BOUNDED RATIONALITY AND LEGAL SCHOLARSHIP

Matthew D. Adler[+]

Any normative framework which has the structure of recommending that decisionmakers advance certain goals, and that they do so in accordance with decision theory (DT), runs into the problem of bounded rationality. The problem is how to refine DT so as to be usable by a bounded decisionmaker -- someone with limited cognitive resources, for whom the *full* evaluation of her choices is impossible or at least very expensive.

This chapter has two aims. The first, pursued in Parts I and II, is to discuss the problem of bounded rationality in general terms. The second, pursued in Part III, is to show why the problem creates a gap at the foundations of legal scholarship. In *Fairness versus Welfare* (*FW)*, Louis Kaplow and Steven Shavell propose a welfarist methodology for legal scholarship.[1] They may be wrong to think that morality is wholly welfarist, but a normative program which says that social welfare is *one* of plurality of criteria by which legal scholars should evaluate laws and policies is very plausible. But we have no good normative handle on how legal scholars who are bounded in their cognitive abilities should implement a welfarist or pluralistic program.

I. THE PROBLEM OF BOUNDED RATIONALITY

The term "bounded rationality" is sometimes used by psychologists to describe or explain certain decisionmaking processes without endorsing or criticizing them. By contrast, the problem of "bounded rationality" that I discuss in this chapter is normative. I am interested in the problem that human cognitive limits create for our attempts to flesh out normative frameworks.

A. *Instrumentalist Normative Frameworks and Decision Theory*

A normative framework furnishes recommendations for some class of decisionmakers. Some normative frameworks furnish *moral* recommendations, others non-moral recommendations (for example, recommendations concerning what the decisionmaker should do in light of her interests). In either event, let us call a normative framework "instrumentalist" if it is oriented around goals.[2] One such framework is

---

[+] Leon Meltzer Professor, University of Pennsylvania Law School. Thanks to Christoph Engel, Jeff Rachlinski, Adrian Vermeule, and Mark White for comments.

[1] Louis Kaplow & Steven Shavell, 2002, *Fairness Versus Welfare*, Cambridge: Harvard University Press.

[2] What exactly this means is a complicated question. Within moral theory, that question has been extensively discussed with reference to the distinction between "consequentialist" and "non-consequentialist" moral frameworks. For purposes of this chapter, I don't need to take a firm stand on the question. The crucial point is that some such frameworks, such as Humeanism, or welfarism or consequentialist moral frameworks more generally, *are* thus oriented and therefore naturally employ DT as their procedural component -- in turn leading to the problem of bounded rationality.

"Humeanism," which recommends to each individual that she pursue whatever preferences she happens to have. Another is welfarism, a moral framework, which recommends that decisionmakers promote the goal of social welfare.

An instrumentalist normative framework will not be very helpful unless it furnishes procedural norms for the decisionmakers within its scope, explaining how they should go about implementing the stipulated goals. Call these procedural norms the "procedural component" of the normative theory.

The argument I will be developing in the next two Parts is this: human cognitive limitations pose serious difficulties for using DT as the procedural component for an instrumentalist normative framework.

Understood as a set of procedural norms, DT enjoins the decisionmaker to think of any choice situation at some time t as an outcome set $\boldsymbol{O} = \{o_1, o_2, \ldots\}$, an action set $\boldsymbol{A} = \{a_1, a_2, \ldots\}$, and a state set $\boldsymbol{S} = \{s_1, s_2, \ldots\}$.[3] An outcome is a normatively relevant description of some way the world might be -- a description in terms of the goals of the normative theory that DT is fleshing out. Equivalently, an outcome is a proposition or a set of possible worlds. Actions are what the decisionmaker can do at t.

States, like outcomes, are propositions. But, in a well-framed description of a choice situation, states are causally independent of actions -- it should not be possible for an action to cause a state[4] -- and so each state is most naturally thought of as a possible past history of the world, together with causal laws. Further, in a well-framed choice situation, the states are mutually exclusive and collectively exhaustive. The decisionmaker does not know which state is the true one -- but one must be, only one can be, and the decisionmaker knows that. Finally, there are necessary connections between the states, the actions, and the outcomes, which again the decisionmaker knows. The conjunction of a state and an action entails one and only one outcome in the outcome set.[5]

<u>How DT Conceptualizes a Choice Situation</u>

States

s_1 s_2 s_3 …

[3] In my discussion of DT, I have relied heavily upon James M. Joyce, 1999, *The Foundations of Causal Decision Theory*, Cambridge: Cambridge University Press.

[4] See Joyce, *The Foundations of Causal Decision Theory*; James M. Joyce & Allan Gibbard, 1998,"Causal Decision Theory," in Salvador Barbera et al eds., *Handbook of Utility Theory, Vol. 1*, Dordrecht: Kluwer, pp. 627-66.

[5] This formulation assumes that the decisionmaker gives zero probability to the possibility of indeterministic causal laws. Refining DT for the possibility of indeterminism raises complicated issues that I cannot consider here. *See* David Lewis, 1981, "Causal Decision Theory," *Australasian J. Philosophy*, 69, pp. 5-30. But it is hard to see why DT, thus refined, would not confront problems of bounded rationality similar to those discussed below.

	a_1	$o(a_1, s_1)$	$o(a_1, s_2)$	$o(a_1, s_3)$
Actions	a_2	$o(a_2, s_1)$	$o(a_2, s_2)$	$o(a_2, s_3)$
	a_3	$o(a_3, s_1)$	$o(a_3, s_2)$	$o(a_3, s_3)$
	…			

Note: $o(a_i, s_j)$ is the outcome in the outcome set $\boldsymbol{O}$ that would result if action a_i were performed in state s_j.

DT also enjoins the decisionmaker to follow certain norms in ranking outcomes and actions. To begin, the decisionmaker is enjoined to have a *complete* and *transitive* ranking of the outcomes. Another such norm is *consequentialism*: if, in each state, two actions produce the very same outcome, then the decisionmaker should be indifferent between the two actions. Another is *dominance*: if one action produces no worse outcomes than a second action in all states, and better outcomes in some, then the first action must be ranked better.[6] Yet another is *independence*: If two actions produce the very same outcome o^* in some state of the world s', the very same outcome o^{**} in state s'', and so forth, and are ranked a certain way, then two actions which are identical to the first pair except that they produce $o+$ rather than o^* in s', $o{+}{+}$ rather than o^{**} in s'', and so forth, should be ranked the same way as the first pair.

DT is sometimes offered, not as a set of procedural norms, but as a set of substantive criteria for normatively appropriate choice.[7] In this guise, the theory takes an "as if" form. It is agnostic about the procedure decisionmaker actually uses, and instead says that -- whatever that procedure might be -- he should pick the choices which he would have selected, had he used DT as his procedure.

I don't know whether DT, in its substantive version, runs afoul of human cognitive limits. But, as already stated, any normative framework is incomplete if it lacks a procedural component. How helpful will it be to state some goal for some agent, but not tell her what to do to implement the goal? And DT, in its procedural form, clearly does run afoul of bounded rationality, as I will show in a moment.

But why even think, then, of using DT as the procedural component of a normative framework? The answer is that DT is far and away the best developed account of rational choice. DT meshes particularly well with instrumentalist normative frameworks. A goal is just a norm for identifying a set of outcomes that matter and ordering its elements. DT then tells the agent how to orient his choices towards these outcomes, given uncertainty, in a plausible way: by thinking systematically about how his possible actions and possible prior states might interact to cause different outcomes.

In addition, Leonard Savage showed that the choices of a decisionmaker who satisfies DT's norms (and several other more technical axioms) can be represented as

[6] More precisely, some of the states in which the first action produces better outcomes must be non-"null," i.e., have positive probability.

[7] See Frederic Laville, 2000, "Foundations of Procedural Rationality: Cognitive Limits and Decision Processes," *Economics and Philosophy*, 16, pp. 117-138.

maximizing the expectation of a utility function.[8] There is some probability function assigning numerical probabilities to states, and some utility function assigning numerical utilities to outcomes, such that the decisionmaker's ranking of actions corresponds to their expected utility (where the expected utility of an action is the sum of the utilities of its possible outcomes, each utility discounted by the probability of the state that, together with the action, produces that outcome). Reciprocally, if the decisionmaker consciously maximizes the expectation of a utility function, then she automatically satisfies the axioms of DT. So it is not too much of an exaggeration to say that DT and expected utility theory are one and the same account of rational choice.

B. *DT and the Specification of Outcomes, Actions and States*

What, then, is the problem of bounded rationality for DT? Humans are cognitively limited, in various ways. They have limited memories, limited computational abilities (even when aided by computers), make logical and mathematical errors, and so forth. But how does this implicate the framework for decisionmaking set forth by DT?

The key difficulty, as I see it, has to do with the *specification* of outcomes, actions, and states. The sets of outcomes, actions, and states can be *fully specified* or *incompletely specified*.

Start with outcomes. Remember that an outcome is a proposition or set of possible worlds. An outcome is "fully specified," for purposes of some instrumentalist normative framework, if any two worlds within the outcome are equally good in terms of the goals of that framework. In other words, a fully specified outcome provides a description of what might occur which is sufficiently complete that nothing of relevance to the framework is left out.

Consider, for example, a Humean framework which tells decisionmakers to maximize their preferences. Imagine that the decisionmaker, Jim, cares only about his annual income and the number of friends he has each year. Then if a given outcome *o* describes Jim's annual income in each year that he is alive, it is incompletely specified. If it describes his annual income and the number of his friends, it is fully specified.

A fully specified outcome set is a set of mutually exclusive and collectively exhaustive outcomes, each of which is fully specified. A fully specified action set consists of every possible action the decisionmaker could undertake at the time of choice. A fully specified state set consists of states which are (again) mutually exclusive and collectively exhaustive, and are sufficiently well described to do what state sets are supposed to do in DT: Each action, together with a state, entails one and only one outcome in the outcome set.

An unbounded decisionmaker can produce a fully specified state, action, and outcome set instantly and at zero cost (she can do so "internally" without mental strain

[8] For an accessible discussion, see David M. Kreps, 1988, *Notes on the Theory of Choice*, Westview Press: Boulder and London.

and without paying for expensive computers and software). DT is therefore most plausibly understood as providing the following specification norm for an unbounded decisionmaker: *Use a fully specified outcome, action, and state set*. Why is this so? Consider, first, the unbounded decisionmaker who uses an incompletely specified action set. In other words, she considers some of the actions that might perform at this moment, but not all of them. Isn't this arbitrary? The benefits of adding an action to the set of actions she already considers is that the new action might be preferable to those she already considers; the costs (given her unbounded abilities) are nil. So why not consider it?

Consider, next, the unbounded decisionmaker who uses a fully specified action set, but an incompletely specified outcome set. Where outcomes are incompletely specified, the basic norms of DT lose their intuitive plausibility. I will show why this is so, using the norm of consequentialism. A similar analysis applies to the other basic norms.

Let us continue with the example of Jim. $\boldsymbol{O'}$ is an incompletely specified outcome set, where outcomes are described just in terms of Jim's income, not his friends. Imagine that Jim reasons in terms of this outcome set; identifies a matching state set (sufficiently detailed to determine which member of $\boldsymbol{O'}$ would result, for every action he might perform at present); and ascertains that there are two actions, a and $a+$, which result in the very same member of $\boldsymbol{O'}$ in each state. In s_1, a results in o^* and $a+$ also results in o^*. In s_2, a results in o^{**} and $a+$ also results in o^{**}. And so forth for every state. The norm of consequentialism tells Jim to be indifferent between a and $a+$. But why should he be? After all, it is possible that - in some states -- a leads to a different number of friends for Jim than $a+$. And Jim cares about the size of his group of friends, not just his annual income. So why should Jim consider a and $a+$ as equally good, given that it is costless for him to replace $\boldsymbol{O'}$ with a fully specified set $\boldsymbol{O}$ whose members specify what happens along both the income and the friendship dimensions, and to think about his choice in terms of that outcome set?

Consider, finally, an unbounded decisionmaker who uses a fully specified action and outcome set, but an incompletely specified state set. This could mean two things. First, it might mean that the state set is insufficiently rich for a conjunction of an action and a state to always entail one and only one outcome in the outcome set. A choice situation, thus framed, is one in which the decisionmaker lacks relevant information about his choice -- and, in the case of the unbounded decisionmaker, this is a pointless gap, one which he could costlessly rectify through a more detailed consideration of possible prior histories of the world together with causal laws.[9] Second, an incompletely specified state set might not be exhaustive. Although each member of the action set, together with each member of the state set, leads to some determinate member of the outcome set, there is some possible state that the decisionmaker ignores. Were he to

[9] Again, this assumes that the decisionmaker gives zero credence to the possibility of indeterminism. See supra note 5.

consider that state, he might find that he accords it non-zero probability, and that adding it to the analysis changes his evaluation of the possible actions. So why not consider it?

C. *Specification and Bounded Decisionmakers*

It is therefore quite straightforward to determine the content of DT's specification norm, in the case of unbounded decisionmakers. Unbounded decisionmakers should use fully specified outcome, action, and state sets. The difficulty arises once we allow that it may be costly or plain impossible for the decisionmaker to produce a fully specified outcome, action, and state set. What specification norm should DT contain for this sort of decisionmaker -- a bounded one? The specification norm cannot be: Use fully specified outcome, action, and state sets. More precisely, given any instrumentalist normative framework with a sufficiently "rich" goal to generate a very large set of fully specified outcomes, DT coupled with a full-specification norm will be a highly implausible candidate to be the procedural component of this framework.

Consider the example of welfarism. For purposes of the general discussion of bounded rationality, this is simply one example of a normative framework that cannot be plausibly fleshed out as advising decisionmakers to use a fully specified outcome, action, and state set. For purposes of the methodology of legal scholarship, as we shall see in Part III, this is a *key* example.

Welfarism is really a family of moral theories, which tell the decisionmaker to aim at some mix of overall well-being and the equitable distribution of well-being. [10] The most plausible variants of welfarism understand individual well-being as the attainment of various "objective goods" or (what is essentially the same thing) as the attainment of those things everyone with full information would converge in preferring. The most plausible variants of welfarism adopt an impartiality constraint: everyone's well-being matters equally, or at least the well-being of everyone within a large population of interest does (all members of the society, for example). In characterizing outcomes, we shouldn't limit ourselves to the well-being of a single dictator or members of a small oligarchy, and if two outcomes differ only with respect to the identity (proper names) of individuals attaining particular welfare levels, they should be seen as equally good.

Consider DT as the procedural component of welfarism with an impartiality constraint and a reasonably extensive list of objective goods. Imagine, for example, a U.S. governmental official June who attends to the well-being of all current and future U.S. citizens and considers the following goods: longevity, consumption, health, happiness, and social relations. June presides over a pollution control agency and, at the present moment, has the power to regulate some toxin.

---

[10] *See, e.g.*, Matthew D. Adler & Chris William Sanchirico, 2006, "Inequality and Uncertainty: Theory and Legal Applications," *University of Pennsylvania Law Review*, 155, pp. 279-377; Matthew D. Adler & Eric A. Posner, 2006, *New Foundations of Cost-Benefit Analysis*, Cambridge: Harvard University Press, pp. 25-61.

A fully specified outcome set ***O***, for June, would be such that each o_i in this set consists of a list of individuals who are U.S. citizens now or in the future; a consumption history for each individual, specifying what goods she consumes for every moment she is alive; a health history for each individual, specifying her health state for every moment she is alive; a happiness history, specifying her affective states for every moment she is alive; and a social history, specifying how many friends she has at each moment, what her social status is, and so forth. A fully specified action set ***A***, for June, would consist of every possible regulation she might issue to limit the toxin. The outcome (in ***O***) produced by a given regulation would depend on the causal laws of toxicology (which determine the direct impacts of the toxin on health and longevity); the laws of economics (which determine how firms, consumers, and workers will react to a given regulation, thereby affecting consumption and other items); and the laws of psychology and sociology (which determine how an individual's health, longevity, and consumption will impact his affective states, and how the distribution of consumption and other items affects social relations). Thus, each state will consist of a combination of a particular toxicological model, economic model, psychological model, and sociological model, plus a specification of prior facts relevant to all the models, sufficiently detailed to lead determinately to one member of ***O*** for each member of the action set.

This level of specification massively exceeds the detail about possible regulations, outcomes, and causal models currently employed by regulators in even the largest and most deliberate rulemakings.[11] I suggest that it would be normatively unwarranted for June to attempt to follow a full specification norm in implementing her welfarist goals. The effort might never end and, if it did, would surely consume large amounts of time and resources. So what *is* the normatively appropriate procedure for June to follow? That, as I see it, is the question of bounded rationality, to which DT currently offers no good answer.

To be clear, I do not mean to suggest that the problem of providing a procedural norm that bounded decisionmakers should follow in implementing an instrumentalist normative framework is unanswerable. "Cognitivists" believe that there are normative truths and facts. "Noncognitivists" deny this, and think that normativity is at bottom just a matter of what norms we wish to endorse. Although my sympathies are cognitivist, this chapter can remain neutral on the issue, because the problem of bounded rationality arises for both cognitivists and noncognitivists. The cognitivist must decide which normative framework she believes to be correct and, within that framework, which procedural component she believes to be correct, given humans' cognitive limitations. The noncognitivist must decide which normative framework she wishes to endorse and, within that framework, which procedural component she wishes to endorse.

How do the cognitivist and noncognitivist settle these issues? Both can, I think, avail themselves of John Rawls' method of "reflective equilibrium." We should decide what normative framework we believe is correct (the cognitivist spin), or what normative

[11] For a discussion of the ways in which cost-benefit analysis and other regulatory decision procedures need to be -- and, in practice, are -- structured to be sensitive to decision costs and delay costs, see Adler & Posner, *New Foundations of Cost-Benefit Analysis*, pp. 62-100.

framework we wish to endorse (the non-cognitivist spin) by striving for an internally coherent framework that respects (as much as possible) both our pre-equilibrium beliefs regarding general normative principles or our pre-equilibrium endorsements of such principles, and our pre-equilibrium intuitions about particular cases. "Solving" the problem of bounded rationality, for a given instrumentalist framework, means modifying DT as a procedural norm, or perhaps replacing it with an entirely different kind of norm, so that the framework now has a procedural component which is acceptable in reflective equilibrium.

Such a solution may well exist; I do not claim otherwise. Rather, my claim is that the scholarly literature on decisionmaking has not yet arrived at a solution, or even come close. Let us now survey current candidate solutions and see where they fall short.

II. "SOLUTIONS" TO THE PROBLEM OF BOUNDED RATIONALITY?

A. *Uncertainty versus Bounded Rationality*

Uncertainty and bounded rationality are distinct problems. Although DT has solved the problem of providing attractive norms for choice where the only obstacle to good choice is uncertainty, these features of the view give us zero traction in addressing bounded rationality.

To see the distinction between uncertainty and bounded rationality, observe that the decisionmaker with limitless ability to conceptualize actions, states, and outcomes might be uncertain which state obtains. Indeed, the central contribution of DT is to help us see what the procedural component of an instrumentalist normative framework appropriate for *partly* ideal agents -- ideal in the sense of having limitless mental abilities, nonideal in the sense of being nonomniscient -- would consist of. In the static case, this decisionmaker has a fully specified outcome set, a complete and transitive ranking of outcomes, and a fully specified action set. But she doesn't know which outcome would, in fact, result from a given action. So she does just what DT suggests. She asks herself, how might the history of the world, including causal laws, have proceeded up to this point? Which outcome would each action map onto, given each such state? And how strongly do I believe in the various states?

In the dynamic case, this unbounded but nonomniscient decisionmaker is considering securing more information about the states -- at some cost. For this case, one well-established aspect of DT -- "value of information analysis" (VOI) --comes into play.[12] The basic idea is that the decisionmaker can use his outcome set and an appropriately enriched state set[13] to evaluate information-seeking actions as much as other sorts of actions. To see the basic idea in a very simple case, imagine that the

[12]On VOI, see, e.g., Robert L. Winkler, 2003, *An Introduction to Bayesian Inference and Decision*, 2nd ed., Gainesville: Probabilistic Publishing, pp. 267-350; Jack Hirshleifer & John G. Riley, 1992,*The Analytics of Uncertainty and Information*, Cambridge: Cambridge University Press, pp. 167-208.

[13] That is, a state set sufficiently specified so that each state and each experiment entails one and only one outcome in the outcome set.

outcome set has four outcomes, one with utility *L*(ow), one with utility *H*(igh), one with utility *L-K*, one with utility *H-K*. There are two ordinary actions, a_1 and a_2, that the decisionmaker, Phil, can take: in the first state, s_1, a_1 leads to the outcome with utility *H* and a_2 leads to the outcome with utility *L*. In the second state, s_2, it is a_2 that produces the *H*-utility outcome and a_1 that produces the *L*-utility outcome. Phil gives the first state a probability of .6 and the second a probability of .4. There is an experiment Phil can undertake to get better information about which state obtains, at some fixed cost *K* in utility. The experiment is action a_3. Phil believes that, if s_1 obtains, the experiment sends signal "X" with probability .9 and "Y" with probability .1. If s_2 obtains, the experiment sends signal "Y" with probability .9 and "X" with probability .1. In effect, there are four states: s_1 with a propensity to send signal X, s_1 with a propensity to send signal Y, s_2 with a propensity to send signal X, and s_2 with a propensity to send signal Y. The probabilities of these states are, respectively, .54, .06, .04, and .36. This enriched state set (together with the assumption that Phil would respond to the experiment by updating his beliefs about the states in a Bayesian fashion, and then choose a_1 or a_2 using these updated beliefs) suffices to produce a determinate outcome in the outcome set for each of the three actions, that is, for the two ordinary actions and the experiment. If *K* is sufficiently small and the difference between *H* and *L* sufficiently large, the experiment is worth undertaking --otherwise not.

Value of Information (VOI) Analysis

	States				
	s_1 and signal X if experiment	s_1 and signal Y if experiment	s_2 and signal X if experiment	s_2 and signal Y if experiment	**Expected Utility**
	P= .54	P =.06	P=.04	P=.36	
Actions					
a_1	utility *H*	utility *H*	utility *L*	utility *L*	.6*H* +.4*L*
a_2	utility *L*	utility *L*	utility *H*	utility *H*	.6*L*+.4*H*
a_3	utility *H-K*	utility *L-K*	utility *L-K*	utility *H-K*	[.9*H* + .1*L*] -*K*

Phil attaches utility to a_3 as follows. He reasons that, if he were to undertake the experiment and the true state of the world is s_1 with a propensity to send signal X, he would receive signal X. He would at that point update his probabilities in Bayesian fashion and ascribe probability .54/.58 to s_1 with a propensity to send X, 0 to s_1 with a propensity to send Y, .04/.58 to s_2 with a propensity to send X, and 0 to s_2 with a propensity to send Y. With these new probabilities, he would choose action a_1, which leads to utility *H-K* in state s_1 with a propensity to send X. Similar reasoning allows him to attach utilities to a_3 in the other three states.

VOI is a very potent tool, but it simply does not address the problem of specification. It takes as given an outcome set; take as given an action set, one that includes some "experiments" and other information-gathering measures; and gives guidance in thinking through how a particular information-gathering measure might be understood (like an ordinary action) as a distribution across different elements of the outcome set, producing different outcomes in different states.

Nor does it seem fruitful to try to "adapt" VOI for the task of determining optimal specification. Consider the problem of optimally specifying an outcome set. Imagine starting with some underspecified set of outcomes ***O***. Among the actions that the decisionmaker could take at *t* are not only ordinary actions a_1, a_2, and so forth, but also the deliberative action *d* of thinking about these ordinary actions using a more fully specified outcome set ***O'***. So we might imagine a "value of specification" analysis that marries ***O*** with this expanded action set -- in other words, it predicts the effects of ordinary actions $\{a_1, a_2, \ldots\}$ *and* deliberative action *d* in the initial outcome set ***O.***

We would need to enrich the state set so that each state, combined with *d*, produces one and only one outcome in ***O.*** One way to do that is to combine the initial state set $\{s_1, s_2, \ldots\}$, sufficient for determining the outcomes in ***O*** of the ordinary actions, with statements about what would occur if the decisionmaker were to think about the ordinary actions with a more refined outcome set ***O'***. So the state set becomes something like: {s_1 and if I were to use outcome set ***O'*** in choosing among the ordinary actions I would pick a_1; s_1 and if I were to use outcome set ***O'*** in choosing among the ordinary actions I would pick a_2; ….. ; s_2 and if I were to use outcome set ***O'*** in choosing among the ordinary actions I would pick a_1; s_2 and if I were to use outcome set ***O'*** in choosing among the ordinary actions I would pick a_2; …}. And the decision matrix now looks like this.

<u>"Value of Specification" Analysis?</u>

Original Choice Situation, with Ordinary Actions

	States	
	s_1	s_2
Actions		
a_1	*o*	*o**
a_2	*o***	*o****
…		

*New Choice Situation, with Ordinary Actions plus Deliberative Action d of Refining Outcome Set from **O** to **O'**.*

States

	s_1 and if I were to use $\boldsymbol{O'}$ to choose among $\{a_1, a_2, \ldots\}$ I would pick a_1	s_1 and if I were to use $\boldsymbol{O'}$ to choose among $\{a_1, a_2, \ldots\}$ I would pick a_2.	…	s_2 and if I were to use $\boldsymbol{O'}$ to choose among $\{a_1, a_2, \ldots\}$ I would pick a_1	s_2 and if I were to use $\boldsymbol{O'}$ to choose among $\{a_1, a_2, \ldots\}$ I would pick a_2.	…
Actions						
a_1	o	o		$o*$	$o*$	
a_2	$o**$	$o**$		$o***$	$o***$	
….						
d	$o+$	$o++$		$o+++$	$o++++$	

Every outcome in this matrix is an outcome in the original outcome set $\boldsymbol{O}$. Note that $o+$ and o need not be the same outcome. That is, the outcome in $\boldsymbol{O}$ of simply picking a_1 (outcome o) need not be the same as the outcome in $\boldsymbol{O}$ of picking a_1 after deliberating among the ordinary actions using a refined outcome set $\boldsymbol{O'}$ (outcome $o+$). Outcome $o+$ may well reflect the increased deliberational expenses that flow from choosing d rather than one of the ordinary actions. Similarly, $o++$ need not be the same outcome in $\boldsymbol{O}$ as $o**$, $o+++$ need not be the same outcome as o*, and so on.

Note that this device allows the decisionmaker to consider undertaking action d and using a more refined outcome set, without actually incurring the costs of doing d. (If she decides just to undertake one of the ordinary actions $\{a_1, a_2, \ldots\}$ rather than doing d, she never actually chooses among the ordinary actions using the more refined outcome set.) So this seems like a promising start to the problem of analyzing the value of refining an outcome set.

But there is a problem. From an initial outcome set $\boldsymbol{O}$, there may be many different refinements. Consider our decisionmaker June. She cares about individuals' well-being, understood to depend on their longevity and moment-to-moment consumption sequences, health sequences, happiness, and social relations for all moments individuals are alive. She starts with an outcome set where each outcome specifies individuals' longevity and average annual consumption. She could undertake action d_1, which means using a refined outcome set that specifies individual' moment-to-moment (rather than annual) consumption; or do d_2 and use a different refined set, that specifies individuals' annual consumption and average annual health; or do d_3 and use yet a different refined set that specifies individuals' moment-to-moment consumption and moment-to-moment happiness; and so forth. There are a very large number of such d_i. Considering all of them will be costly for the bounded decisionmaker; but there is nothing in the VOI framework, either directly or by analogy, which suggests how to limit the set of deliberative actions d_1, d_2 … that the decisionmaker considers.

A similar point is that the decisionmaker can start with different outcome sets.

June might start with ordinary actions $\{a_1, a_2, \ldots\}$ and a initial outcome set $\boldsymbol{O}$ that specifies individuals' consumption, and from that starting point think about the outcomes in $\boldsymbol{O}$ of deliberative actions which pick among the ordinary actions using some refinement of $\boldsymbol{O}$. Or, June might start with ordinary actions $\{a_1, a_2, \ldots\}$ and an initial outcome set $\boldsymbol{O^*}$ that specifies individuals' happiness, and from *that* starting point think about the outcomes in $\boldsymbol{O^*}$ of deliberative actions which pick among the ordinary actions using some refinement of $\boldsymbol{O^*}$. Again, there is nothing in VOI, directly or by analogy, which helps guide June in determining which outcome set she should start with.

B. *Optimal search*

Economists since George Stigler have entertained the notion of optimal search. Consider the individual who wants to buy a good, has found some willing sellers with various asking prices, and is trying to decide whether to look for a seller with a lower asking price, given that the costs of searching are positive. Stigler's suggestion is that, in this sort of case, the rational individual searches only if doing so increases expected utility -- given expected search costs, the probability of finding a seller with a lower asking price, and the expected benefit of doing so.[14]

The optimal-search notion seems sensible enough in the case of a decisionmaker with zero analytic costs but positive costs of "searching" in the sense of making physical efforts to locate goods. Such a decisionmaker can consider every possible sequence of bodily movements he might make, every possible choice situation that might eventuate along each sequence, the choices he would make in each such situation, and the possible results of each such choice - and then decide whether to initiate some sequence of bodily movements, as opposed to engaging in an immediate act of consumption, by maximizing expected utility. It is harder to see how the optimal-search notion is helpful for the decisionmaker with positive analytic costs.

Consider adapting the notion to the specification of choice sets (where it seems, at first blush, applicable). How exactly does this work? Building on the discussion in the previous section, perhaps we might say this. The decisionmaker's outcome set $\boldsymbol{O}$ is fixed. He has some initial choice set $\{a_1, a_2, \ldots\}$. There are additional options $\{x_1, x_2, \ldots\}$ not in the initial choice set. Rather than just considering those, he considers a "choice-set-expansion" choice e. Performing a choice-set-expansion choice such as e means expanding the initial choice set $\{a_1, a_2, \ldots\}$ in some way. For example, e might mean expanding $\{a_1, a_2, \ldots\}$ to $\{a_1, a_2, \ldots, x_1\}$. The decisionmaker evaluates the outcomes of a choice such as e in the outcome set $\boldsymbol{O}$ with an appropriately enriched state set -- one with conditionals of the form, "were I to consider this expanded choice set, I would choose this action".

<u>Analyzing the Value of Expanding a Choice Set</u>

[14] See George J. Stigler, 1961, "The Economics of Information," *Journal of Political Economy*, 69, pp. 213-225.

States

	s_1 and if I were to add x_1 to my choice set I would choose a_1	s_1 and if I were to add x_1 to my choice set I would choose a_2 …	s_1 and if I were to add x_1 to my choice set I would choose x_1 …	s_2 and if I were to add x_1 to my choice set I would choose a_1 …
Actions				
a_1	o	o	o	o'
a_2	$o*$	$o*$	$o*$	$o**$
…				
e	$o+$	$o++$	$o+++$	$o++++$

But this procedure is perverse for a decisionmaker with positive analytic costs (at least absent an intrinsic preference for the very process of decision analysis). The analytic costs of considering an expansion option e -- the option of supplementing the initial choice set $\{a_1, a_2, \ldots\}$ with some choice x_1 -- are *greater* than just adding x_1 to the choice set without thinking about it and then choosing among $\{a_1, a_2, \ldots x_1\}$. In order to evaluate $\{a_1, a_2, \ldots, e\}$in terms of our fixed outcome set ***O*,** the decisionmaker must have a set of mutually exclusive and collectively exhaustive states sufficiently richly described so that: each state maps each of the initial actions $\{a_1, a_2, \ldots\}$onto one and only element of ***O***; each state maps the choice e onto one action in the choice set $\{a_1, a_2, \ldots x_1\}$, that is, each state determines which action in the expanded set the decisionmaker would choose, were she to expand the set; *and*, in each state where expanding the choice set would lead the decisionmaker to choose x_1, x_1 is mapped onto one and only element of ***O***. In order to evaluate the choice set $\{a_1, a_2, \ldots, x_1\}$ in terms of our fixed outcome set ***O***, the decisionmaker needs a less elaborate state set: she needs a set of mutually exclusive and collectively exhaustive states such that each state maps each of the initial actions $\{a_1, a_2, \ldots\}$ onto one and only one element of ***O***, and each state maps x_1 onto one and only one element of ***O***. Simply adding x_1 to the choice set without thinking and then choosing from the expanded set is less analytically laborious than thinking about adding x_1 to the choice set -- so why not just take the first course?

A second and equally serious difficulty for the proposal now under discussion is this: If we start with some subset $\{a_1, a_2, \ldots\}$ of the totality of actions open to the decisionmaker right now, there are a huge number of choice-set-expansion options e_1, e_2, e_3, … where each e_i expands the initial subset in a different way. And there is nothing in the optimal-search literature that I am aware of which explains how the bounded decisionmaker should narrow the choice-set-expansion options being considered -- in effect, how he should narrow down the set of possible searches, and think about only some of them.

C. *Tversky and Kahneman and the "Heuristics and Biases" Program*

Amos Tversky, Daniel Kahneman, and co-workers have shown that real-world individuals are characterized by a variety of "heuristics and biases,"[15] and this now-famous research has, in recent years, generated much legal scholarship, under the banner of "behavioral law and economics." One heuristic identified by Tversky and Kahneman, "prospect theory," involves a framework for decisionmaking which directly violates DT. In brief: decisionmakers do not think in terms of the value of outcomes, but rather in terms of the value of losses and gains relative to some reference point. The value function for losses is convex, the value function for gains is concave, and the loss function is steeper than the gain function. Where the possible losses and gains resulting from a choice are uncertain, individuals do not weight the value of possible losses and gains by their probabilities, but rather by a weighting function which is an S-shaped transformation of the probabilities.

A different set of heuristics involve "judgments" rather than choice. These are processes by which individuals form beliefs -- in particular, processes by which individuals make probability judgments. The three judgment heuristics originally identified by Tversky and Kahneman were "representativeness", "availability", and "anchoring and adjustment." In brief: the representativeness heuristic involves judging whether some possible state of the world is true, given some data, by focusing on whether the data is representative of the state. The prior probability of the state is ignored. "Availability" means ascribing higher probabilities to states whose instances are easier to recall from memory. "Anchoring and adjustment" means making probability judgments (or other magnitude estimates) by starting from some initial, arbitrary value.

Sometimes, the heuristics identified by Tversky, Kahneman and their successors are termed instances of "bounded rationality." But do these heuristics in fact help solve the *normative* problem of "bounded rationality" as I have described it here? To begin, it should be noted that Tversky, Kahneman and others working in this tradition generally seem to see the heuristics and biases they identify as *departures* from rationality -- as deviations from the true norms of choice, embodied in DT, and the true norms of judgment, embodied in probability theory.[16]

To be sure, prospect-theoretic choice, and probability judgments driven by representativeness, availability, or anchoring and adjustment, *could* really be part of the correct procedural component of the correct normative framework, whatever Tversky and Kahneman think. While that is true, these heuristics are not particularly relevant on the

---

[15] For reviews of this scholarship, see, e.g., Jonathan Baron, 2000, *Thinking and Deciding*, 3d ed., Cambridge: Cambridge University Press; Reid Hastie & Robyn M. Dawes, 2001, *Rational Choice in an Uncertain World*, Thousand Oaks: Sage Publications; Gideon Keren & Karl H. Teigen, 2004, "Yet Another Look at the Heuristics and Biases Approach," in Derek J. Koehler & Nigel Harvey, eds., *Blackwell Handbook of Judgment and Decision Making*, Malden: Blackwell Publishing, pp. 89-109; Daniel Kahneman, 2003, "Maps of Bounded Rationality: Psychology for Bounded Economics," *American Economic Review*, 93, pp. 1449-75.

[16] *See* Gerd Gigerenzer et al., 1999, *Simple Heuristics that Make Us Smart*, New York: Oxford University Press, pp. 25-29; Kahneman, *Maps of Bounded Rationality*, p. 1456.

issue at hand: namely, how to revise DT, and the norm of full specification, so as to be usable by bounded decisionmakers. Prospect theory is just orthogonal to that question. It tells the decisionmaker to transform outcomes into losses and gains from a reference point, and to transform the probabilities of states by an S-shaped function. How to characterize outcomes in the first place (the problem of specifying an outcome set), and which actions to consider (the problem of specifying an action set), are not a part of the theory.

The judgment heuristics, too, give no guidance about how to specify an outcome, state, or action set. However, these heuristics could be seen as an important adjunct to DT revised for the bounded decisionmaker. They might be an appropriate way for that decisionmaker to economize on the analytic costs of assigning probabilities to states, at least in the case of "small" decisions. But, intuitively, all or at least many instrumentalist normative frameworks will identify a class of "large" decisions where it would be unwarranted for the decisionmaker not to think more systematically about her decisions. (I am pointing to intuitions because, on a reflective-equilibrium account of normative reasoning, intuitions matter.) The whole apparatus of probability theory and statistics is a methodology for thinking more systematically about our probability ascriptions. Consider welfarism. Within the framework of welfarism, aren't there at least some decisions -- in particular, governmental decisions to issue statutes or regulations that govern many individuals -- where the deployment of this apparatus is the right thing to do?

I will not try to identify this class of "large" decisions more precisely, using criteria that a bounded decisionmaker could employ. I don't see how I could do so except as part of a more general normative account of decisionmaking by bounded actors -- the very problem that no one has yet managed to solve. But it is implausible, I suggest, to think that the class is null or small, and thus implausible to take representativeness, availability, anchoring and adjustment, and similar "quick and dirty" judgment heuristics as anything like full solutions to the question of how bounded individuals should make probability ascriptions.

D. *Simon and Satisficing*

Herbert Simon is *the* pioneer in normative work on choice by bounded agents.[17] He rightly and famously argued that computational costs undermine expected-utility theory as a general account of rational choice. But -- notwithstanding all the credit Simon is owed for initiating this field of inquiry -- he did not, I suggest, succeed in solving the problem of bounded rationality.

[17] See generally Herbert Simon, 1982, 1997, *Models of Bounded Rationality, Vols. 1-3*, Cambridge: MIT Press; Herbert Simon, 1979, 1989, *Models of Thought, Vols. 1-2*, New Haven: Yale University Press.

Simon proposed that bounded agents rationally "satisfice."[18] Satisficing, generally, means setting a threshold or aspiration level for a choice. Rather than constructing a full action set, consisting of every option open to her, the decisionmaker begins to enumerate these options (systematically or not) and chooses the first one that is "good enough" -- that meets the aspiration level. Satisficing is, therefore, directly responsive to the problem that constructing a full action set may be infeasible or hugely expensive for the bounded agent.

But is satisficing a normatively attractive solution to this problem? That depends on how the aspiration level is set. Some scholars (not Simon) suggest that an aspiration level might be set in terms of expected utility.[19] For this proposal to be workable, we would need to solve the problem of simplifying the outcome set. Calculating the expected utility of an action by identifying its possible outcomes in a fully specified outcome set, developing a matching state set, and then assigning utilities to the outcomes and probabilities to the states, would be infeasible or at least unwarranted if the fully specified outcome set is very large. We would also, of course, need to provide a normatively attractive proposal for setting the aspiration level while economizing on computation costs (relative to doing DT with a fully specified outcome set). I am not aware of plausible suggestions on these two fronts.

A different possibility is to set an aspiration level for choices in terms of some non-utility features of choices -- for example, deciding which house to purchase by looking for houses until one meets a price cutoff and a size cutoff. But how do we do this where the impact of choices on the outcomes we care about is uncertain?

Simon, in his original work on satisficing, offered a very simple suggestion. Take every *outcome* and make a binary judgment that the outcome is either "satisfactory" (a value of 1) or "unsatisfactory" (a value of 0). A choice, then, meets the aspiration level if it has a satisfactory outcome in every state.

This solution (upon examination) leaves untouched the problem of specifying an outcome set. Are we supposed to consider whether the possible outcomes of an action in the fully-specified outcome set are always 1 or 0? In many cases, this will be infeasible or hugely expensive. Reciprocally, once we fix an incompletely specified outcome set appropriate for the problem at hand, it will sometimes -- for some "large" decisions -- be appropriate to order these outcomes on some finer scale than a binary scale.

A different solution is to pick certain aspects or dimensions of outcomes that are relevant to the goals adopted by our normative framework; produce a simplified set of outcomes that differ with respect to *these* dimensions; and set an aspiration level for choice in terms of the expected value of choices with respect to the selected dimensions.

---

[18] See Herbert Simon, 1955, "A Behavioral Model of Rational Choice," in both *Models of Bounded Rationality* and *Models of Thought*. For discussions of satisficing by other scholars, see Michael Byron, 1998, "Satisficing and Optimality," *Ethics*, 109, pp. 67-93; Michael Byron, ed., 2004, *Satisficing and Maximizing*, Cambridge: Cambridge University Press.

[19] See Byron, "Satisficing and Optimality."

(For example, Fred may care about his income, his living space, his aesthetic experience, his happiness, and many other things. He focuses just on income and living space; and decides to bid on a house whose *expected* cost is no more than $300,000 and whose *expected* living space is at least 2000 square feet.) If the simplified outcome set is sufficiently simplified, this may reduce to a case of choice under certainty. (Fred will, in fact, know what the contract price and the living space of his house is.)

One difficulty here is identifying the dimensions of outcomes to focus upon. Intuitively, we want to pick the most "important" dimension -- but identifying that dimension, while economizing on analytic costs, may prove tricky, as I shall discuss in a moment. Quite apart from this issue, the variant of satisficing now under consideration still faces the problem of specifying a non-arbitrary aspiration level.

E. *Gigerenezer, "Take the Best" and Noncompensatory Choice Procedures*

A body of recent work on bounded rationality has focused on "noncompensatory" heuristics. Much of this scholarship has been undertaken by Gerd Gigerenzer,[20] who is notable for his endorsement of heuristics as *rational*. He sees their use, not as a deviation from rationality, but as a rational response to computational demands.

Gigerenzer and his co-workers have focused on judgment rather than choice -- in particular, on prediction tasks -- and have analyzed a heuristic they call "Take the Best"(TTB).[21] Imagine that there is a population of objects. The individual is given pairs of objects, and tries to predict which object has a higher value on some criterion. Each object in the population has a series of *k* binary cues. The predictor (in the simplest case) knows whether each object he is presented has a negative or positive value for each of the *k* cues, and knows the ecological validity of each cue -- that is, the frequency, across all pairs of objects where one object has a positive cue value and the other not, with which the positive value object has the higher criterion value. TTB tells the individual to follow this procedure in predicting which object in a pair has a higher criterion value: Order the cues in the order of their validity; if the most valid cue assigns a positive value to one object and a negative value to the other, choose the first object; otherwise, move on to the next-most-valid cue and, if that doesn't differentiate, the next, ultimately picking randomly between the objects if no cue differentiates.

TTB is a noncompensatory rule in the sense that the information about a pair of objects provided by a higher-validity cue can never be outweighed by the information provided by lower-validity cues. Gigerenzer's research finds TTB to be surprisingly accurate in a variety of prediction tasks.

---

[20] See Gigerenzer, *Simple Heuristics that Make Us Smart*; Gerd Gigerenzer, 2000, *Adaptive Thinking: Rationality in the Real Worl*d, New York: Oxford University Press; Gerd Gigerenzer & Reinhard Selten, eds., 1999, *Bounded Rationality: The Adaptive Toolbox*, Cambridge: MIT Press; Gerd Gigerenzer, 2006, "Heuristics," in Gerd Gigerenezer & Christoph Engel, eds., *Heuristics and the Law*, Cambridge: MIT Press, pp. 17-44.

[21] See, e.g., *Simple Heuristics that Make Us Smart*, pp. 75-95.

TTB is closely related to noncompensatory choice procedures that other scholars have studied.[22] Much of this work has focused on choice under certainty. Now, there are multiple choices, not necessarily a pair; the choices have values on *k* dimensions; the decisionmaker's true preference is some weighted sum of the values on the *k* dimensions. One sort of noncompensatory procedure in this context is the "lexicographic" rule -- the analogy of TTB in the prediction context. Order the dimensions by their weights, top to bottom. Given a choice situation, pick the choice with the highest value on the top dimension. If there are multiple choices tied for the highest value on the top dimension, pick among these using the next-top dimension; and so forth. A different, even simpler noncompensatory procedure is to choose using only the top dimension -- and then pick randomly among choices tied for the highest value on this dimension.

These ideas, in turn, can be generalized to the case of choice under uncertainty, and to the case of any instrumentalist normative framework (not just the Humean framework), along something like the following lines. The goals of the framework define a fully-specified outcome set. The goals also order the outcomes and allow the assignment of a utility to each outcome, representing its place in the ordering. Assume, further, that these utilities are the weighted sum of an outcome's "subutilities" along *k* dimensions or aspects of the outcomes. The top-value rule says this: define a simplified outcome set, where outcomes differ only with respect to the top dimension. Use a DT framework, framed in terms of that simplified outcome set, to choose between the alternatives -- and if actions are equally good in terms of that set, choose randomly. The lexicographic rule says: Define a simplified outcome set, where outcomes differ only with respect to the top dimension. Use a DT framework to choose between the alternatives, using that simplified outcome set. If two or more are tied for best, move on to the next-best dimension, and frame a DT problem in terms of an outcome set specified with respect to that dimension. If two or more remain tied for best, move on to the third-best dimension to choose between these ….

An initial difficulty with such lexicographic or single-value procedures is that they do not help with the problem of specifying an action set.

But don't they at least give us some traction in simplifying the outcome set? One difficulty is that the utilities which the normative framework assigns to outcomes may not be representable as the linear sum of *k* dimensions. The literature on multiattribute utility theory shows that a linear decomposition of preferences (and, more generally, goals) presupposes special conditions which are not, plausibly, general requirements of rationality.[23] Absent a linear decomposition, it may not be meaningful to say that one dimension or aspect of outcomes is weightier than another in terms of the framework. A different and obvious problem is that, even if the utility numbers that the framework

---

[22] See John W. Payne et al., 1993, *The Adaptive Decisionmaker*, Cambridge: Cambridge University Press; Robin M. Hogarth & Natalia Karelaia, 2007, "Heuristic and Linear Models of Judgment: Matching Rules and Environments," *Psychological Review*, 114, pp. 733-58; Barbara Fasolo et al., 2007, "Escaping the Tyranny of Choice: When Fewer Attributes Make Choice Easier," *Marketing Theory*, 7, pp. 13-26.
[23] See Detlof von Winterfeldt & Ward Edwards, 1986, *Decision Analysis and Behavioral Research*, Cambridge: Cambridge University Press, pp 331-34.

assigns to fully specified outcomes are linearly decomposable into the sum of weighted dimensions, the bounded individual may not know what those weights are.

A third and distinct problem is this: Noncompensatory procedures are not perfectly accurate and, in some environments, may be sufficiently less accurate than compensatory procedures to be inadvisable. Robin Hogarth and Natalia Karelaia have undertaken substantial research on the accuracy of noncompensatory prediction and choice procedures.[24] Although the answer is complicated, one important determinant of whether noncompensatory procedures are much less accurate than compensatory procedures is (not surprisingly) the true weights of the cues or dimensions. If the true value or utility of an item is the weighted sum of *k* cues or dimensions, and if one cue or dimension has a much larger weight, then a noncompensatory procedure using that cue or dimension will tend to be quite accurate. If the weights are closer to equal, then a noncompensatory procedure will tend to be less accurate.

Consider, then, a decisionmaker facing a large choice: one in which, it seems, it is rational to expend substantial decisional costs to get the decision right. And imagine that our decisionmaker believes that there is no one dimension of fully specified outcomes that dominates the others, in terms of the goals she is trying to attain. Imagine, for example, our official June, who cares about individuals' well-being, and believes that well-being is a matter of longevity, consumption, health, happiness, and social relations, with none taking lexical priority over the others. June is overseeing the issuance of a major rule, which (she thinks) will change the well-being of many individuals over many years. June finds that one version of the rule maximizes population attainments with respect to the dimension of well-being she takes to be most important -- say, health. But the difference between the best rule and the next best in terms of population health is small, and June is wondering whether to consider how the rules compare on the other dimensions. Wouldn't she be rational to consider that -- at least if the other dimensions are measurable without huge expense and there is no emergency requiring immediate issuance of the rule?

In short, while noncompensatory strategies may well be normatively advisable in an important range of cases -- for small decisions, or where the decisionmaker believes there is a dominant dimension in the framework's ordering -- they are more problematic in others (large decisions without dominant dimensions).

F. *Departing from DT: Dropping the Axioms or Dropping the Structure*

DT consists of a certain structure for thinking about decisions, plus a set of norms. The attractiveness of these norms has been challenged. It has been suggested that the rational decisionmaker might only have a partial rather than complete ordering of outcomes.[25] It has also been argued (particularly in the context of social choice) that it

[24] See Hogarth & Karelaia, "Heuristic and Linear Models," and their prior scholarship cited therein.
[25] See, e.g., Adler & Posner, *New Foundations of Cost-Benefit Analysis*, pp. 161-62.

might be appropriate to violate the independence norm.[26] Finally, some have even suggested that violations of transitivity are rational.[27]

But these discussions are largely orthogonal to the problem of bounded rationality. As long as the decisionmaker is supposed to think about her choice in terms of outcomes -- whether or not her ranking of those is supposed to be complete or even transitive, and whether or not she is supposed to comply with independence -- the problem remains that we have no good theory for how to simplify the outcome set to make it tractable for bounded decisionmakers.

A more radical revision of DT would be to change its structure. Might we develop a procedural component for our normative frameworks that enjoins or permits the decisionmaker to ignore outcomes entirely (and therefore, also, to ignore states), and to focus just on her actions? There *is* a literature on non-instrumentalist approaches to decisionmaking, which I will not attempt to survey here.[28] Such approaches *might* avoid or mitigate the problem of bounded human cognitive abilities. But they would sit very uneasily with instrumentalist normative frameworks, such as welfarism, other consequentialist moral frameworks, or any other normative framework that is oriented around some goal. Indeed, they would sit very uneasily with pluralist normative frameworks, which incorporate both goals and other requirements (e.g., a hybrid moral view that tells the actor to maximize social welfare within deontological "side constraints"). Anyone attracted to any instrumentalist or pluralist normative framework should hesitate to adopt a procedural norm that tells the actor to ignore outcomes.

III. LEGAL SCHOLARSHIP

At this point, the reader is surely growing impatient. What on earth does the problem of bounded rationality have to do with the methodology of legal scholarship?

The connection is this. *The problem of bounded rationality will stymie any attempt to provide an instrumentalist methodology for prescriptive legal scholarship*. Legal scholarship may be prescriptive, providing recommendations to legislators, judges, and other legal officials; or it may be nonprescriptive, seeking to describe or explain the behaviors of various actors. A methodology for legal scholarship furnishes norms for legal scholars themselves. In the case of prescriptive legal scholarship, it tells legal scholars how to go about making recommendations to legislators, judges, and other legal officials. A methodology for prescriptive legal scholarship is instrumentalist if it tells legal scholars to make those recommendations by determining which laws and policies maximize some goal.

---

[26] *See* Adler & Sanchirico, "Inequality and Uncertainty," pp. 334-50.
[27] *See* Larry S. Temkin, 1996, "A Continuum Argument for Intransitivity," *Philosophy & Public Affairs*, 25, pp. 175-210.
[28] See Elijah Millgram, ed., 2001, *Varieties of Practical Reasoning*, Cambridge: MIT Press.

If legal scholars were cognitively unbounded, an instrumentalist methodology for prescriptive legal scholarship would naturally incorporate DT as its guidance for legal scholars -- with the further guidance, on the crucial issue of specification, that legal scholars employ fully specified outcome, action, and state sets. It would instruct the scholar making recommendations to some class of legal officials to consider all possible actions that the officials might undertake; to evaluate each action with reference to an outcome set that is fully specified in light of the goal of the methodology; to do so using a fully specified state set (containing mutually exclusively and collectively exhaustive states, each of which maps a given action onto a single outcome); and to do so consistent with the norms of DT, i.e., completeness, transitivity, consequentialism, dominance and independence.

However, because legal scholars are not cognitively unbounded, it is very difficulty to say what norms an instrumentalist methodology for prescriptive legal scholarship should contain on this crucial issue of specification. This is one implication (among many) of the general analysis presented in Parts I and II. The general problem is that reflection about appropriate decisionmaking -- by economists, philosophers, decision theorists, and others who study decisionmaking norms -- has currently reached no point of reflective equilibrium with respect to how bounded decisionmakers should use the apparatus of DT.[29] The implications of this problem for *legal scholarship* are that anyone reflecting on the appropriate methodology for legal scholarship will have difficulty reaching reflective equilibrium with respect to how legal scholars giving advice to decisionmakers should use DT to formulate that advice.

It would be absurd to suggest that the implications of bounded rationality for legal scholarship are its most important or interesting implications. I certainly do not claim that! Cognitive limitations are a pervasive feature of humans, and thus bounded rationality is a pervasive problem for attempts to specify normative frameworks for aspects of human life -- including but hardly limited to legal scholarship. Still, legal scholars should understand that bounded rationality *is* a pervasive problem, which will stymie their own efforts to rationalize their own activities. I do not believe this point has been sufficiently understood. And it is one worth articulating in a book on the methodology of law and economics.

I will elaborate the implications of bounded rationality for legal scholarship by focusing on *Fairness versus Welfare* (*FW).* *FW* is the most sustained attempt, in recent years, to provide a methodology for prescriptive legal scholarship. Although *FW* is, to some extent, addressed to governmental officials, it consists, first and foremost, of guidance for legal scholars and other policy analysts.

> Our object has been to convince legal policy analysts to pursue a research agenda focused on identifying which legal rules best promote individuals' well-being … It should be clear … that our claim concerns legal academics and other policy analysts, not ordinary individuals who must make decisions in everyday life. … We [also] acknowledge that the problem of government officials is complicated by the fact that their constituents may not always be able to understand proper analyses of legal rules (or of many other government policies). Nevertheless, we believe

[29] On reflective equilibrium, see above, text accompanying note 10.

> that responsible government decisionmakers will be able to make better policy decisions if those who analyze legal policy devote themselves to identifying the effects of legal rules on individuals' well-being -- that is, if they employ welfare economics rather than base their analysis on notions of fairness.[30]

The normative framework that *FW* adopts is welfarism. *FW* argues that legal scholars should identify and recommend the legal rules that advance the goal of social welfare - meaning the overall well-being of some population of interest but potentially also the distribution of well-being in the population.[31] To put this formally, legal scholars should identify the legal rules that maximize a "social welfare function": a utility function that assigns a utility number to each outcome as a function of individuals' well-being numbers in that outcome, and that does so without reference to the identity of the individuals (no dictatorship or proper names), and with the understanding that these well-being numbers are interpersonally comparable.[32] By well-being, *FW* means the satisfaction of fully informed preferences.[33] *FW* criticizes the traditional approach to law-and-economics scholarship, which is to evaluate laws and policies with reference to wealth-maximization or Kaldor-Hicks efficiency.[34] *FW* also famously recommends that legal scholars should ignore extra-welfarist or "fairness" considerations.

FW, although lengthy, is incomplete. A fully developed welfarist methodology for prescriptive legal scholarship would have a full procedural component, explaining in greater detail how scholars should pursue the task of identifying welfare-maximizing legal rules.[35] But any attempt to do that, I suggest, would run headlong into the problem of bounded rationality.

The conundrum of a legal scholar, trying to identify the legal rules that maximize social welfare, is quite analogous to that of our hypothetical welfarist governmental official June. Just as June needs to somehow narrow down the set of possible legal rules that she might *promulgate*, so the scholar must somehow narrow down the set of possible legal rules that he might *investigate*. The scholar, like June, wants to evaluate a given legal rule by considering the different outcomes that the rule might have, depending on the state of the world -- ignoring non-well-being facts about outcomes and focusing just on facts about well-being. But, again, an outcome set consisting "just" of outcomes fully specified with respect to well-being facts would be huge -- given that each individual's well-being is multidimensional, and given that social welfare is a function of everyone's well-being (not just the well-being of a single dictator). As Kaplow and Shavell explain:

> The notion of well-being used in welfare economics is comprehensive in nature. It incorporates in a positive way everything that an individual might value -- goods and services that the individual can consume, social and environmental amenities, personally held notions of

---

[30] P. 472. See generally pp. 382-402.
[31] *FW* does not take a position on the membership of this population. See p. 26 n.19.
[32] See pp. 24-28.
[33] See pp. 18-24, 410-13.
[34] See pp. 35-38, 458-61.
[35] *FW* itself, it should be noted, does not discuss the problem -- which I take to be crucial -- of how legal scholars should construct outcome, action and state sets. There is a brief discussion of how welfarist scholars should reach conclusions under uncertainty, which does not address the problem. See pp. 457-58

> fulfillment, sympathetic feelings for others, and so forth. Similarly, an individual's well-being reflects in a negative way harms to his or her person and property, costs and inconveniences, and anything else that the individual might find distasteful. Well-being is not restricted to hedonistic and materialistic enjoyment or to any other named class of pleasures and pains. [36]

Finally, like June, the scholar must somehow whittle down his state set. Policies produce outcomes (inter alia) by affecting human behavior. So a computationally unbounded scholar predicting the effect of a given policy on a given set of outcomes would consider every possible model of human behavior to which he ascribes some nonzero probability -- not just the traditional model of expected utility maximization, and prospect theory, but every alternative. More precisely, if there are N individuals in the population, and a (presumably very large) finite number M of behavioral models to which he ascribes a nonzero probability, he should consider every composite model which says that individual i behaves in accordance with model j -- leading to M^N composite models. This is an overwhelming task for bounded scholars.

As an illustration of the difficulty I am describing, consider optimal tax scholarship -- the main area in which scholars have explicitly used social welfare functions to evaluate policies and, on Kaplow and Shavell's view, the model for normative legal scholarship.[37] James Mirrlees, in his original, Nobel-prize winning work, sought to determine the optimal income tax schedule, given the following assumptions:

> Imagine an economy where individuals have the innate ability to transform working-time into a single consumption good, which is called income. Each individual's utility is a numerical function which depends only on his net income and the quantity of labour he supplies. Thus his preferences are personal. Individuals are regarded as identical except that they vary in their ability to supply labor. Thus they can be grouped by productivity types. For the same number of hours worked, a more able person naturally can produce more income. Each individual decides how much labour to supply, calculating what will maximize his utility. All these labour supply decisions taken together determine the output of the economy …. [Government] cannot monitor the number of working hours a person chooses to work, but can only observe a person's income. For this reason the only policy the government can execute is to impose a tax schedule. The government chooses the income tax schedule which maximizes its social welfare function, knowing the manner in which individuals of any productivity type will respond. [38]

We see, here, a radical simplification in the characterization of outcomes. Well-being is a function of income and leisure, nothing else. We also see a radical simplification in the state set. In particular, prospect theory and every other non-expected utility model of individual behavior are ignored. And a very simplified model of the economy is adopted (each worker cannot change his productivity, and his wage rate is solely a function of his productivity). Finally, Mirrlees in his original work ignored the possibility of optimizing a social welfare function by combining an income tax schedule with other policies.

---

[36] P. 18.

[37] See *FW* pp. 31 n.31.

[38] Matti Tuomala, 1990, *Optimal Income Tax and Redistribution*, Oxford: Oxford University Press, pp. 4-5. See also pp. 86-105. For the original paper, see James Mirrlees, 2006, *Welfare, Incentives, and Taxation*, Oxford: Oxford University Press, pp. 131-73.

These were understandable simplifications. Even a Mirrlees is human and faces computational limits! The problem is that we lack a persuasive normative account to tell us whether the simplifications Mirrlees adopted were the right ones -- whether he simplified too much, or too little, or along the right dimensions.[39]

Some readers of early drafts of this chapter have been puzzled by my focus, here, on the bounded rationality of *scholars*. Isn't that a secondary or meta-problem? *FW* instructs legal scholars to identify the laws and policies that maximize social welfare. In order to do so, scholars must predict how different laws and policies will affect the well-being of individuals in the population. Those effects will depend, inter alia, on the behavior of legal officials and private citizens. Officials and citizens may have cognitive limitations, and may deviate from the model of expected utility maximization posited by traditional economics. Isn't it this deviation -- not the cognitive limitations of legal scholars themselves -- that constitutes the primary threat to *FW*'s welfarist methodology for legal scholars?

I think not. If legal scholars were unboundedly rational, the fact that non-scholars -- private citizens or legal officials -- might deviate from expected utility maximization would not frustrate *FW*'s program for legal scholarship. In general, DT imposes norms of rationality on the decisionmaker using DT; but it is agnostic about the nature of the casual models driving the human actors or nonhuman processes in the decisionmaker's environment.[40] DT as a framework for prescriptive legal scholarship imposes norms of rationality on the scholar herself; but it is agnostic about the nature of the causal models that determine the effects of legal rules, including causal models of the behavior of individuals and government officials. Unboundedly rational legal scholars using DT's state-outcome-action framework could, without difficulty, be sensitive both to the possibility that some or all citizens and officials might conform to expected utility maximization, and to the possibility that some or all citizens or officials might fail to conform. In one (improbable!) state within the state set, all conform; in another state, none do; in other states, some do and some don't.

Of course, prescriptive legal scholars are addressing their recommendations to government officials. But I don't see why the bounded rationality of the addressees would stymie *FW*'s program if the legal scholars themselves were unboundedly rational.

In short, it is the bounded rationality of *legal scholars* -- not citizens and officials --- that poses a grave difficulty for *FW*'s welfarist program for legal scholarship.[41]

---

[39] To be sure, some of the simplifications have been dropped in subsequent optimal tax work, but subsequent work still contains many simplifications that would not characterize scholarship by unbounded scholars. For a survey of subsequent work, see Tuomala, *Optimal Income Tax*.

[40] There is a different and larger question here, about whether DT can be used for "strategic" rather than "parametric" choice -- where the effects of the decisionmaker's choices depend on other actors who are themselves choosing with reference to the decisionmaker's choices. I believe the answer is "yes," but cannot pursue the point here.

[41] It might be objected that *FW* is arbitrary to address its recommendations just to legal scholars. Welfarism is either the right moral framework for everyone or for no one. Still, the point remains that the

What are the possible responses to this difficulty? One response, of course, is to reject *FW*'s welfarist program for legal scholarship. But this is trickier than it sounds. Much of the criticism of *FW* involves its rejection of fairness criteria. If this indeed is a deficiency in *FW*, and we correct it by adopting a pluralist program -- recommending that legal scholars evaluate policies with respect to both social welfare and some set of fairness criteria -- the problem of bounded rationality remains (at least) with respect to the welfarist part of this program. A different response is to adopt a wholly non-welfarist framework, which enjoins legal scholars to focus solely on fairness criteria, or other nonwelfarist considerations, and ignore overall well-being or the distribution of well-being. Such a framework might be easier to square with legal scholars' cognitive limits, but it would be normatively unattractive: social welfare is surely one significant part of the moral landscape. Finally, we might return to the traditional criteria adopted by law-and-economics: wealth maximization or Kaldor-Hicks efficiency. But these criteria are much less attractive (on their own, or as part of a pluralistic framework) than social welfare as *FW* understands it. And, in any event, cost-benefit analysis -- the traditional technique for implementing wealth-maximization or Kaldor-Hicks efficiency -- faces exactly the same problems of explaining how to simplify action, state, and outcome sets that the application of a social welfare function does.[42]

Nor will the problem be solved by shifting away from *FW*'s preferentialist theory of well-being. What compounds the problem of bounded rationality is not preferentialism but multidimensionality. Any theory which recognizes multiple sources of well-being (be it multiple objective goods, multiple kinds of good mental states, or the multiple things people prefer), and cares about the well-being of many people, will tend to generate a particularly large fully specified outcome set.

A different response is to endorse welfarist goals for legal scholars (either solo, or as part of a larger pluralist program), but to adopt one of the possible solutions to the problem of bounded rationality discussed in Part II. I have criticized those solutions already, and will not repeat the criticism here.

Yet another response is to point to the collective nature of legal scholarship. One might say, "Each legal scholar should try to identify laws and policies that maximize social welfare. But, in doing so, each scholar should coordinate with other scholars, rather than acting as if she were the sole researcher pursuing the welfare-maximization goal." This leads us into the murky waters of rule-consequentialist revisions to DT. Rather than asking which policy she should investigate, so as to maximize one or another social welfare function, the scholar might ask something like the following: what rule for the community of scholars, if followed by everyone in the community, would maximize that social welfare function? But, whatever the attractions of rule consequentialism, it does not seem a fruitful solution to the problem of bounded rationality. There is a large

---

viability of *FW*, *qua* methodology for legal scholarship, is chiefly threatened by the cognitive limitations of legal scholars.

[42] *See* Adler & Posner, *New Foundations of Cost-Benefit Analysis*, pp. 68-73.

set of possible rules to structure the community's research that might be followed, each to be evaluated in light of a large set of outcomes and states.

Finally, it might be pointed out some legal scholars are engaged in nonprescriptive scholarship. They aim, not to recommend laws and policies to government officials, but simply to describe and explain how individuals behave (specifically, to understand how individuals respond to legal rules).

There is a clear sense in which descriptive/explanatory legal scholarship does *not* need an account of bounded rationality. Imagine that, as a descriptive/explanatory scholar, I believe that human behavior in some domain is produced by some computationally cheap heuristic such as those identified by Tversky and Kahneman or Gigerenzer: availability, representativeness, anchoring and adjustment, Take the Best, and so forth. I need not take a position about whether this behavior is rational.[43] My aim is not to endorse or criticize the behavior, but to explain it. The heuristic, I claim, does so. Why do I need to take a position on whether compliance with the heuristic is normatively appropriate?

A harder question is whether legal scholars engaged in descriptive/explanatory scholarship confront the problem of bounded rationality in structuring their own research activities. Will any methodology for nonprescriptive legal scholarship be stymied by the cognitive limitations of legal scholars? Any such methodology will give guidance to legal scholars in deciding which questions about human behavior to investigate. Odd as it may sound, one such methodology -- at least in principle -- is welfarist. That methodology tells scholars to pursue those research projects into human behavior that have the greatest expected positive impact on social welfare -- in virtue of the information the projects might produce, the probability of producing that information, and the relevance of the information to well-being .[44] Such a methodology would, pretty clearly, be hampered by scholars' bounded rationality. However, there may be plausible nonwelfarist accounts of how scholars should undertake descriptive/explanatory work that are feasibly implemented by bounded scholars.

I will not try to address these questions about the foundations of descriptive/explanatory legal scholarship. Bounded rationality may not be a gap in the foundations of descriptive/explanatory scholarship, but it *is* a large and unresolved gap in our understanding of how prescriptive legal scholars should conduct their activities. *FW* may be right that prescriptive legal scholarship should focus solely on identifying legal rules that maximize social welfare. At a minimum, it is surely true that social welfare is one of the criteria that prescriptive legal scholarship should use to evaluate legal rules. But we currently lack any normative handle on how cognitively limited scholars should undertake that analysis: on how incomplete action, state, and outcome sets should be structured so that the analysis is both sufficiently intensive, and yet neither

[43] See Allan Gibbard, 2002, "Normative Explanations: Invoking Rationality to Explain Happenings," in Jose Luis Bermudez & Alan Millar, eds., *Reason and Nature: Essays in the Theory of Rationality*, pp. 265-82.

[44] Cf. Philip Kitcher, 2001, *Science, Truth and Democracy*, New York: Oxford University Press, pp. 117-35.

computationally infeasible nor overly expensive, given the welfare stakes of the rules being evaluated.