Normative analysis asks how we (as a society) should make trade-offs between individuals. Behavioral welfare economics extends the scope of this analysis to a single individual, asking how the individual should trade off potential motives. We argue that any faith in economists’ (behavioral or otherwise) abilities to resolve such problems in moral philosophy cannot be based on past accomplishments of welfare economics, and that the best way to understand welfare economics is to view it as a part of positive economics with no normative content.

We distinguish three types of welfare analysis. We refer to them as welfare I, II, and III. Welfare I identifies Pareto efficient outcomes within a particular model and relates those to equilibrium outcomes. Welfare I includes results such as the first theorem of welfare economics or the Myerson and Satterthwaite theorem on two-sided bargaining—results that show when an efficient outcome must, or cannot, be attained in equilibrium. Such results are used to explain the persistence of certain economic institutions (such as competitive markets) or derive limitations on what can be achieved within any institutional framework (such as bargaining or auctions).

Welfare II posits an objective function for a policymaker and evaluates different policies or institutions according to that objective function. In other words, welfare II analyzes the implications of a policymaker’s preferences. Policymakers often pursue sophisticated objectives that cannot be captured by the standard functional forms used to describe simple, self-interested economic agents. Moreover, evidence regarding policymakers’ preferences is typically limited. Hence, discussions of policymaker preferences tend to be somewhat speculative and typically fall outside standard utility theory.

Welfare III tries to identify, advocate, or show the impossibility of normative criteria for the aggregation of preferences. While welfare II asks “which of these two options might the policymaker choose?” welfare III asks “which of these two options should the policymaker choose?” Since the actual preferences of the policymaker are of no relevance for welfare III, the policymaker may be eliminated from the discussion and the appropriate normative criterion may be attributed to an idealized “we” or “society.” For the purpose of this essay, the term “normative economics” is synonymous with welfare III.

We will refer to the following position as the behavioral welfare thesis—revealed preference is an inadequate tool for uncovering true preferences. (For example, we would not use it for infants or the mentally ill.) We know that even adults make mistakes, engage in dysfunctional behavior, suffer from biases, etc. Given these problems, true utility cannot be identified from revealed preference. Understanding physiological and psychological mechanisms can help us distinguish mistakes and true utility. The behavioral welfare thesis is concerned with welfare III. It distinguishes true utility, i.e., the appropriate measure of happiness and choice utility, the determinant of behavior, and concludes that we cannot determine which actions are likely to increase others’ happiness by observing their behavior. Identifying and quantifying true utility is important not only as a prerequisite for trading off one individual’s happiness against another’s, but also for determining the right course of action when no such trade-off exists. Behavioral economists view identifying welfare with choice to be an uncritical acceptance of libertarian moral philosophy.

Behavioral economists offer three supporting arguments for their thesis. They appeal to various psychological mechanisms, biases, and mistakes. They note that recent developments in neuroscience and psychology have made (or will make) the identification of true utility easier. Finally, they argue that even if these
developments do not lead to a determinate theory of true utility, developing criteria, based on some combination of the state of the art psychology and individual judgment, is essential because individuals often find themselves in positions to make decisions that will influence others’ well-being.

We are skeptical that economists will be able to come up with resolutions for age-old questions in moral philosophy regarding what constitutes true happiness. It also seems unclear how economists plan to bring the preferences of policymakers in line with the newly developed measures of happiness. Our key point, however, is independent of this skepticism. The vast majority of the welfare analysis of typical journal articles can best be understood as welfare I or II rather than welfare III, and unlike welfare III, neither welfare I nor welfare II relies on libertarian philosophy. For the purposes of welfare I and welfare II, a theory of true utility, whether it defines true utility as choice utility or not, is unnecessary. Thus, the potential role for a new behavioral moral philosophy in economics is limited in existing research, just like the role of libertarian philosophy.

I. Welfare I and Welfare II as Positive Economics

The most emphatic argument for the behavioral welfare thesis is the observation that human beings engage in behaviors that obviously conflict with their interests. Tourists get run over in London because they walk off the curb without looking. People become addicts, save too little for retirement, make computation errors, take unreasonable risks, and fail to carry out reasonable plans. To behavioral economists, all this suggests a need for a measure of happiness that is, to some extent, independent of revealed preference. Such a measure would have no role in welfare I or welfare II, however.

Welfare I, as defined in the previous section, is a tool for analyzing economic institutions and models. Pareto optimality is a stability test. If an institution is not optimal, then there are mutually agreeable changes. For example, economists have found that US farm subsidies are inefficient. Farm subsidies could be eliminated and farmers could be compensated in a way that would increase the economic welfare of all US households. Yet, farm subsidies persist. What are the mechanisms (political and economic) that lead to the persistence of inefficient policies? In this sense, welfare I is part of positive economics. Economists use strategic and competitive models to analyze behavior within particular institutional settings, while they use Pareto optimality (or the lack thereof) as a tool for predicting institutional stability (or change).

The template for welfare I analysis might be as follows: Pareto statements (farm subsidies are inefficient) are used to define new questions (what makes farm subsidies persist?) that lead to better models of the underlying institution. Welfare I is suitable for this task because the welfare of an agent is defined to be synonymous with choice. We can expect agents to gravitate toward welfare improving policies and institutions only if welfare is synonymous with what the agents perceive to be in their self-interest. True utilities that are distinct from choice utilities are of no use for analyzing and predicting such institutional change.

In welfare II, the analyst posits an objective function and asks which policy maximizes that function. For example, a monetary authority might care about the stability of macroeconomic variables. The monetary authority’s welfare function might be the negative of the deviation from target output and inflation levels. In this analysis, the economist infers or makes assumptions about the objectives of the decision maker. The “right” objective function for welfare analysis II is the objective function of the decision maker. The economist’s contribution is to find the mapping between objectives and optimal policies.

II. Welfare III

Introductory economics textbooks typically describe the distinction between positive and normative economics as follows.

Normative economics involves ethical precepts and value judgments. Should the government give money to poor people? Should the budget deficit be reduced by higher taxes or lower spending? Should the socialist countries introduce private property and stock markets? These issues can be debated, but they can never be settled by science or appeal to facts. There are no
right and wrong answers to these questions because they involve ethics and value judgments rather than facts. (Paul A. Samuelson and William D. Nordhaus 2005)

As Samuelson and Nordhaus describe above, welfare III corresponds to normative economics. It asks questions such as “should government give money to poor people?” or attempts to devise criteria for answering such questions. This is also the type of welfare economics that the behavioral thesis has in mind.

The history of an individual through time can be described as a succession of separate selves …. Which one of these selves should be granted authority over outcomes in the future? (Daniel Kahneman 1994)

Note that Kahneman’s question is similar to “should government give money to poor people?” The difference is that Kahneman sees a role for welfare III analysis even in situations where Samuelson and Nordhaus do not, and even when no interpersonal trade-offs are involved. Kahneman assumes that the identification of true utilities or criteria for trading off the welfare of the individual’s various selves is to be a part, perhaps central part, of behavioral economics.

Thus, the behavioral welfare thesis can be interpreted as two distinct assertions. First, normative economics as defined by Samuelson and Nordhaus (welfare III) is important, but its perceived libertarian premises must be rejected because individuals make mistakes and have other limitations that create a gap between behavior (revealed preference) and well-being (true utility). Second, behavioral welfare economics must identify policies that maximize the true utility of individuals. This may involve deciding which of the selves should be granted authority, identifying mistakes, or studying the brain to determine which processes are “misguided.”

A discussion of the intrinsic importance of welfare III, or the merits of any particular ethical theory, is far beyond the scope of this paper. Instead, we make the following simple points. First, welfare analysis of an overwhelming majority of economics research papers can be understood best as welfare I or welfare II and not welfare III. Second, for welfare I and welfare II, mistakes and other human limitations do not create the type of methodological problems that behavioral economists describe.

To see the first point, note that welfare analysis in the typical economics journal article entails one of the following two approaches: identifying Pareto efficient outcomes, or identifying the maximizers of a more or less arbitrary (social) welfare function defined as the weighted sum of arbitrary representations of individuals’ ordinal preferences or of consumer surplus.

We explain, above, how Pareto optimality is used in welfare I analysis. Of course, instead of the welfare I interpretation of Pareto optimality, we can interpret it as an ethical principle. Consumer sovereignty plus a minimal efficiency requirement to not make A worse off if doing so does not make anyone else better off. Even for someone who wholeheartedly embraces the underlying ethical principle, the minimal efficiency criterion is almost never decisive. One could not answer questions like “should the government give money to the poor?” with this criterion. For those who do not accept the underlying ethical principle, Pareto optimality as a criterion for welfare III is even less useful. Hence, interpreted as welfare III, the question of whether a particular economic model generates Pareto optimal allocations seems uninteresting. Only through its welfare I interpretation does Pareto optimality gain its central place in economic analysis.

Next, consider models that maximize a particular social welfare function. Here again, two interpretations are possible. The modeler may be using the particular social welfare function as an abstract moral criterion (welfare III) or as a representation of some policymaker’s (or median voter’s) preferences for analyzing what he would do (welfare II). As an exercise in welfare II, a social welfare function is a testable hypothesis. It may be verified or rejected by observing the behavior of the policymaker.

The welfare III interpretation must confront a series of questions addressed rarely in economic research. In what sense does writing down an arbitrary social welfare function (without any analysis or discussion) constitute an ethical argument? Does the modeler anticipate a challenge from readers or referees regarding his particular welfare function or the particular weights? What arguments would this challenge entail? What would compel or convince others with different preferences over social outcomes to abandon
their in favor of the welfare function presented in the paper? Who do we have in mind as the potential consumer of this moral philosophy? As an exercise in welfare III, positing a welfare function is a moral judgment without an argument.

There are some, but relatively few, examples of economics papers that undertake normative analysis of the sort we have been calling welfare III. Axiomatic analysis of interpersonal utility comparisons and Arrow’s impossibility theorem can and perhaps must be viewed as welfare III. But work that distinguishes feasible (i.e., implementable) social objective functions from infeasible ones without offering any argument for any particular objective function, such as the work of Leonid Hurwicz (1979) and Eric Maskin (1999) and the literature that followed, is best understood as a part of welfare II.

A research paper with behavioral welfare III analysis is even more difficult to find (or imagine). What neurological or behavioral evidence can the author invoke to argue that behavior is heroic or reckless, a part of the human condition, or deviant? These are moral and philosophical questions not “economic” questions. The idea that neurological techniques can identify particular goals (of an individual or the entire species) that are suitable or unsuitable is itself a (rather unusual) philosophical position.

Suppose behavioral economists happen to agree on a measure of true utility. Suppose they conclude religious observance is a mistake and does not increase true utility, but drug consumption (in moderate doses) increases true utility in some people. Will this new understanding compel individuals (policymakers) to replace their choice utilities with the newfound true utilities? When Kahneman asks, “which one of these selves should be granted authority over outcomes in the future?” who does he have in mind as the authority-granting agency? As we argue above, economic research is engaged rarely in debates about ethics. Questions such as “what is a fair distribution of resources?” are not key questions in economics, nor are there well-established methods for tackling them. It is therefore difficult to see how research could emerge on what is a fair distribution of resources for a single individual over time, and on how the individual should be compelled to utilize these resources.

We conclude this section by noting that mistakes and other limitations do not create methodological problems for welfare I and welfare II. We can imagine two types of mistakes:

(a) Mistakes in which the objective is clear, but actions may fail to achieve it, such as accidents or wrong decisions that arise from incomplete information or temptation.

(b) Mistakes, in which the objective of the decision maker is questioned, such as dangerous consumption, addiction, and inconsistent choice.

For both welfare I and welfare II, mistakes of the first type simply add constraints (or costs) no different from technological or incentive constraints. We can expect that institutions emerge (such as insurance, monitoring, or commitment mechanisms) that deal with mistakes of the first type (welfare I), and that such institutions receive support from policymakers (welfare II).

Mistakes of the second type are irrelevant for both welfare I and welfare II. Economic institutions can arise only as a response to needs that individuals recognize. Any external assertion that drug consumption is bad is irrelevant for the addict and the dealer, and therefore for welfare I. Of course, the policymaker might be paternalistic, but this causes no dilemma for welfare II. The question is not what the policymaker ought to do but what policy options will achieve his actual goals. For example, a policymaker may impose a drug ban. Whether he does this because of religious convictions or an understanding of the brain mechanisms underlying addiction is irrelevant for welfare II.

III. Conclusion

One popular introductory economics text asserts:

Sometimes we want to go beyond explanation and prediction to questions such as ‘What is best?’ This involves normative analysis, which is also important for both managers of firms and those making public policy. (Pindyck and Rubinfeld 2004)
Normative analysis of this sort almost never comes up in more advanced texts or in research papers. Neither upper division undergraduates nor graduate students are told that managers and policymakers look to economics for help in understanding how to achieve goals or in identifying the “right” goals. Discussions of the economists’ role in defining criteria for trading off the well-being of one person for that of another or discussions of the economists’ obligation to ensuring that managers and policymakers adopt these criteria seem to be reserved for introductory textbooks and methodological essays.

Many behavioral economists assume the absence of such discussion in journal articles indicates these questions were, at least partially, resolved by economists’ acceptance of libertarian philosophy. They contemplate enlarging the economists’ role by rejecting or modifying this philosophy. Not only are economists expected to develop and advocate criteria for trading off one person’s happiness for that of another, they must also figure out and propagate criteria for deciding what is happiness. Thus, economists, in collaboration with neuroscientists, must develop principled arguments and criteria for declaring certain decisions as mistakes, even if the individual never recognizes them as such. Economists must identify the different selves of the individual and mediate among these selves. They must analyze the workings of the brain and figure out which processes lead to decisions that enhance true utility and which ones lead to behavior divorced from true utility.

The idea that individuals, including policymakers and managers, have preferences over others’ consumption and behavior is not new. The novel view (or at least the view gaining novel popularity in methodological discussions), is the suggestion that such preferences are fundamentally different from individuals’ preferences over their own consumptions and actions; that these preferences can be derived from moral philosophical and behavioral-neurological considerations; that behavioral economists are qualified and perhaps obligated to undertake this derivation; and that there is a constituency of policymakers, managers, and other influential decision makers ready, or at least willing, to be informed by behavioral economists’ findings and conclusions on these issues.

No matter how impressed we may be by recent developments in psychology and neuroscience, it is difficult to see how these developments have resolved or are likely to resolve age-old problems in moral philosophy about what constitutes true happiness. The standard response to this observation is that recent developments in psychology and neuroeconomics are so significant that now we may legitimately expect better answers to timeless questions in moral philosophy; that, in any event, making decisions, often paternalistic decisions, on others’ behalf is inevitable; and that such decisions are important, and therefore we should be eager to discuss how to make these decisions correctly.

In this essay, we have questioned neither the importance nor the necessity of such decisions. We have not criticized any particular proposal for answering them. We have argued only that any faith in economists’ (behavioral or otherwise) abilities to resolve such problems in moral philosophy cannot be based in the past accomplishments of welfare economics, and that the best way to understand welfare economics is to view it as a part of positive economics.

REFERENCES


