

Marquette University

e-Publications@Marquette

Economics Faculty Research and Publications

Economics, Department of

2005

Neoclassicism, Artificial Intelligence, and the Marginalization of Ethics

John B. Davis

Marquette University, john.davis@marquette.edu

Follow this and additional works at: https://epublications.marquette.edu/econ_fac



Part of the [Economics Commons](#)

Recommended Citation

Davis, John B., "Neoclassicism, Artificial Intelligence, and the Marginalization of Ethics" (2005).

Economics Faculty Research and Publications. 146.

https://epublications.marquette.edu/econ_fac/146

Marquette University

e-Publications@Marquette

Economics Faculty Research and Publications/College of Business

This paper is NOT THE PUBLISHED VERSION; but the author's final, peer-reviewed manuscript. The published version may be accessed by following the link in the citation below.

International Journal of Social Economics, Vol. 32, No. 7 (July 2005): 590-601. [DOI](#). This article is © Emerald Group Publishing Ltd. and permission has been granted for this version to appear in [e-Publications@Marquette](#). Emerald Group Publishing Ltd. does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Ltd.

Neoclassicism, artificial intelligence, and the marginalization of ethics

John B. Davis

University of Amsterdam and Marquette University, Milwaukee, Wisconsin, USA

Abstract

Purpose

The paper examines the dependence of the positivist and welfarist preference satisfaction paradigm of neoclassical economics upon an implicit functionalist philosophy of mind. Functionalism is the doctrine that mental states are strictly materialistic and understandable in cause-effect terms. An important aspect of functionalism is the multiple realizability thesis, namely, that mental states can be realized in any type of hardware, whether human brain or computer.

Design/methodology/approach

The approach used involves investigating the fact-value distinction after Robbins in terms of the positivist meta-ethical view known as emotivism, and then explaining emotivism as inherently functionalist. Functionalist thinking itself is explained in terms of contemporary philosophy of mind.

Findings

An important finding is that the preference satisfaction paradigm can be shown to be as suitable to artificial intelligence systems as to human beings. A consequence of this is that normative concerns are increasingly difficult to address in connection with the neoclassical thinking about economic agents.

Research limitations/implications

The paper does not investigate more recent research programs in economics (such as behavioural economics) that depart from basic neoclassical assumptions.

Practical implications

A practical implication of the paper is that it shifts attention to previously un-emphasized aspects of neoclassical thinking.

Originality/value

The paper's value to explain the relation of economics to ethics in neoclassical economics in connection with functionalist philosophy of mind.

The awkward position of neoclassical economic theory in regard to the relation between economics and ethics is excellently explained and discussed in [Hausman and McPherson's \(1996\)](#) recent volume in the Cambridge Surveys of Economic Literature series, *Economic Analysis and Moral Philosophy*. First, consider the paradox at the theory's very core. Though the theory is always claimed to be positive and value-free, in fact its definition of well-being or welfare exclusively in terms of subjective preference satisfaction, requires making a significant (and questionable) value-judgment, since well-being or welfare is also commonly defined in terms of a host of additional normative matters that the theory suppresses, such as justice, fairness, rights, liberty, and dignity (pp. 71; also cf. [Sen, 1987, pp. 47](#)). This position is now generally referred to as welfarism ([Sen, 1987, pp. 38-40](#)). Second, consider the form that policy recommendation takes, namely Pareto efficiency judgments based on preference satisfaction. Not only are potential improvements of any significance which make absolutely no one worse off virtually non-existent in ordinary experience ([Hausman and McPherson, 1996, p. 88](#)), but the idea of subjective preference satisfaction itself is unworkable in the real world for a variety of reasons: when preferences change, we don't know whether to respect old or new preferences; market outcomes are both determined by and change preferences; preferences are often manipulated and coerced; preferences depend on beliefs, and people often have false beliefs; individual preferences often conflict; preferences may be malevolent, anti-social, irrational, and immoral (pp. 75). Note also that attempts to identify "real," higher, well-informed, or otherwise appropriately laundered preferences, both undermine the basic idea behind subjective preference satisfaction, and simultaneously require recourse to additional normative considerations. Thus it seems that the theory cannot modify its conceptual basis without essentially abandoning it, and yet cannot make practical recommendations on that basis either.

Hausman and McPherson conclude from these problems and others that neoclassicism needs to develop alternative normative frameworks not only for the purpose of evaluating social policy, but even, they emphasize, to be able to formulate appropriate questions for positive research in economics. However, it seems there is little reason to expect any such development soon, since there is hardly any interest among most contemporary economists in re-evaluating the preference satisfaction paradigm upon which the discipline's normative stance rests. Why is this? One answer is that economists since the time when [Marshall \(1885/1956\)](#) changed the name of the discipline from political economy to economics have wanted to be thought of as scientists, and have believed that all science must be a value-free enterprise on the model of

natural science. The preference satisfaction paradigm subsequently appeared to be the most adequate foundation on which to build a purportedly value-free science of economics. Alternatively, it might be argued that attachment to the preference satisfaction paradigm came first, and that this attachment led economists to their subsequent stance on the relation between facts and values. A third, quite likely possibility is that the two conceptions developed jointly. Procedurally, however, I think that before this last, more complex case is explored a better understanding of the second alternative is needed, since while the first answer is entertained by many economic methodologists, few have made a case for saying that attachment to the preference satisfaction paradigm gave an important impetus to the development of positivism and welfarism in contemporary economics. This paper consequently makes a case for this proposition as a step in the direction of understanding the paradoxical position of contemporary neoclassicism on the relation between economics and ethics.

The main argument made below that the preference satisfaction paradigm underlies contemporary welfarism and positivism in economics is ultimately an argument about beliefs that economists have implicitly come to hold with respect to modern philosophy of mind. In a word, it will be argued that from the array of contemporary philosophies of mind the view that neoclassical economists have come to hold about economic agents is *functionalism*. As will emerge, however, my argument that this is the case depends in good part upon showing that functionalism should be understood as the destination at which neoclassicism has arrived in good part because of its earlier philosophical commitments, specifically commitments to the logical positivist views of ethics. Indeed, functionalism is a comparatively recent doctrine first coined in connection with the analysis of artificial intelligence by the [Philosopher Hilary Putnam \(1966/1975\)](#), whereas preference satisfaction as a specifically paradigmatic component of contemporary neoclassical theory dates back to [Robbins' \(1935/1938\)](#) critique of interpersonal utility comparisons, which heralded the beginnings of the collapse of cardinal utility theory with its richer normative framework, and ushered in ordinalist methods and Pareto efficiency judgments. Thus, this paper argues that positivism and welfarism in contemporary neoclassicism initially derive from a logical positivist view of ethics that has since developed into an implicit commitment to a functionalist philosophy of mind of the economic agent.

The Section 1 of this paper provides a discussion of the formation of the preference satisfaction paradigm particularly as it concerned the orientation economists took on the fact-value distinction after Robbins. The philosophical position concerned is the logical positivist, meta-ethical doctrine of emotivism, and the argument of this Section 1 is that an attachment to emotivist doctrines from the time of Robbins' critique of interpersonal utility comparisons laid the basis for both welfarism and the strong view of the fact-value distinction in contemporary neoclassicism. Section 2 then outlines how, given these emotivist commitments, the preference satisfaction paradigm may also be understood as inherently functionalist. This is important because functionalist philosophies of mind assume there are no significant differences between the human mind and artificial intelligence. My assumption is that though artificial intelligence shares some features of human intelligence, it lacks other particularly important ones. Thus, this section also argues that the conception of the agent in neoclassical economics is essentially that appropriate to artificial intelligence as opposed to distinctively human intelligence. In effect, rather than model artificial intelligence on human intelligence, the argument here is that the human intelligence of neoclassical economic agents is increasingly being modelled on artificial intelligence. Section 3 of the paper then argues that those elements of human intelligence excluded from a functionalist conception of mind and the neoclassical understanding of economic agents are central to an adequate understanding of the relation of economics to ethics. Problems with functionalist views of mind from the perspective of this richer view of human intentionality are briefly described and linked to neoclassical welfarism and positivism, and finally Section 4 comments on the marginalization of ethics in neoclassical economics and the relation of economics to ethics.

1 The preference satisfaction paradigm and emotivism

The separation of facts and values in contemporary neoclassicism is generally associated with the proposition that descriptive economic analysis is essentially positive and value-free, but it can also be looked upon from the perspective of what the fact-value separation implies about the realm of values and ethics. Specifically, that the domains of facts and values are separate implies that ethical discourse is not a fact-stating discourse, and that accordingly the statements or propositions of ethics cannot be thought to be true or false. Thus, on this view it cannot be said that it is true (or false) that it is wrong to kill another human being without reason, though it can be said that it is true (or false) that some person believes, feels, or has the opinion that it is wrong to kill another human being without reason. The latter statement has a truth value depending upon whether the person in question has the relevant belief, feeling, or opinion. But for neoclassicists the former statement cannot be true or false, because, on the any strong separation of facts and values, ethical statements simply lack truth values. In the logical positivist doctrines regarding ethics of the 1930s ethical statements were thought to lack truth values, because they were nothing more than expressions of feeling or emotion. This view – a form of non-cognitivism in (meta-)ethics – came to be known as emotivism ([Ayer, 1936/1946](#); [Stevenson, 1937](#)). [Davis \(1990, 1994, pp. 50\)](#) have argued previously that there are strong affinities between logical positivist emotivism and Robbins=ordinalism, and will not repeat those arguments here. Rather here I concentrate on how the ordinalist critique of interpersonal utility comparisons reinforced the fact-value dichotomy and welfarism in neoclassicism after Robbins.

That the chief impetus for neoclassical economists to abandon cardinal utility analysis in favour of ordinal utility analysis came from Robbins' critique of a strategy central to normative recommendation in the cardinal framework should not be overlooked. Cardinal utility theory generally judged ethically superior states of affairs to be those producing the most measurable utility, but because change invariably affects individuals' utilities differently, practical determination of those states typically involved utility tradeoffs that employed interpersonal comparisons of utility. Robbins, of course, rejected interpersonal utility comparisons by arguing that utility was not measurable. Accepting this meant that any statement that compared one person's utility with another's (such as the gain in one person's utility exceeds the loss in another's) could not have a truth value. On the surface, however, saying that positive statements of this sort lack truth value does not necessarily imply that ethical statements about how the world might be morally better as a result of policies meant to increase utility also lack truth value. It could be true (or false) that the world is morally better when utility is higher, even though we lack methods for measuring utility. Saying that utility is not measurable does not by itself imply a commitment to the non-cognitivist view that ethical statements lack truth values. But Robbins in fact believed that utility was not measurable, because it was subjective in nature. Different individuals' utilities in his view were private and incommensurable, and there accordingly were simply no facts about utility in general or across individuals which could figure in ethical statements. From this perspective, ethical statements cast in utility terms cannot be true or false, and one can consequently never say that it is true (or false) that one state of affairs is morally better than another.

This conclusion may strike some as too strong, since Pareto efficiency judgments, which ordinalism regards as a legitimate part of normative economics, seem to make reference to states of affairs that are regarded as superior to Pareto inefficient states of affairs as a matter of fact. How can it be, then, that ordinalism – even though it supposes there are no facts about cardinal utility recommendations – is non-cognitivist in connection with its view of Pareto efficiency judgments? Let us consider what Pareto efficiency judgments involve. When some state of affairs is judged Pareto superior to some other state of affairs, the Pareto superior state makes at least one person better off and no one worse off. This one or more individuals made better off are better off in their own estimation, i.e. they prefer the Pareto superior state. But there is nothing in this that implies that they are in fact better off, since the judgment that the Pareto superior state is preferred is entirely subjective. Using

the distinction above, we can say it is true (or false) that the individuals involved believe themselves better off, but we cannot say that it is true (or false) that they are better off. Put differently, positive statements about individuals=preferences can have truth value, but ethical statements about those preferences cannot have truth values. Pareto efficiency judgments, then, do refer to states of affairs that may hold as a matter of fact, but only because they are not a part of any ethical discourse. Moreover, while economists may independently say that Pareto superior states of affairs ought to be adopted, according to their own theory the recommendation that this occur represents only their own preference. Thus strictly speaking ordinalism as traditionally conceived since Robbins lacks any normative economics whatsoever, despite the fact that many economists see Pareto recommendations as having this role. In effect, economics has no relation to ethics.

It seems fair to say, then, that ordinalism and the preference satisfaction paradigm remains non-cognitivist with respect to ethical discourse. That the form of non-cognitivism involved is emotivism which may be less clear. Emotivism – the view that ethical statements are expressions of feeling or emotion – needs to be distinguished from another form of non-cognitivism, prescriptivism, which is the view that ethical statements are prescriptions ([Hare, 1963](#)). Is ordinalism possibly prescriptivist? While preferences, or acts of preferring, might be thought prescriptions in regard to the recommending of Pareto states, the main use of the concept of preference to explain individual choice seems closer to the idea of desire (cf. [Hollis and Robert, 1993](#)). Though more might be said to justify this conclusion, here I simply take preferences to be expressions of feelings. It is important, however, to take this characterization in a fairly narrow sense. The concept of a feeling *per se* obviously involves many dimensions, most of which, most would agree, are excluded from the neoclassical idea of a feeling as a preference. Specifically, following Robbins' and more recently [Arrow's \(1951/1963\)](#) rejection of interpersonal utility comparisons, any notion of intensity of feelings or preferences needs to be excluded from the concept as it is presently employed. Preferences on the standard view involve determinate relations between the individual and things in the world. Allowing that preferences come in varying intensities would fundamentally complicate these relations, making it virtually impossible to formally map out set relationships between economic agents. Preferences, it might thus be said, can have very little mental or intentional depth in contemporary neoclassical thinking. They are rather conative states of the individual, more representing ways in which individuals tend to respond to developments that impact upon them (such as changes in prices and incomes) than ways in which individuals register experience.

The roots of the fact-value distinction and welfarism in the preference satisfaction paradigm thus arguably have roots in neoclassicism's attachment to logical positivist emotivism. For simplicity I put aside here the question of whether neoclassical economists may have been disposed before Robbins to this attachment, based on an interest in having economics be seen as scientific on the model of natural science. Rather in what follows I go on to concentrate on the relationship between emotivism and functionalism, the postwar philosophy of mind originally associated with Putnam and other philosophers who hoped to demonstrate the identity of mind and brain. Accordingly the purpose of the discussion in Sections 2 and 3, respectively, is to attempt to show two things:

1. that emotivism contained important presuppositions for the development of a functionalist philosophy of mind for agents in neoclassical economics; and
2. that an attachment to functionalism further reinforces the neoclassical understanding of the relation of economics to ethics in the form of the theory's commitment to welfarism and positivism.

2 Functionalism, neoclassicism, and artificial intelligence

Functionalism is the view in contemporary philosophy of mind that defines mental states solely in terms of their causes and effects. It is the view that mental or intentional states are not intrinsic properties of the mind, such

as might be involved in the qualitative experience of pain, but are rather sets of causal relations to both those causes that bring mental states about (such as what might cause a pain) and the effects those mental states are apt to bring about in individuals' consequent behaviour (such as crying out). A mental state is thus just a cause-effect connection in the world that happens to involve beings with minds/brains. Indeed, as a cause-and-effect theory, a conscious motivation behind the development of functionalism was to provide a materialistic account of mind which could contest immaterialist philosophies of mind in the Cartesian tradition. However, despite the general appeal of this type of approach to many postwar philosophers, functionalism is also generally agreed to be unable to account for the qualitative nature or phenomenal characteristics of mental states, or more broadly consciousness, which some philosophers argue is inseparable from human intentionality ([Searle, 1992](#)). Not only are mental states defined solely in relational terms, so that the feel or qualia of mental states is irrelevant to their analysis, but it is possible to imagine cases in which two individuals have the same functional states yet entirely different qualitative mental states ([Block and Fodor, 1972](#); also cf. [Block, 1978](#)). This seems to be counter-intuitive and good reason to suppose that functionalism has serious limitations as a complete philosophy of mind.

Yet more provocative is the functionalist view that mental states exhibit what is termed multiple realizability. In explaining mental states in terms of their causal roles, i.e. according to their function or as functional states, the view distinguishes between a mental state's causal role and the occupant of that mental state. Thus, in principle the same mental or functional state could occupy different types of beings, including humans and forms of artificial intelligences such as Turing machines and other computational devices. Indeed, Turing machine functionalism, as originally developed by [Putnam \(1975\)](#) in light of the development of computers, contained many of the view's initial insights, which were later generalized to other varieties of functionalism. This generates an interesting tension for functionalism between the progress the view arguably makes in incorporating mental states into a general cause-and-effect nexus and the conclusion, unappealing to many, that the human mind is not different in any important respects from inorganic minds. But before addressing this matter, a more immediate question for the discussion here is whether ordinalist neoclassical theory operates with a functionalist conception of the economic agent.

Turning to the first of the two issues at the end of the last section, then, can it be shown that neoclassicism's emotivism predisposes the theory toward operating with a functionalist conception of the neoclassical economic agent? Emotivism takes ethical statements to be expressions of feeling, and was suggested above to be associated in neoclassicism with a narrow understanding of preferences that excludes any notion of preference intensity. Essentially, the argument was that development of ordinalism and the preference satisfaction paradigm followed upon Robbins' critique of interpersonal utility comparisons, and that this critique depended upon a subjectivist conception of preferences (and a non-cognitivist view of ethical statements). Eliminating any notion of preference intensity, however, removes an important dimension of human intentionality from our understanding of economic agents. As already noted, on the standard preference satisfaction view, a preference is simply a relation between an agent and the objects to which the agent's preference applies (agent A prefers object x to object y). Since this relation does not contain information about how strongly a preference is felt, it makes no use of the qualitative or phenomenal characteristics of preferences understood as the "feel" or experiencing of those mental states. Indeed, there is nothing in the idea of preference satisfaction that allows one to say that the agent has a conscious sense that one object is preferred to another. Object x is simply preferred to object y, period. Even allowing agents to have meta-preferences only introduces another preference relation, not the idea of the agent's preferences having any qualitative or phenomenal features associated with their being experienced or felt. Thus, the preference satisfaction paradigm appears to be a type of functionalist philosophy of mind, at least in regard to explaining preferences on the model of mental states that are strictly relational in character, and that ignore the feel or conscious nature of mental states.

This says nothing, of course, about agent's beliefs as the other type of mental states ascribed to economic agents. Are these too understood as essentially relational in the manner of preferences? Here the preference satisfaction paradigm offers little guidance, since the principal emphasis in the theory is on the nature and organization of agent preferences. Indeed, the theory maintains a sharp distinction between preferences and beliefs based on the idea of a subjective preference as something that is not reflected upon or otherwise reasoned about. Economic agents do have beliefs about probability distributions of future events and about the characteristics of goods. These beliefs might be termed informational beliefs. Agents also have beliefs about the way economic institutions and markets are thought to work, as required, for example, by the rational expectations hypothesis. These beliefs might be termed theoretical beliefs. But both types of belief seem to be understood solely as learned cognitive content, and thus as a sort of mental states that may be fully characterized as functional states of the agent. While in the absence of a more developed account of beliefs in the theory, it would be too strong to say that the theory is clearly functionalist in this regard to both preferences and beliefs, that the theory does not offer a clearly non-functionalist view of belief suggests that it sees this type of mental state in the same terms as preferences.

Returning to the nature of preferences as mental states, it seems fair to say that the preference satisfaction approach also shares the materialist cause-and-effect understanding of mental states that additionally defines functionalism. First, agents act when a causal event in the form of a change in prices or income produces a preference for a set of objects previously not preferred. Second, their action in choosing that new set of objects is the effect of the original causal event, the mental state of preferring the new objects to other available objects, and whatever relevant beliefs agents hold about the situation. This seems clearly to imply that agents' preferences are mental states in the functionalist sense. Though on the contemporary axiomatic understanding of preferences there is a formal structure to an agent's preferences, this syntax, as it were, only accounts for the form of agents' preferences in response to causal events. It does not account for the act of preferring a new set of objects over other available objects, which only occurs subsequent to price and income changes. Also, while agents' beliefs on the occasion of a new preference are not typically formed by these same causal events, that these beliefs seem to be nothing more than learned cognitive content derived from acquaintance with a history of such events suggests that they too may be understood to be functional states of the economic agent.

How, then, does this all bear on the relationship between artificial intelligence and the conception of decision-making in economics, a subject of growing interest in recent decades (cf. [Sent, 1998](#))? Perhaps the main approach, dating from [Simon's \(1969\)](#) early work on artificial intelligence, relies on the idea that intelligence is little more than the ability to serially process symbols in a rule-governed manner. Putting aside the limitations this involves associated with exclusive attention to serial processing, a fundamental difficulty with this conception is that it excludes the idea of understanding from intelligence. [Searle \(1980\)](#) has made this point especially clear in his Chinese room argument. Suppose that we designed an intelligent system that perfectly simulated some human cognitive capacity, such as the capacity to translate English into Chinese. Were we to give an individual who did not know Chinese but who knew how to operate that system questions to be translated into Chinese, though that individual would indeed be able to perfectly answer English questions in Chinese, we still would have to allow that he or she did not actually understand Chinese. Understanding, then, is something distinct from symbol processing, and though systems that simply process symbols can be very useful, they should not be assumed to fully replicate human intelligence. Put differently, there is something more to intelligence than purely formal or syntactical structures. Minds also have semantic contents that we represent to ourselves when we say we understand what something means. Meaning something is a dimension to many mental states, then, but nothing in the view of mental states as functional states, whether for human or artificial intelligences, requires any discussion of meaning.

Now it would be a mistake to say that functionalism everywhere implies a commitment to a theory of intelligence modelled on artificial intelligence, much less to any particular approach to artificial intelligence. And it would also be incautious to say that the tendency toward functionalist philosophy of mind in neoclassical economics commits all neoclassical economists to the idea that human intelligence is best modelled on artificial intelligence. But it is not unhelpful to point out that the development of the preference satisfaction paradigm through emotivism to functionalism tends to eliminate those elements in human intentionality that are the least tractable for theories of artificial intelligence. It might be better to put this way: the development of the subjective preference conception of the individual offers very little in the way of an account of forms of consciousness and understanding that are not immediately implicated in an analysis of individuals as cause-and-effect agents suitable to artificial intelligence. With this in mind, let us turn to the second of the two issues identified at the end of Section 1: whether attachment to a functionalist philosophy of mind reinforces ordinalist neoclassicism's welfarism and positivism.

3 Functionalism, welfarism, and positivism

Welfarism, recall, is the view that the only things to be considered in neoclassical normative argument are individual utilities as registered by subjective preferences ([Sen, 1987, p. 40](#)). Since, as Hausman and McPherson show, very little can be said about ethics in terms of individual preferences alone, it is instructive to briefly consider how neoclassical economists have attempted to develop the subjective preference paradigm to produce a broader normative vision. One such avenue is to employ an analysis of interdependent preferences or utility functions to explain social principles that appear to have moral significance, especially well-recognized phenomena that do not fit comfortably within the subjective preference paradigm but often have economic importance, such as altruism. Essentially, individuals are said to be altruistic if they have preferences for others' well-being ([Stark, 1995](#)). Alternatively, it might be said that individuals who have preferences over others' well-being exhibit sympathy for others, and this suggests that some preferences are associated with moral motives. Against this as a view of altruism, however, is the fact that having a preference for another's well-being must still be understood as being the interest of the individual having that preference. This seems at odds with our ordinary understanding of altruism as something which does not presuppose own interest. Relatedly, one might argue that if sympathy is to be understood as arising out a moral motive, it is better understood as empathy, an interpersonal capacity to "stand in another's shoes" (cf. [Fontaine, 1997](#)). Yet, the idea of taking up another's position hardly seems compatible with the subjective preference framework either.

Note, then, the fundamental difficulty for neoclassicism in attempting to develop the preference satisfaction paradigm in any way that makes use of the idea of an interpersonal comparisons. As [Hausman \(1995\)](#), when one operates within the subjective preference framework, interpersonal comparisons of well-being need to be formulated as interpersonal comparisons of the extent to which different individuals' preferences are satisfied. But there seems little reason to think that any sense can be made of the idea of comparative preference satisfaction. Alternatively, empathy as a capacity to put oneself in another's place has the connotation of involving more than an orientation toward another, as might be expressed in the form of a preference (A prefers that B be happy rather than sad), since the idea of empathetic identification with another draws on the empathizer imagining just how the other feels. That is, empathy extends beyond sympathy in adding to the notion of having an orientation towards another a sense of also having a felt concern for the other. Of course it is just this additional layer of intentionality that subjective preferences exclude on the traditional view since Robbins. A set of preference rankings, it might be said, is only a syntactical device or formal structure for processing symbols. In contrast, having a felt concern for others rather refers us to what a given state of affairs means to another. Indeed, it does not seem that one can explain the idea of putting oneself in another's place without employing the concept of meaning, since putting oneself in another's place explicitly requires a change of perspective regarding how things are understood.

Welfarism, then, is limited in regard to the range of normative phenomena it may accommodate, because as an analysis solely in terms of preferences it excludes important dimensions of our ordinary experience that we make use of in making normative judgments. From the perspective of the subjective preference paradigm, one cannot speak coherently of the intensity or depth of preferences. But this excludes any treatment of the qualitative or phenomenal character of mental states, as would be needed were we to rely on an interpersonal capacity to understand how others feel when we make normative judgments that take us beyond the Pareto framework. Thus, the functionalist development of the preference satisfaction paradigm as a syntax of preferences closes off neoclassicism from serious ethical argument. This reinforces the positivist-emotivist character of the theory that emerged in Robbins' critique of interpersonal utility comparisons. Looking upon welfarism in an emotivist light as a set of value-free, positive descriptions of possible preference patterns (for example, that one pattern of preferences can be characterized as Pareto superior to another), with a parallel account of ethical discourse as purely expressive, means that welfarism actually lacks any significant normative content whatsoever. The functionalist development of the preference satisfaction paradigm in the form of welfarism does nothing to alter this situation.

4 The marginalization of ethics in neoclassical economics

At the beginning of this paper, I distinguished two possible historical accounts of why neoclassical economics has increasingly marginalized ethics. The more familiar account is that economists followed Marshall's lead in seeking to make economics value-free on model of the natural sciences. On this view, the preference satisfaction paradigm was subsequently found to be the best vehicle for developing this conception of economics, so that the preference satisfaction paradigm and its commitments to welfarism and positivism are largely the effect of *a priori* disposition on the part of economists to emulate the natural sciences. The alternative account argued for above is that it was the development of the preference satisfaction paradigm that gave the chief impetus to the development of welfarism and positivism in contemporary neoclassicism. On this view, Robbins' critique of interpersonal utility comparisons initiated the turn away from any meaningful understanding of ethics and economics relation to it, and laid the foundation for the strong separation of facts and values that currently characterizes the discipline. Of course there may be truth to both accounts, but in closing I will simply comment on the nature of the motives involved in each account, since there are interesting differences between the two cases in this regard that bear on the issue of the marginalization of ethics in economics.

Note first that in both cases the desire to produce a more scientific discipline lies in the foreground. When Marshall re-labelled political economy economics in 1885, and when Robbins argued against interpersonal utility comparisons in the 1930s, both believed they were helping economics become a more scientific enterprise. But Marshall was principally concerned with damage done to the reputation of the discipline on account of its internal controversies at the end of the 19th century (cf. [Groenewegen, 1995](#)). Not only was there the polar opposition between Ricardo's supply-side legacy and Jevons' new marginal utility contributions, but the Ricardian socialists and Marx (and perhaps the English Historical school as well) had in his view done much to discredit the discipline. Essentially, then, Marshall was concerned with the perceived integrity of the field *per se*, especially as he initiated his efforts to have the economics tripos established at Cambridge. In contrast, Robbins' efforts to elevate economics' status as a science were specifically tied to reconceptualizing the psychology of utility and thereby eliminating a particular set of normative recommendations from economics, namely those made possible using interpersonal utility comparisons. Thus whereas Marshall was concerned that the field as a whole might be thought to be in professional disarray, Robbins' rather believed economics was in need of reform where it came into contact with ethics. Unlike Marshall who was not directly concerned with the relation of economics to ethics in his effort to elevate economics as a science, Robbins' effort to change the status of the discipline depended directly upon changing its relation to ethics.

Robbins, of course, represented his arguments against interpersonal utility comparisons as being based on irresolvable ambiguities in the concept of utility, and it may well be that he was not motivated by any special desire to eliminate the egalitarian types of policy recommendations from economics that interpersonal comparisons allowed, or narrow the space in which ethical discourse could operate in relation to economics. Nonetheless, the attachment of subsequent neoclassical economists to the preference satisfaction paradigm has had both of these effects, so much so that it is difficult not to draw the conclusion that the scientific status of economics today has been achieved by lowering the status of ethics and by economics' distancing itself from ethics. Most economists are not disturbed by this, because, it seems, they believe that this distancing ensures rather than undermines the scientific status of economics. But Hausman and McPherson believe that the poor relation of economics to ethics even jeopardizes economics' more positive projects, i.e. undermines its status as a science. For a number of reasons they believe that economics is at least partly a moral science, that the ideal of a value-free economics is seriously misconceived, and that economists need to have a good understanding of society's moral questions to know how to formulate worthwhile research questions. Unfortunately, their arguments will likely influence few. More likely to have the result they seek is a continued marginalization of economics as a discipline perceived to have little to contribute to answering real questions confronting society.

References

- Arrow, K. (1963), *Social Choice and Individual Values*, 2nd ed., 1st ed., 1951, Yale University Press, New Haven, CT.
- Ayer, A.J. (1946), *Language, Truth and Logic*, 2nd ed., 1st ed., 1936, Dover, New York, NY.
- Block, N. (1978), "Troubles with functionalism", *Minnesota Studies in the Philosophy of Science*, Vol. IX, University of Minnesota Press, Minneapolis, MN, pp. 261-325.
- Block, N. and Fodor, J. (1972), "What psychological states are not", *Philosophical Review*, Vol. 81, pp. 159-81.
- Davis, J. (1990), "Cooter and Rappoport on the normative", *Economics and Philosophy*, Vol. 6, pp. 139-46.
- Davis, J. (1994), *Keynes's Philosophical Development*, Cambridge University Press, Cambridge.
- Fontaine, P. (1997), "Identification and economic behavior: sympathy and empathy in historical perspective", *Economics and Philosophy*, Vol. 13, pp. 261-80.
- Groenewegen, P. (1995), *A Soaring Eagle: Alfred Marshall*, Edward Elgar, Aldershot, pp. 1842-924.
- Hare, R.M. (1963), *Freedom and Reason*, Oxford University Press, Oxford.
- Hausman, D. (1995), "The impossibility of interpersonal utility comparisons", *Mind*, Vol. 104, pp. 473-90.
- Hausman, D. and McPherson, M. (1996), *Economic Analysis and Moral Philosophy*, Cambridge University Press, Cambridge.
- Hollis, M. and Robert, S. (1993), "Rationality in action", *Mind*, Vol. 102, pp. 1-35.
- Marshall, A. (1956), "The present position of economics", in Pigou, A.C. (Ed.), *Memorials of Alfred Marshall*, first published 1885, Kelly and Millman, New York, NY.
- Putnam, H. (1975), "The mental life of some machines", *Mind, Language and Reality: Philosophical Papers*, first published 1966 Vol. 2, Cambridge University Press, London.
- Robbins, L. (1938), *An Essay on the Nature and Significance of Economic Science*, 2nd ed., Macmillan, London.

- Searle, J. (1980), "*Minds, brains, and programs*", *Behavioral and Brain Sciences*, Vol. 3, pp. 417-24.
- Searle, J. (1992), *The Rediscovery of Mind*, MIT Press, Cambridge, MA.
- Sen, A. (1987), *On Ethics and Economics*, Basil Blackwell, Oxford.
- Sent, E-M. (1998), "*Artificial intelligence*", in Davis, J., Hands, W. and Mäki, U. (Eds), *The Handbook of Economic Methodology*, Edward Elgar, Aldershot.
- Simon, H. (1969), *The Sciences of the Artificial*, MIT Press, Cambridge, MA.
- Stark, O. (1995), *Altruism and Beyond*, Cambridge University Press, Cambridge.
- Stevenson, C.L. (1937), "*The emotive meaning of ethical terms*", *Mind*, Vol. 46, pp. 10-31.