A strikingly small fraction of the scholarly literature on judging and law asks how well judges perform their work and how their performance might be improved. The aim of this essay is to explore the prospects for an ambitious normative, empirical, and prescriptive project that would set out criteria for good judging, evaluate actual judicial performance against these criteria, and suggest reforms to produce higher or steadier levels of performance. As of now, although there are scattered writings (discussed below) that would seem to belong to such a project, there is hardly any evidence that a project like this has captured the attention of the field as a whole. It's true that we write a lot about how judges decide cases, which decisions we think are wise or foolish, and even what methods we think judges should use to decide cases. But these are not the same thing.

Why hasn't the problem of judicial performance attracted more sustained attention from researchers? It does not seem plausible that people who spend their lives studying judges are uninterested in how well judges do their work. Rather, I suspect that the difficulties involved in such project---from laying out decisional norms capable of wide acceptance to familiarizing ourselves with huge chunks of literatures in which we lack expertise--appear insurmountable. I am not prepared to say with certainty that the obstacles can be overcome, and I readily concede that the job is too big for one study, scholar, or small team of scholars, if only because the potentially relevant literatures from psychology, law, and political science are enormous. And yet I am inclined to believe that a community of scholars working steadily on the project and content to divide the
labor and move incrementally could make real progress over time and am eager to hear the opinions of the other workshop participants.

The motivating questions of the essay, then, are largely practical. Can the study of psychology and judging reveal ways to enhance judicial performance? More specifically, is it possible to identify traits that make some people better equipped to be judges than others? If so, is there a measurement strategy that would allow us to ascertain the extent to which candidates for judgeships possess these traits? Even if significant individual differences cannot reliably be measured, might it be possible to teach judges cognitive strategies for good judging? Are there certain institutional rules and structures that would be expected to promote or impede good judging?

The essay begins with individual differences, first considering different possible methods for identifying important individual traits, then discussing and raising questions about some of the most interesting traits, and finally exploring obstacles to testing propositions about individual differences in judicial ability. In the second part of the paper I offer a few remarks about institutional arrangements that might have significant effects on the quality of judging. The essay offers almost nothing in the way of firm conclusions, instead posing a large number of questions in the hope that at least some will pique the interest of other participants and stimulate discussion at the workshop.

I. Picking Better Judges

Individual Differences in Judicial Ability
Even someone without the slightest exposure to psychological research could tell you that some people are better at solving problems, reaching sound judgments, or making wise decisions than others. Now judges are a fairly homogenous subset of the population, on average smarter and better trained to make decisions than a typical layperson, so it may be that we shouldn't expect to find nearly the variation among judges that there is in the population as a whole. But there is also no reason to assume we wouldn't find any meaningful variation. Surely most of us have noticed that we respect or admire some judges and judicial opinions more than others. This gives us enough reason to explore differences in abilities across judges. But how should we go about this exploration? That is the question I begin with.

**Research strategies.** It may be useful to begin by thinking about two very different approaches researchers might take to identifying important individual differences across judges. One strategy, employed at least in part by Viscusi (1999), Guthrie, et al. (2001), Hirsch (2003), and Englich (2006), begins with the premise that judges are human beings and draws primarily from the vast psychological literature on cognitive pitfalls. The first tasks of researchers taking this approach are to identify kinds of suboptimal reasoning processes\(^1\) that have the greatest potential to undermine judicial performance\(^2\) and

\(^{1}\) I readily accept the argument, advanced most persistently by Gerd Gigerenzer (see Gigerenzer 2006), that certain reasoning processes that appear to be regarded as cognitive errors by many psychologists in fact serve human beings well in many situations. When I refer to reasoning processes as suboptimal, I do not mean to imply that they should always be avoided. All I mean is that they are likely to produce inferior performance on the kind of tasks judges face. Suboptimal processes are ones we don't want our judges relying on to decide cases.

\(^{2}\) Not all cognitive problems that have been uncovered are equally good candidates. For example, that people often fail to make good predictions as to how happy or sad possible outcomes will make them (Gilbert and Wilson 2000) is fascinating and could matter
determine whether judges share the general tendency to fall into these ways of thinking. For an illustration, take Englich's (2006) report on judges' sentencing decisions. Englich chose to investigate anchoring effects, where people who are asked to give a numerical answer to some question or problem are influenced by a number mentioned to them, reasoning that anchoring effects could have an important impact on sentencing decision. In experiments asking real judges to decide hypothetical issues, she found that judges presented with prosecutorial recommendations for longer sentences tended to impose longer sentences than judges presented with more lenient sentencing recommendations. Strikingly, this effect held even when judges rolled dice to generate the "prosecutor's recommendation" and so had to know that the information contained in the recommendation was purely random and entitled to no weight in their thinking. In their similar but broader study, Guthrie, Rachlinski, and Wistrich (2001) too found that judges were susceptible to anchoring effects (in assessing damages in a civil case) as well as to several other suboptimal ways of thinking, including hindsight bias, where people considering some event that is now past underestimate how difficult it would have been to predict that event in advance.

Ideally, researchers adopting this approach would be able to take two further steps, testing a) whether judges suffer these kind of cognitive lapses often in their professional work and b) whether and why some judges are less prone to them than others. Serious obstacles stand in the way, of course, the most serious being access to judges at work. I will return shortly to the question of how we might study judges in action, but even if it were to turn out to be impossible to do, it would not seem greatly in some parts of life. But we would not expect it to affect judicial decision
irresponsible to draw inferences about real-world judging from studies of the sort just described. We can reasonably suppose that judges who make slips in their responses to brief experiments will not be able to avoid such slips entirely in judging real cases. And if we find individual traits among the general population that make such slips more or less likely, we can conclude with some confidence that these traits can distinguish in the same way between judges.

The chief drawback to this strategy is the danger that it will produce candidate lists of reasoning problems that are overinclusive or underinclusive--or both at the same time. Psychologists have so far identified scores, if not hundreds, of ways in which human thinking in certain situations manages to fall short of the ideal. For a project on improving judging to have any practical applications, it must be able to isolate a few important traits to favor in judicial selection or a few key reforms that could enhance judging. It might turn out that a host of reasoning problems have common roots and so can be ameliorated all at once. But not knowing if this is so, we would be better off trying to ensure that the processes we study are those likely to be most central to professional judging.

Underinclusion is the more worrisome prospect. At least from the perspective of this outsider, the psychological literature on sub-optimal reasoning seems to have developed rather haphazardly, with attention determined more by which issues happened to come up earliest and generate the most controversy than by which are most central to the most important reasoning tasks people have to perform.\textsuperscript{3} Even if this assessment is making often.

\textsuperscript{3} If I had a dollar for every time I read about Linda the bank teller in researching this topic, I would be able to open my own feminist bank. But I am not sure I have learned something crucial about human thinking.
unfair, we cannot ignore the possibility that there are important elements to the judging process that have not yet been studied systematically by psychologists. If we consider only those cognitive pitfalls that have already been identified by psychologists, we may miss others that matter at least as much for judicial performance.

This brings us to the other very different way of approaching the question of variation in judicial performance. This approach, as illustrated by Sherry (2003) and Solum (2003) begins by focusing not on judges' humanity but on their jobs. To follow this strategy, we begin by thinking carefully about what it means to be a judge and what is typically thought to distinguish good from poor judicial performance. We then ask ourselves what individual characteristics would be most likely to promote good performance.

The most important advantage of this strategy over the one just discussed is that it can more easily escape the confines of existing work in psychology. Judging a legal dispute is not the kind of cognitive activity the average person engages in regularly. Some of the psychological processes involved in it may be quite different from the quotidian processes typically studied by psychologists, and by starting with our focus on judging, we may notice processes that have been overlooked. But the more limited grounding in psychological theory also carries risks, especially that we will pay undue attention to characteristics that psychologists already know to be less influential or relevant than we suspect or that we will conceptualize them inadequately.

In my view, our best hope for progress lies in somehow combining the two approaches. Ideally, we would start with a fully fleshed out model of judging that not only provided a basis for evaluating overall judicial performance in a case or set of cases
but also specified the discrete tasks involved in deciding a single case. We could then turn to psychology to see how it accounts for variation in performance on tasks like those. Of the work I have seen, Dan Simon's (1998; 2004) best exemplifies this integrated approach. If a group of scholars were to critically engage his theories and ponder how to expand on them and test their empirical implications, I think significant progress could be made.

Until we have gone farther in combining insights from studies and judging and psychology, we can benefit from looking to similar work in related areas. Political science is a sensible place to start. As frequently as psychology has been incorporated into studies of mass attitudes and behavior, I think those of us who study judging would actually gain more by examining applications of psychology to foreign policy decision making (e.g., Herrmann, et al. 1997; Geva, Mayhar, and Skorick 2000; Boettcher 2005) and other elite decision making. I have found Tetlock's (2005) recent work on expert forecasting enormously helpful for thinking about judging. Of course the specific reasoning tasks and psychological issues they examine will not always be those most relevant for judging. For instance, foreign policy studies give a great deal of attention to assessing risk, and Tetlock's book is mostly about predictions. Assessing risk and predicting the future will sometimes be very important for judges (as when they're deciding on cases or considering the wider consequences of modifying doctrine), but there will also be plenty of instances where they play little or no role in judges' thinking. But these works can still provide us with psychological insights and with models of theorizing and empirical testing. Work on decision making even further removed from law and politics could be equally valuable. I have only caught glimpses of such work so
far but suspect that there is a good deal of it out there and would appreciate guidance from those more familiar with it.

**More on good and bad thinking.** So we have not yet performed the rigorous groundwork necessary to define a set of individual-level variables likely to have important effects on the quality of judicial decision making. But we still may be able to identify some strong candidates for inclusion in that set, and some preliminary thinking about those characteristics could be valuable, even if it is a bit premature. To move us in this direction, let us return to the first strategy and consider more fully some of the kinds of cognitive pitfalls we would like to see our judges stay out of.

We have already encountered the pitfall of overusing heuristics, cognitive shortcuts that can serve people very well on a daily basis but are less appropriate where, as in judging, it is very important to reach a correct decision and there is time for thought. There are many more heuristics that might matter for judging, but I want to think instead about some more pervasive influences on cognition. The set of influences I will discuss are very different in some ways, but they all have two important things in common: they generally occur outside of people's awareness, and if people were aware of them, they would typically regard them as unwelcome.

If for some reason it were only possible to study one such influence, I would almost certainly choose confirmation bias. As explained by Nickerson (1998) in his excellent overview, the term "refers usually to unwitting selectivity in the acquisition and use of evidence." (175) When subject to confirmation bias, although we may mean to "impartially evaluat[e] evidence in order to come to an unbiased conclusion," we wind up
"building a case to justify a conclusion already drawn." (175) The threat to good judging is obvious. Consider a judge who must determine whether a legal claim has a sound statutory basis by considering the language of the statute, the history of the statute's adoption, and/or prior judicial opinions interpreting the statute. If at some level the judge is convinced that the plaintiff's claim has merit, confirmation bias could lead him to give too much weight to evidence suggesting that the claim is allowed by the statute and too little to evidence in the other direction. For instance, in his thought processes he might move quickly past language in the statute itself that could be read as excluding this kind of claim, while paying careful attention to comments in the legislative record indicating that some of the bill's supporters expected it to have broad applications.

Confirmation bias would seem to be an equally dangerous threat for both appellate and trial judging. The next pitfall to consider probably has a fairly minor effect in appellate courts but could play a major role in trial judges' decisions. Psychologists have found considerable evidence that attitudes toward other people or things and certain ways of thinking about them can enter into our thinking without our knowing it, unbidden and even unwanted (Greenwald and Banaji, 1995, 8). These "implicit" or "automatic" cognitions can influence how we evaluate other people (Devine, et al. 2002) and how we behave, even in contexts that have little to do with what we're actively thinking about (Bargh, et al. 1996).

Often the effects will be harmless or even helpful, as when my response to the stereotype "bears are dangerous" spares me an unpleasant encounter with the one investigating my trash can. But when trial judges act on such influences, the result can be injustice. The most obvious danger is for sentencing decisions, where we would be especially
concerned that implicit attitudes related to race, ethnicity, or gender could cause a judge to treat people differently without a valid reason. But they could also inappropriately affect how judges respond to parties in civil cases, how they assess witnesses and even how they behave toward attorneys.

To say that these cognitive interlopers are automatic or implicit does not mean that we can never keep them from infecting our reasoning where we don't want them to. Indeed, there is evidence that they sometimes can be controlled and that some people are better at controlling them than others (Devine, et al. 2002) So just as the importance of avoiding confirmation bias encourages us ask whether and why some people are better able than others to avoid it, it makes sense to ask about individual differences in control over things like stereotypes and implicit attitudes and to suppose that those with more control would make better judges.

Better judges also might have more control over another often unnoticed influence on thinking, namely affect. I confess that I began my research on this project with a view of affect that now seems sadly naive, supposing that affect was best seen as a hindrance to good thinking. This simplistic view has been widely rejected in recent years psychologists. Many argue not only that feelings, emotions, and moods pervade normal cognitive activity (see generally Forgas 2000; Dai and Sternberg 2004) but that they are essential to healthy and effective thinking (Damasio 1994; Thagard 2006). If so, then the challenge for someone hoping to think well is not to keep affect from entering into cognitions at all but to let it infuse cognition where it is likely to be helpful and block it out where it is not.
A related theme in the literature is the importance of intuition. Unlike analytical thought, intuition "somehow produces an answer, solution, or idea without the use of a conscious, logically defensible, step-by-step process" Hammond (1996, 60). We often describe solutions that come to us through intuition as "feeling" right. Hammond argues forcefully, and to my mind, persuasively, that we can frequently find better solutions by relying on both intuition and analysis than through analysis alone. While adopting a less explicitly prescriptive position, Haidt (2001) argues that intuitions are central to the kinds of moral judgments we make in daily life. He does not claim that good moral judgments must have intuitive bases, but his insistence that intuition is integral to moral thinking at least implies that someone who tried to reach moral judgments without ever consulting her intuitions might find herself on shaky ground. I suspect that in many instances the process of deciding legal questions closely resembles the process of reaching moral judgments.

My readings in these areas have not progressed far enough--nor, I suspect, have psychologists reached enough of a consensus--for me to have a clear picture of the role affect and intuition would play in the best judging. In the concluding section of the essay, I come back to this issue with some questions for psychologists. But for now we can be content with the insight that people are likely to do a better job of thinking when they allow the right feelings to have the right influences on them.

**What individual traits?** As preliminary as it is, the preceding discussion of desirable and undesirable elements in thinking gives us something to work with. We can provisionally define the best judge as one who avoids confirmation bias and the
inappropriate use of heuristics, is able to take note of and fight against implicit cognitions that should not enter into her decisions, and incorporates affect and intuitions into her thinking insofar, but only insofar, as they aid good judgment. With these criteria in mind, we can then ask whether there are individual human traits that make some people more likely than others to satisfy the criteria.

Someone skeptical of this project might point out that there could be a nearly infinite number of human traits that affect how well people tend to think. That might well be true, but we can hope to make some progress by starting with characteristics that have been studied most extensively by psychologists. In my readings, I have encountered one kind of individual characteristic far more than all others combined, namely people's cognitive dispositions, or their customary ways of thinking. Each of the following characteristics, or tendencies, has been found to predict success at overcoming some combination of the cognitive challenges we are focusing on:

A) Need for cognition, defined by Cacioppo et al. (1996) as "a stable individual difference in people's tendency to engage in and enjoy effortful cognitive activity."(198) On average, people high in need for cognition should be more likely to overcome the challenges.

B) Dogmatism, or closed-mindedness (Rokeach 1960). According to Davies (1998),

[D]ogmatism is characterized by the relatively closed cognitive organization of beliefs and disbeliefs about reality. The closed nature of the belief systems of individuals high in dogmatism can be observed in their tendency to compartmentalize and isolate their beliefs and disbeliefs, whereas the more open belief systems of individuals low in dogmatism can be observed in their readiness to make connections between disparate...
beliefs. Individuals high in dogmatism strive to avoid inconsistency in their attitude and belief systems, and they react to inconsistent information by minimizing or ignoring it.

(456, citations omitted)

On average, less dogmatic people should meet the challenges more successfully.

C) Need for cognitive closure. As defined by Kruglanski and Webster (1996), this "refers to individuals' desire for a firm answer to a question and an aversion toward ambiguity." To put this somewhat crudely, individuals high in need for cognitive closure care more about getting an answer than getting the answer right. Unlike need for cognition, need for cognitive closure can be situational, meaning that a given person could experience a greater need for closure in one situation than another. But it also understood to vary across persons, with some individuals routinely experiencing a greater need than others. On average, those lower in need for cognitive closure should perform better.

D) Integrative complexity. People who engage in integratively complex reasoning recognize the legitimacy of opposing viewpoints and work toward solutions that incorporate insights from different viewpoints. More formally,

Integrative complexity is formally defined in terms of two cognitive stylistic attributes: evaluative differentiation and conceptual integration. Evaluative differentiation requires the emergence of dialectical (thesis–antithesis) reasoning. A speaker is evaluatively undifferentiated if he or she sees the world in rigid, good–bad terms that preclude the existence of reasonable alternative perspectives and that deny the possibility of legitimate trade-offs... The second cognitive stylistic attribute, conceptual integration, requires reasoning that builds on earlier evaluative differentiations ... Common forms of integration include developing explanations for why reasonable people view the same
Like the need for cognitive closure, integrative complexity is understood to vary across both individuals and situations. Higher levels of integrative complexity should tend to be associated with better performance.

While these four concepts are distinct, there are clear similarities between them, especially the last three, and they are sometimes analyzed together (e.g., Stanovich 1999, Parker and Fischhoff 2005). That they are closely related and produce similar findings as far as people's ability to avoid cognitive pitfalls gives us reason to think that we are on to something with them. So does the fact that they predict performance in tasks and professions similar to judging. Most important here is Tetlock's (2005) finding that "hedgehogs", who adhere rigidly to a favorite theory in trying to explain various phenomena, tend not to be as good at forecasting political and economic events as "foxes," who approach problems from a more flexible, ad hoc theoretical perspective. It is hard to resist the conclusion that people who enjoy thinking and looking at problems from multiple perspectives are likely to make better judges.

But the conceptual similarities also create difficulties. The scholars who originally developed these concepts attempted to differentiate them, but at least to this non-expert the contrasts do not seem terribly clear, nor have subsequent studies applying these concepts had as firm a theoretical grounding as one could wish. I am frequently

---

4 As noted, the need for cognition appears to be fairly distinct from the other three concepts. Yet Sargent (2004) found that an apparent relationship between need for cognition and support for tough punitive measures (those higher in need for cognition are less supportive) disappeared once he controlled for what he calls "attributional complexity." Curiously, attributional complexity, itself very complex (comprising seven subscales!), would seem at first glance to be more closely akin to the other three concepts.
unable to see why, say, a researcher interested in explaining performance on a given task chooses to focus on need for cognition rather than need for cognitive closure. Studies that systematically test these various dispositions against each other seem hard to find.

So at the same time that my readings to this point make me confident that cognitive dispositions of this sort matter for the kind of judge someone would be, they leave me with many unanswered questions. Are some dispositional variables better predictors of reasoning performance--or, better, of judicial performance--than others? If so, why? Is what distinguishes the variables or what they have in common more important for performance? What is it that they have in common? Are these dispositions fundamental or do they arise from something deeper in the psyche? To what extent, if at all, are these dispositions shaped by experience? To take one example, are people simply born foxes or hedgehogs? Might foxiness grow from experiences of mistakes arising from inflexible theorizing? Might one's prospects for learning from mistakes be conditional on some deeper ability to recognize and admit them?

I recognize that these questions are irritatingly vague but feel that we would gain more insight into judging if we could answer them. Among other things, they could help us predict whether efforts to improve judging through education or institutional restructuring were likely to bear fruit. Having a better sense of the fundamental things at work could help us focus on what is most consequential. A better understanding of how different characteristics relate to each other would allow us to construct a fuller and more nuanced picture of which characteristics are likely to be associated with the best performance in which circumstances.
For as Tetlock (2005; Tetlock, Armor, and Peterson 1994) and Sherry (2003) remind us, a trait that seems generally desirable or is typically associated with stronger performance will not always serve us well in all situations. We may prefer the humble judge to the arrogant one for most cases, but when important principles are at stake, society isn't always served best by judicial humility. Foxes may generally do better in thinking about events, but overly flexible or open-minded thinking can lead to dithering or even paralysis. As any trial judge would tell us, there are times when one's responsibility to reach some decision must trump the desire to reach the best possible decision.

Admitting that few if any characteristics are likely to be uniformly salutary may at first seem like a step backward. But this admission points the way to another area of inquiry. Note the similarity to the proposition (discussed above) that we may reason best when we allow affect and intuition to play some part in our thinking. If the best cognitive performance comes from reining in certain normally commendable tendencies where necessary and from blending different kinds of cognitive inputs, perhaps the best judges are those who are best able to perceive and evaluate their mental processes as they occur. If this is right, then

"[g]ood judgment now becomes a metacognitive skill...Good judges need to eavesdrop on the mental conversations they have with themselves as they decide how to decide, and determine whether they approve of the trade-offs they are striking in the classic exploitation-exploration balancing act, that between exploiting existing knowledge and exploring new possibilities."  Tetlock (2005, 23)

I have not progressed very far into the psychological literature on meta-cognition, but my impression is that here too we have reached ground that is firm but not
necessarily stable. For while being aware of what is happening in your mind can improve performance (e.g., Parker and Fischoff 2005), it can also backfire, for instance by bringing things to mind that you would prefer to keep out (Bargh, et al. 1996) or can have both desirable and undesirable effects (Forgas and George 2001). This suggests two additional lines of questioning. First, how deeply can and should thinking about our own thinking go? Can we in fact perceive how our awareness of our cognitive processes itself affects them? If so, how helpful is this ability? Is there a point at which the dangers of interfering with effective thinking outweigh the possible benefits of heightened introspection? Second, how useful is self-awareness without self-control? Do the two tend to go together? How much control over their own cognitive processes can people have? Do some people have much more than others?

**Other individual traits.** I end this section with some brief speculation about other traits that might be desirable in judges. This speculation derives not from particular findings in psychology nor even from truly systematic thinking about the nature of judging. They reflect little more than intuitions that have come to me as these issues have rolled around my head. The main justification for including them here is that I would enjoy hearing other participants' thoughts about these possibilities and others that occur to them.

The other justification is that they move us beyond a fairly narrow focus on avoiding cognitive pitfalls. When we think about what it means to be a good judge, we have in mind more than just the ability to keep from making serious errors. Let us consider two somewhat conflicting propositions that most people would probably agree to.
The first is that judges should promote fairness and justice through their decisions and, therefore, need to show good moral judgment. While recognizing the absurdity of dealing so cursorily with a topic that has troubled philosophers for thousands of years, I would point to one quality that strikes me as indispensable if one is to be a good moral judge, namely the ability to look beyond one's own interests and beliefs and take into account those of others. The concept of sympathy comes to mind, but it does not seem quite right. In a nicely balanced discussion of sympathy, Feigenson (1997) points out that our sympathies frequently do not lie where justice dictates they should and, therefore, that the dangers of acting on them as a judge will frequently outweigh its benefits. What I have in mind is something subtler and more controlled that we might call sympathetic (or perhaps empathetic) imagination. A judge possessed of this quality will have the ability to imagine what the world might look like and seem like through the eyes and mind of someone very different from the judge, whether that person is one of the parties to a case before the judge or, perhaps even more importantly, an utter stranger whose life will be affected by a doctrine announced in the judge's opinion. Is this a quality that varies greatly among people, among people in the pool from whom judges are shown? Is there any way we can measure it? Is it as integral to good moral judgment as I suppose?

The second proposition is that judges should not base decisions entirely on their own moral views. Rather, they should make sure their decisions are grounded in legal texts and established legal principles. Judges have the power to determine the outcomes of cases and announce law. The temptation to use that power to promote what seems right to the judge when the law seems to point in the opposite direction must often be
very strong. So one important characteristic separating superior judges from inferior ones may be the capacity to resist such temptation. Less obviously, differences in how people engage with rules and norms might matter for how well judges deal with this challenge. The norm that judges should decide according to the law is so basic, so widely accepted and taught, that all judges must be deeply aware of it. We would expect judges to have a range of opinions as to the norm's validity, and perhaps this is enough to explain variation in adherence to it. But might a deeper psychological force be at work here? Do people differ in how thoroughly they internalize norms, so that one person could experience more discomfort at transgressing a norm than another, even though they agree about the norm's validity? If so, then following the logic of this discussion, should we give preference for judgeships to the first type of person? Might there be a downside to this character trait?

**Demonstrating applications to judging.** As noted earlier, even if we can identify a number of characteristics likely to have important effects on the kinds of reasoning tasks judges are required to perform, it is another thing to prove that these actually matter for judicial performance or that knowing about them can help us pick better judges. To go further, we would have to measure judges' possession of these characteristics and analyze the relationships between these characteristics and the judges' performance in real cases.

There are two serious challenges here. One is creating a valid and reliable measure of judicial performance. This very difficult task is the subject of another workshop session, so I will say no more about it here. The other challenge is conducting psychological assessments of judges. The most common method of assessment for the
kinds of characteristics we are interested in relies on subjects' responses to standard questions. Perhaps a fair number of judges would be willing to fill out psychological questionnaires, but it is hard to feel optimistic about that prospect. And even if they were willing to, we would have to view the results with some skepticism, since judges may be tempted to portray themselves in ways that comport with traditional standards of judicial temperament.

A more promising approach is to assess judges remotely, through content analysis of their speeches or writings. Scholars have successfully applied this method to a variety of political figures to measure personality traits and cognitive dispositions, including integrative complexity. (For references and several examples, see Dyson and Preston 2006.) To gauge its utility for the project under discussion here, we need to ask two additional questions. First, how broad a range of psychological variables can content analysis be used to measure? Could we use it to assess such traits as self-awareness, the ability to regulate one's mental processes, or sympathetic imagination? Second, what materials would we analyze? Judicial opinions would be the natural choice. But the fact that they are intended for public consumption means that we have to be wary; judges may choose language reflecting how they wish to be seen, not how their minds really work. And the fact that they are written after the judge has ascended the bench make them less useful for identifying people likely to make good judges. Another option, for federal judges at least, is to analyze nominees' responses at Senate confirmation hearings, but here self-presentation is an even bigger problem. Could we devise methods of analysis subtle enough to discern judges true character through whatever smokescreens they throw up in their opinions or answers? Are there any plausible alternatives? I do not
have space to pursue these questions further here but hope we will have a chance to do so at the workshop.

II. Making Judges Better

Although this essay has been concerned primarily with the possibility of selecting people who are likely to be good judges, this is only one route to promoting strong judicial performance. Another is to find ways to help existing judges do their jobs better. We might be able to do this through either training or institutional reform.

Training

Just as not all people are equally prone to cognitive lapses, psychologists have found that even people who are susceptible to them to do not necessarily succumb. Training in reasoning can improve people's performance (Mitchell 2002), as can "debiasing" instructions given at the time of performance (Arkes 1991; Simon 2004). It may well be that judges could benefit from reading studies of decision making or from receiving tips from experts. On the other hand, since judges typically have already received fairly extensive training in decision making by the time they reach the bench, there may be little room for improvement. I have no more to say on this topic but would be interested to hear other participants' thoughts about it at the workshop.

Rules and Institutions
Another way we can aid people's performance is by structuring the situations in which they operate to allow them to do their best. To take a trivial example, many of the tasks judges need to perform are easier to do well if one has ample time. Performance is likely to suffer if judges' caseloads are allowed to get out of hand.

As with our attempt to identify important individual differences, here too we would ideally proceed in several steps, specifying the kinds of mental processes we wish to discourage or encourage in judges, determining the circumstances most likely to do so, and thinking about the kinds of rules and institutions most likely to create those circumstances. That is more than we can accomplish here, but having already discussed desirable and undesirable processes at some length, we can jump right to the consideration of circumstances.

If we ask what situations can enhance cognitive performance generally and are likely to promote good judging in particular, one concept leaps out from the literature: accountability. When people are put in situations where they believe they will have to defend their decisions to others, they often employ sounder reasoning techniques. As Lerner and Tetlock (1998) explain, accountability is by no means a panacea: it does not counter all threats to good reasoning, and it sometimes backfires (see also Pelham and Neter 1995). Still, considering the conditions under which accountability has the best chance of reducing cognitive errors gives us good reasons for thinking it could aid judicial performance. According to Lerner and Tetlock, accountability is most likely to work where poor performance results not from a dearth of reasoning skills, but from insufficient attention. Furthermore,
Self-critical and effortful thinking is most likely to be activated when decision makers learn prior to forming any opinions that they will be accountable to an audience (a) whose views are unknown, (b) who is interested in accuracy, (c) who is interested in processes rather than specific outcomes, (d) who is reasonably well-informed, and (e) who has a legitimate reason for inquiring into the reasons behind participants’ judgments. (259)

Judges already have special training in decision making, and the characteristics listed above would seem to describe the primary audiences for judges' opinions--bench and bar--quite well. So although we should not assume that accountability will always have salutary effects on judicial decision making, it is well worth thinking about.

Engel's (2004) discussion of how accountability might affect judicial performance is excellent. Rather than duplicate it here, I would ask us to focus on two questions: To whom should judges be accountable? How can and should we induce accountability?

As the quotation from Lerner and Tetlock suggests, while accountability may encourage people to do a better job of thinking, the definition of "better" is contextual, with the context shaped at least in part by the thinker's likely audiences. Naturally, different audiences can have widely divergent interests and values, and a judge motivated to please one audience might act quite differently from one motivated to please another (Baum 2006). But even audiences who put aside their own interests and cared only about decisional accuracy might have different views about what information and considerations should properly enter into a decision.

The most obvious audience for judicial decisions is the bench and bar. It is easy to see why it should be useful for judges to feel some accountability to this audience and hard to imagine an argument otherwise. The same can probably be said about legal
academics as an audience. Judges aware that their decisions will have to be justified to members of these audience, who possess the expertise to evaluate those decisions and share professional standards for evaluating them, should more readily avoid important cognitive errors and try harder to ensure that their decisions are adequately grounded in the law.

More interestingly, is there an argument to be made that accountability to the other branches of government or even to the general public can do more than bolster the democratic legitimacy of judicial decisions--that it can actually enhance the quality of reasoning that goes into them? At first glance, this seems highly unlikely. The responses of these audiences will often be based overwhelmingly on their agreement or disagreement with what a judge decided and have little or nothing to do with the quality of reasoning that went into the decision. But if we believe that stepping outside one's own narrow area of experience and trying to see the issues involved in a case from other perspectives is a critical step in good judging, then the argument seems more plausible. We probably do not want our judges to worry too much about whether the average person would agree with a given decision, but what would be the effect on judicial reasoning if judges forced themselves to write opinions that the average person could understand and accept as well reasoned? Might it help them recognize and question modes of reasoning that feel comfortable and familiar to members of the legal profession but in fact are artificial or otherwise inadequate? Might it give them a better chance of escaping confirmation bias or overcoming stereotypes or implicit attitudes that are not obviously repugnant but still should not enter into judging?
Returning to the legal audiences, we may readily agree that judges should be accountable to them, but how do we ensure that they are? An evident solution is to require judges to justify their decisions in writing. A written record increases the likelihood that the judge's actions in that case will be taken into account in any evaluation of his work. Of course, written opinions are standard procedure in the United States, but they are not everywhere. And inside and outside of the U.S., there are practices that undercut the accountability effect of putting reasons in writing. One such practice is non-publication, where the court elects not to submit an opinion to the official reporter who collects and prints the opinions in the bound volumes that go to libraries and offices. In itself, the decision not to publish is fairly inconsequential today, since even "unpublished" opinions are widely available on electronic services. But judges commonly view unpublished opinions as second-class opinions entitled to less weight. This view is formalized in a number of courts with the rule that unpublished opinions are "non-precedential," meaning that judges deciding a similar case in the future have no obligation to follow the logic of the first case or even seriously consider it. A judge writing an opinion that will be unpublished therefore knows that it is far less likely to be carefully reviewed or actively criticized by other judges, lawyers, or academics. I do not mean to suggest that a policy of denying precedential value to unpublished opinions is always unwise, but the cost in terms of reduced accountability should be recognized.

A practice that allows judges to evade accountability even more effectively is the per curiam, or unsigned, opinion. It is true that a panel of judges as a whole can come in for criticism or praise when no individual author is identified, but the motivation to do one's best thinking must surely be less when one is accountable only as part of a group,
not as an individual. Taking this logic to its extreme, we could argue that the best reasoning is likely to emerge when every judge on a panel is required to justify his or her decision. There is much to be said for the American practice of producing a single opinion that speaks for the court, driven home to us whenever the Supreme Court undercuts this practice by issuing a badly fractured set of opinions and we struggle to make sense of the case and its implications. But the possibility that issuing opinions seriatim could lead to a greater overall quality of reasoning by making each judge on a panel feel accountable should be weighed on the other side.

Whether or not each judge on a panel writes separately, the need to persuade colleagues if one's position is to prevail creates another form of accountability, this one found only on collegial courts. This may be an especially effective form of accountability. Engel (2004) makes an intriguing distinction between accountability at different stages of the decisional process, suggesting that being forced to justify one's reasoning at earlier stages can do more to enhance the quality of that reasoning. Engel has in mind stages of a trial, such as decisions on motions, but his insight applies equally well to collegial appellate courts. In these courts, judges routinely interact with each other before final decisions are reached and have to be justified to the outside world: through questions and comments at oral arguments, remarks made during conferences, and criticisms and suggestions offered to each other as opinion drafts are circulated.

The presence of sometimes skeptical colleagues at these interim stages does not only serve as a continual reminder of the need to justify one's decisions. Perhaps more importantly, the exposure to forcefully expressed competing positions can have a powerful effect on the content and quality of judges' thinking. If this logic is correct,
then we may expect, ceteris paribus, that collegial courts will produce more careful and typically more effective thinking than courts where judges operate alone. Recognizing that solo trial judging is unlikely to be replaced by panel judging in those countries, including the U.S., where it is firmly entrenched, is there anything that can make the experience of judging alone more like that of judging on a collegial court? One possibility that comes to mind is to encourage judges to use clerks who are capable of serving as consistently tough critics. Inexperienced clerks who serve at the pleasure of their judges can sometimes fulfill this role, but clerks with longer tenures who are accountable not to individual judges but to the court as a whole would probably do it better.

Having descended into bald speculation, it is time for this essay to come to a close. But as it does, I would ask you to join me in speculating. Debates pitting the adversarial system of judging, where inquiries and the presentation of evidence are driven almost entirely by the parties, against the inquisitorial system, where those with the responsibility of deciding the case take an active part in bringing out the facts, typically focus on questions of fairness to the parties and the likelihood that all relevant facts will come out. What other considerations might emerge if we approached the issue from a psychological perspective? Might it be that active participation at all stages encourages deeper and more effective thinking? Does the presentation of evidence by the two sides significantly enhance the judge's ability to view the issues in the case from more than one perspective? Asking questions like these and thinking through the implications for many other kinds of rules and structures that have not occurred to me could allow us to make valuable contributions to debates on institutional reform.
III. Concluding Thoughts

I would be surprised, to put it mildly, if this essay convinced any reader that the project described in its introduction is worth tackling. But it will have accomplished something if it shows that both psychologists and judicial scholars would have important parts to play in such a project. Most of the essay has been addressed to my fellow students of courts. It ends with a few thoughts about how psychologists could help.

Research into how affect influences thinking is fascinating and holds out the promise of huge implications for judging. I hope that as psychologists pursue this research they seek to learn not only what role affect plays in thinking but also when and how the infusion of affect makes different kinds of thinking more or less effective.

The literature on cognitive dispositions is productive and interesting and may serve the needs of most psychologists very well as it now stands. But it can easily frustrate an outsider hoping to find crisply defined concepts with clear practical implications. It would be helpful to read more systematic comparisons of cognitive dispositions designed to elucidate not only their similarities and differences, but also the ways in which and reasons why they affect people's reasoning. My sense of the literature is that it tends to take reasoning problems as given (e.g., some people use heuristic X even where it is unhelpful) and ask whether a certain disposition might help account for the occurrences of those problems. Perhaps we could learn more by turning the question around, focusing hard on the nature of the dispositions and moving outward from there, imagining all kinds of important effects those dispositions could have on reasoning and
then testing for those effects, whether or not they have been previously identified or studied.

Finally and relatedly, psychologists might be able to offer more to those of us hoping to export insights from the laboratory to the workplace if they were to give more attention to different kinds of reasoning problems. In my reading, I have been surprised at how few of the numerous carefully constructed and otherwise interesting studies of human thinking I have seen have contributed substantially to my thinking about judicial decision making. Probably this stems primarily from my unfamiliarity with this huge literature or the fact that judicial decision is far removed from everyday thinking in many ways. Still, there may be something to my impression that psychologists focus disproportionately on narrow, discrete cognitive tasks. As Nair and Ramnarayan (2000) point out in their study of business management decisions, many of the important mental challenges we that we regularly confront are big, complicated, and messy. Of course we should not ignore smaller challenges and indeed must recognize that large complex tasks will often be made up of smaller simpler ones. But, to my mind, studies of complex processes like Nair and Ramnarayan's and Tetlock's (2005) can be invaluable, and a greater emphasis on such studies within psychology would be very welcome.
REFERENCES


