensions of economic decision-making. From an economics of scientific knowledge perspective (Hands, 2001, pp. 365ff), economists themselves can be said to engage in economic decision-making to the extent that their research and scientific activity have economic motivations. Economists clearly recognize that some types of research are more conducive to career advancement than others, so it seems reasonable to say that economic motivations have an influence on economists' research choices. Accordingly it makes sense to ask whether the economically motivated decisions economists make also have psychological dimensions. Taking prospect theory (Kahneman and Tversky, 1979) as our guide for how to understand the psychological dimensions of choice means that we should see choices as being framed by reference points.

An essential feature of the present theory is that the carriers of value are changes in wealth or welfare, rather than final states. This assumption is compatible with basic principles of perception and judgment. Our perceptual apparatus is attuned to the evaluation of changes or differences rather than to the evaluation of absolute magnitudes. When we respond to attributes such as brightness, loudness, or temperature, the past and present context of experience defines an adaptation level, or reference point, and stimuli are perceived in relation to this reference point (Kahneman and Tversky, 1979, p. 277).

But while current wealth or welfare constitute a generic sort of adaptation level or reference point appropriate to economic choices in general, in more specific choice contexts reference points presumably more closely reflect the nature of those choices. What sorts of reference points, then, might operate in regard to economists' research choices?

When we ask how economists make choices as economists, that is, with an awareness of being economists, then their professional social identity would function as a reference point for the choices they make. In social identity theory, when one acts on a social identity, one acts as a representative agent of that social identity (Tajfel *et al.*, 1971). In behavioral economics terms, then, when economists are economically motivated to make research choices which they believe a representative economist would make, one such research choice they would likely make would be to support the standard value neutrality view, since that view is representative of standard economics. In effect, just as economists recognize that some types of research are more conducive to career advancement than others, so they would recognize that contesting the standard value neutrality view would likely be inimical to career advancement. This explanation offers one way of understanding the odd circumstance that economists remain attached to a position that philosophers and other social scientists have long rejected as indefensible. That their position is indefensible is irrelevant for them, since the choice to support value neutrality is a psychological one motivated by career considerations, not one motivated by a reasoned evaluation of the position itself.

Returning to prospect theory, we also find that the theory provides a way of saying something about economists' degree of attachment to value neutrality. "Strictly speaking, value should be treated as a function in two arguments: the asset position that serves as reference point, and the magnitude of the change (positive or negative) from that reference point" (Kahneman and Tversky, 1979, p. 277). The interpretation then given to the value function which substitutes for the standard utility function is that people generally exhibit loss aversion.

A salient characteristic of attitudes to changes in welfare is that losses loom larger than gains. The aggravation that one experiences in losing a sum of money appears to be greater than the pleasure associated with gaining the same amount (Kahneman and Tversky, 1979, p. 279).

Economists' value function, when we address the research choices they make which concern value neutrality, would exhibit loss aversion if they felt that that arguments they might make that would potentially undermine value neutrality would "loom larger" than arguments they might make that would reinforce the position. That is, when their social identity as economists acts as a reference point for the choices economists make, and an important component of this social identity is the value neutrality position, then a way of understanding the strength of their attachment to the position is to say economists feel that their standard economics social identity would likely be jeopardized by giving any more than cursory attention to the subject. Indeed, the sense that the position could suffer losses is probably well-founded in that economists are presumably at least aware that their view of value neutrality is not shared by many outside economics. Consequently investing in potential gains to the position by making further arguments for it is not attractive.

To avoid potential losses relative to their social identity as economists, the best course of action is to ignore the subject altogether.

A behavioral economics explanation of economists' attachment to value neutrality accordingly not only tells us why economists remain attached to an indefensible position, namely, that as a social identity reference point its content and meaning are simply irrelevant. It also tells us, when we impute a prospect theory representative value function to economists, that on loss aversion grounds any discussion of the subject should be avoided if at all possible.

Ironically, behavioral economists also generally subscribe to the standard value neutrality view, and see prospect theory as providing an "adequate descriptive model" for explaining choice behavior under risk (Kahneman and Tversky, 1979, p. 263). Indeed, the value function prospect theory employs as a replacement for the standard utility function basically has the same preference-based welfare characteristics, so that prospect theory also makes the four ethical value judgments reviewed in the previous section. This does not mean that a behavioral economic analysis of economists' attachment to value neutrality should be set aside. It rather means that behavioral economics' analysis applies reflexively to itself as well, so that following the expanded reading of Putnam's argument provided above we should say that prospect theory is no less value-entangled than standard utility theory.

7. Concluding comments on the identity of economics

Previously I argued (Davis, 2008) that over its history economics has regularly reproduced itself as a core-periphery structure in which rival approaches compete in the periphery to occupy the core and be identified with economics, and core positions in economics change over time. A core-periphery account accordingly provides an account of the identity of economics whereby the discipline can undergo change but retain recognizably distinct core content. One issue this view leaves open is why some approaches and not others come to occupy the core of the discipline. Presumably part of the answer concerns the social uses of economics (though surely not all successful approaches are socially useful). But another part of the answer might concern how dominant approaches selectively admit and filter content adopted from peripheral approaches. For example, game theory came into the core of economics in the 1980s but only in the form of non-cooperative game theory. Thus though over the long history of economics the core of the field has been occupied by very different types of approaches, in the short run changes in the core are generally gradual. On a filter view, however, the adoption of content from peripheral approaches in the long run can be either consistent with core views or ultimately disruptive of them. If game theory is an example of the former case,

then experimentalism might be an example of the latter, as evidenced by the evolution of Vernon Smith's thinking about standard theory. Smith began his experimental work expecting to confirm standard assumptions, and this helped introduce experimental research into economics, but over time he increasingly found himself disconfirming them (Smith, 2010), and this can be seen as having had a disruptive effect on standard economics. This state of affairs can possibly be due to the filtering processes in a discipline working non-monolithically on a variety of levels with limited forward-lookingness. Thus it is perfectly possible that new ideas seen initially to be consistent with existing core conceptions later turn out not to be.

What might we say in this regard with respect to value neutrality? It is probably fair to say that all the new approaches in economics today share standard theory's value neutrality position. Behavioral economics is not unique in this regard. Thus with this barrier to entry missing some of their contents (or indeed entire approaches) could enter core of economics in the future. But there is an important difference between the new approaches in economics and standard theory on value neutrality. Whereas the basis for value neutrality in standard economics lies in Robbins' arguments, the new approaches in economics with their different conceptual origins not only lack clear connection to Robbins' thinking, but also appear to lack any specific arguments in favor of value neutrality. The value neutrality doctrine in economics has a specific history and a tight set of connections to general equilibrium theory and Pareto judgments. But the new approaches in economics generally operate with different theoretical foundations, and their attachment to the doctrine appears to be largely conventional. So it is not clear what basis the doctrine might retain in economics should these new approaches or elements from them become part of core economics. Of course this does not imply that economics will change its odd stand on value neutrality. But it might be a mistake to assume things will never change in this regard. What is called for, it thus seems, from those who would like to see economics adopt a more enlightened view of values in economics is careful attention to how the new approaches in the field address the issue, on the assumption that their interpretation of the matter might ultimately supplant the standard view.

References

Ayer, A. J. (1936) *Language, Truth, and Logic*, London: Gollancz, 2nd Edition, 1946.

Blaug, Mark (1998) "The Positive-Normative Distinction," in J. Davis, W. Hands, and U. Mäki, eds., *The Handbook of Economic Methodology*, Cheltenham: Edward Elgar, 1998, pp. 370-4.

- Boumans, Marcel and John Davis (2010) *Economic Methodology: Understanding Economics as a Science*, Basingstoke: Palgrave Macmillan.
- Carnap, Rudolph (1928) *Der Logische Aufbau der Welt*, Leipzig: Felix Meiner Verlag, English translation by Rolf A. George, *The Logical Structure of the World. Pseudoproblems in Philosophy*. University of California Press, 1967.
- Carnap, Rudolph (1938) “The Foundations of Logic and Mathematics,” in *International Encyclopedia of Unified Science*, vol. 1, part 1, Chicago: University of Chicago Press, pp. 139-214.
- Davis, John (1990) “Cooter and Rappaport on the Normative,” *Economics and Philosophy* 6 (1): 136-146.
- Davis, John (2005) “Robbins, textbooks, and value neutrality,” *History of Political Economy* 37 (2): pp. 191-6.
- Davis, John (2008) “The turn in recent economics and return of orthodoxy,” *Cambridge Journal of Economics*, 32 (May): 349-366.
- Guala, Francesco (2012) “Are preferences for real? Choice theory, folk psychology, and the hard case for commonsense realism,” in A. Lehtinen, J. Kuorikoski, and P. Ylikoski, eds., *Economics for Real: Uskali Mäki and the place of truth in economics*, London: Routledge, pp. 137-155.
- Gul, Faruk and Wolfgang Pesendorfer (2008) “The Case for Mindless Economics,” in A. Caplin and A. Schotter, eds., *The Foundations of Positive and Normative Economics: A Handbook*, Oxford: Oxford University Press, pp. 3-39.
- Hands, D. Wade (2001) *Reflection without Rules: Economic Methodology and Contemporary Science Theory*, Cambridge: Cambridge University Press.
- Hands, D. Wade (2012a) “Realism, commonsensibles, and economics: the case of contemporary revealed preference theory,” in A. Lehtinen, J. Kuorikoski, and P. Ylikoski, eds., *Economics for Real: Uskali Mäki and the place of truth in economics*, London: Routledge, pp. 156-178.
- Hands, D. Wade (2012b) “The Positive-Normative Dichotomy and Economics,” in Philosophy of Economics, U. Mäki (ed.), Vol. 13 of D. Gabbay, P. Thagard and J. Woods (eds.), *Handbook of the Philosophy of Science*, Amsterdam: Elsevier: pp. 219-239.
- Hume, David (1739) *A Treatise of Human Nature*, edited by L. A. Selby-Bigge, 2nd ed. revised by P.H. Nidditch, Oxford: Clarendon Press, 1978.
- Kahneman, Daniel and Amos Tversky (1979) “Prospect theory: An analysis of decision under risk,” *Econometrica* 47: 263-291.

Mas-Colell, Andreu, Michael D. Whinston, and Jerry R. Green (1995) *Microeconomic Theory*, Oxford University Press.

Putnam, Hilary (2002) *The Collapse of the Fact/Value Dichotomy and Other Essays*, Cambridge, MA: Harvard University Press.

Quine, W. V. (1953) "Two Dogmas of Empiricism," in *From a Logical Point of View*, Cambridge, MA: Harvard University Press.

Robbins, Lionel (1932) *An Essay on the Nature and Significance of Economic Science*, 2nd ed., London: Macmillan, 1935.

Robbins, Lionel (1938) "Interpersonal Comparisons of Utility: A Comment," *Economic Journal*, 48: 635-41.

Smith, Vernon (2010) "Theory and Experiment: What are the questions?" *Journal of Economic Behavior and Organization* 73: 3-15.

Stevenson, Charles (1944) *Ethics and Language*, New Haven: Yale University Press.

Tajfel, Henri, Michael Billig, Robert Bundy, and Claude Flament (1971) "Social Categorization and Intergroup Behavior," *European Journal of Social Psychology* 1: 149-178.