How Bayesian are Judges?

Jack Knight, Mitu Gulati & David Levi

Abstract
Richard Posner famously modeled judges as Bayesians in his book, *How Judges Think*. A key element of being Bayesian is that one constantly updates with new information. This model of the judge who is constantly learning and updating, particularly about local conditions, also is one of the reasons why the factual determinations of trial judges are given deference on appeal. But do judges in fact act like Bayesian updaters? Judicial evaluations of search warrant requests for probable cause provides an ideal setting to examine this question because the judges in this context have access to information on how well they did on their probabilistic calculations (the officers who conduct the search have to file, in every case, a “return” detailing what was found in their search). Based on detailed interviews with thirty judges our answer to the “How Bayesian are Judges?” question is: Not at all. The puzzle we are left with, given that acting in a Bayesian fashion is normal behavior for the rest of us, is why we get these puzzling results for judges in the search warrant context?
How Bayesian are Judges?

Jack Knight, Mitu Gulati & David Levi*

I. Introduction

A fundamental element of the U.S. judicial system is that trial judges find facts. Judges at the trial level are the first to see a case. They live in the local area, interact with the litigants, the witnesses, the lawyers, and are generally assumed to have a deep understanding of the local context. That understanding is supposed to enable them to make factual determinations with a high degree of accuracy. When we say factual determinations, we don’t literally mean facts in the sense of “the bus was painted blue”. We mean something along the lines of “in this area, the fact that a building has multiple teenagers constantly going in and out, with numerous fast Japanese made motorcycles parked on the curb” suggests a high probability of a drug selling operation. In a different area, those same facts might suggest that the building is the headquarters for a motorcycle racing team. The assumption that judges at the local level develop this kind of deep understanding of context is part of the reason why when cases get appealed, the appellate courts are required to defer to the trial courts on their fact findings.¹ There are, of course, other reasons for deference to the trial judge as well—the trial judge is the one who gets to interact with the witnesses, observe demeanor, hear variations in tone, and so on—much of

* Faculty, Duke Law School. For comments on prior drafts, thanks to Michelle Childs, Andre Davis, Paul Grimm, Don Molloy, Nancy Joseph, Mark Ramseyer, Philip Pro, Richard Posner, Joe Webster, Andrew Wistrich, the students in the Duke Judicial LLM program in 2013 and to participants at the 2015 UNLV conference on the study of federal magistrate judges. Thanks also to Dean Dan Hamilton, Associate Dean Thom Main and the editors of the UNLV Law Review for having put this conference together.

¹ This is explicitly true as a matter of doctrine in the area that is the focus of our research, grants of search warrants. See Illinois v. Gates, 462 U.S. 213, 236 (1983) (directing that the magistrate judge’s determinations on sufficiency of evidence for the grant of a warrant be given “great deference”); Massachusetts v. Upton, 466 U.S. 727, 733 (1984); United States v. Leon, 468 U.S. 897, 914 (1984) (similar). For an articulation of this rationale by Judge Posner, that draws analogy to the negligence context, see U.S. v. McKinney, 919 F.2d 405, 419-420 (7th Cir. 1990) (Posner, J., concurring).
which does not come through in a trial transcript. For our purposes though, what is relevant is that the trial judge is supposed to be able to better predict whether motorcycles plus teenagers equals drug den or racing team.

Implicit in the foregoing—and this is particularly important for purposes of our project—is that judges are assumed to be Bayesians. That is, the model of the judge is one where she is constantly learning; and particularly so as she sees more and more disputes. And as judges improve their understandings of the facts on the ground, they are better able to make legal determinations. They learn things like, for example, that a car with Florida license plates driving around in North Carolina, and that is always careful to drive no more than five miles above the speed limit, is most likely transporting cocaine. They also learn to recognize when it is that that set of facts no longer indicates drug carrying and that instead it is the peppy cheerleader type who tries to make chit chat at the airport with the TSA officer who is likely the newest drug mule. Judges at the appellate level, by contrast, who are often located in some distant big city and never interact with litigants, are not expected to have these deep understandings of the facts on the ground. Their specialty is applying the substantive law to the facts that the trial judge determines.

The foregoing articulation of how judges, and particularly trial judges, behave is often how many of us describe the legal system to our students. It is the model of the fact-finding judge used by Justice Rehnquist to justify his decision to limit the review of magistrate judges granting warrants in *Illinois v. Gates*.

Justice Rehnquist explained:

> [T]he central teaching of our decisions bearing on the probable cause standard is that it is a "practical, nontechnical conception." *Brinegar v. United States*, 338 U.S. 160, 176 (1949). In dealing with probable cause, . . . as the very name implies, we deal with probabilities. These are not technical; they are the factual and practical considerations of everyday life on which reasonable and prudent men, not legal technicians, act." *Id.* at 175 . . . The process does not deal with hard certainties, but with probabilities. Long before the law of probabilities was articulated as such, practical people formulated certain common sense conclusions about human

---

behavior; jurors as factfinders are permitted to do the same – and so are law enforcement officers. Finally, the evidence thus collected must be seen and weighed not in terms of library analysis by scholars, but as understood by those versed in the field of law enforcement.”

Further, the foregoing is the conception of the judge in that has undergirded the work of legal luminaries as diverse as Karl Llewellyn and Richard Posner. Indeed, the description of judges as Bayesian is one that we take from Posner’s book, How Judges Think.

But do judges actually behave in this fashion, like quasi social scientists, where they are constantly learning and updating and making better decisions? Posner does, we suspect. But do others? At first cut, it seems obvious that it must be the case. After all, more interactions with the world improve understanding and judgment for all of us. But judges making legal determinations are not necessarily making decisions in the way that ordinary individuals do. Take Joe, a law student, who is contemplating whether to buy a burger at the law school cafeteria. On the last two occasions that Joe ordered burgers there the meat was overcooked and the buns were soggy. Joe now has a sense that he will suffer a cost, an unpleasant eating experience, if he orders a burger from the law school cafeteria a third time. The judge, by contrast, may not suffer any cost from his incorrectly concluding that a building with broken windows in his town is an indicator of a drug den. Put differently, someone else has to eat the soggy bun that the judge has purchased. And if there is no cost from failing to update and instead there is a benefit--maybe

---

5 Posner’s book, to be fair, emphasizes that different aspect of the Bayesian calculus than we do. And that is the part about how probability estimations are made as a function of “priors” or “preconceptions” – that is, one’s prior understandings of the world. Richard A. Posner, How Judges Think 67 (2010). In a Bayesian model, the priors though get constructed as a function of observation and experience and then get updated as a function of new information that the agent receives. For more, see Sharon McGrayne, The Theory That Would Not Die: How Bayes Rule Cracked the Enigma Code, Hunted Down Russian Submarines and Emerged Triumphant from Two Centuries of Controversy (2011).
the store with the soggy buns is very conveniently located--then it is even more likely that updating will not occur.

Now, it is not quite the case that judges do not internalize any of the costs of failing to act like Bayesians. We suspect that judges do not like getting reversed by a higher court. And even if the risk of reversal is small because the appeals courts have to defer to the trial judges on their fact-findings, there is still some risk. But what if there was no risk of reversal? Would judges still work on developing the kind of deep local knowledge of the facts? After all, there are still other pressures that would be in operation--such as the judge's concern about reputation and the desire for promotion. And this is to say nothing of the potential social costs of failing to update the factual criteria of what constitutes probable cause in these criminal investigations.

This article reports on the results of a study into the question of the degree to which judges seek to update and improve their knowledge of the local circumstances? To do this study, we take advantage of an unusual aspect of the rules relating to search warrants in the federal system. To obtain a search warrant, the law enforcement officer in question has to present the judge--usually a magistrate judge, but sometimes a district judge--with a set of facts that the officer thinks add up to "probable cause". No one quite knows what exactly probable cause is, but we do know generally that it is a probabilistic decision that the local magistrate makes based on information that she is given by a law enforcement officer. The types of information that officers typically present to the judge are affidavits asserting facts observed by the officer or reports from informants who work with the officer. From that information, the judge has to determine whether there is enough of a probability that contraband will be found to justify allowing an intrusion into someone's private space. In doing so the judge must, as she would

---


8 This is as contrasted with what one might call a legalistic decision, where the appellate courts might set down a formula for what kind of evidence it is appropriate to consider while calculating probable cause. This latter approach was rejected by the Court in Illinois v. Gates, supra.
with any witness, take account of the trustworthiness and the reliability of both the officer and the informants.

The most relevant aspect of these rules for purposes of our study is that the officer, after the search is completed, has to file a “return” with the court, where the officer is required to report on the results of the search. This particular aspect is unusual in the federal court system in that judges, in most cases (for example, in sentencing), are not given after-the-fact information about how their decisions have turned out (did the defendant who was shown leniency because of the sterling character references from local politicians continue to misbehave?). In the warrant context, however, the judge has ready access to this kind of information. The officer has to file a document with the court detailing what stuff was found in during the search. To reiterate, that means that the judge has information available to evaluate how well she did in making her probabilistic decision. And it provides information about the reliability of the officer and the related informants on whose assertions the original probable cause decision was based. With enough of this kind of data--judges in busy districts can have over a few hundred warrants that they approve in any given year--the judge can begin to determine which of the officers and informants are more trustworthy and reliable as well as which factors that went into her decision were good predictors and which ones were not.\(^9\) The end result should be, over time, improved predictions regarding whether contraband will be found.

A very important feature of the context that we examine here is that the judges are unlikely to be concerned about the risk of reversal. As a result of the Supreme Court’s decision in *U.S. v. Leon* three decades ago, the decision of a judge to grant a warrant is essentially immunized from appeal.\(^10\) The Supreme Court said that unless it could be shown that the judge acted in bad faith, the decision to grant the warrant would stand even if ex post it seemed like there was not enough evidence for probable cause. The rationale being that since the officers in question

---

\(^9\) As the paper by Lee and Davis for this conference describes, there are many hundreds of thousands of warrant requests evaluated by magistrate judges every year. [Cite to the Lee & Davis paper for this conference]

would have performed the search in reliance on the judge’s determination, no deterrent purpose vis-à-vis the officers would be served.

Our initial plan, when we set up the project was to look at the actual returns for individual judges in a handful of states and examine whether the quality of decision making improved over time (including before and after the decision in U.S. v. Leon which, as noted, essentially immunized the judge’s decision from being overturned on appeal). After a few months of attempting to obtain the data from courts in our state, however, we gave up—we were able to get some data, but most of it was either lost, misplaced or unavailable to us for some other unspecified reason. During the process of trying to find the hard data though, and in particular in talking to various court officers about these returns, we began to realize that our original premise—that judges were surely interested in these returns and were updating their probabilistic calculations with information from them, was off base. And among the clues here were the spiders and dust; it seemed as if no one, let alone judges, had ever looked at these returns.\(^\text{11}\) We turned our focus, therefore, from the returns themselves to the judges and the question of whether they were using that information to update their knowledge bases.

II. The Interviews

During the 2012-14 period, we conducted thirty interviews with magistrate judges and district judges from four states.\(^\text{12}\) We also spoke to roughly a dozen prosecutors, defense lawyers, higher court judges, state judges and court administrators in the states that were our focus. In total, just in terms of the judges, we spoke to thirty judges; twenty-eight of who had been federal magistrate judges at some point in their careers. We identified our judges from the federal court

\(^\text{11}\) The court clerks told us on multiple occasions that they were “thrilled” that “someone” was finally attempting to use the data in the returns. Although, as noted in the text, we ultimately could not get enough of the data.

\(^\text{12}\) The majority of these judges (over 80%) were located in two states in the southern United States. The handful of other judges were from states in the Midwest and the West, whom we spoke to primarily to find out whether the phenomenon we were observing was idiosyncratic to the states we had focused on. That did not appear to be the case.
websites for the states in question. For the lawyers to whom we spoke, we tried to speak to respected senior lawyers within the system.

We conducted almost all our interviews together, with the three of us present and participating. So as to make our interviewees comfortable, we did not record the interviews and instead took notes. The three of us--a former judge and prosecutor (now a law school dean), a political scientist and a law professor--had different priors and expectations about judges, which was in part why we decided to do this project together. What we found surprised and puzzled all three of us.

We conducted the interviews without a fixed set of questions but instead began by explaining our interest in understanding the warrant process, the data we had attempted to collect, and the context of the seminar on judicial behavior that we were teaching. Our first question always was to ask for the subject’s view of the operation of the warrant application process. That is, the mechanics of how the process worked. Our subsequent questions encouraged respondents to fill out their stories with additional detail. The interviews ranged from roughly an hour to two hours.

In the next section, we report on the themes from the narratives, as they pertain to the question at the heart of our paper—how do judges update their views as a result of the information they obtain through the warrant review process. Our impressions are necessarily subjective. In other words, we know what they said they did; not what they actually did. First, however, some background on the judges who constituted the vast majority of our sample; federal magistrate judges.

a. Magistrate Judges

The overwhelming majority of warrant applications in the federal system are reviewed by magistrate judges. In theory district judges can also review these applications, but they rarely do so unless a magistrate judge is unavailable. Given

---

13 We formally began each interview with a statement about the subject’s rights of confidentiality and anonymity.
14 The interviews were almost all conducted in the judges’ offices.
that magistrate judges are roughly 90% of our sample, some background on who 
they are and what role they play may be helpful.

Magistrate judges are an understudied group of judges; the literature on 
judging largely ignores them.15 Their importance within the federal system, 
however, has been consistently increasing over recent years as the workload of the 
federal trial courts has increased.16 Structurally, these judges operate under a 
different set of conditions than do their district court judge colleagues on the trial 
courts in terms of the appointment process they go through, the possibility of job 
loss, salary, pension, support staff, and overall responsibilities.

Magistrate judges in the U.S. federal court system perform a supplementary 
or assisting function to the district court judges who sit above them within the 
hierarchy of the judiciary and who have the ultimate responsibility to handle all of 
the cases filed in the district. One sees and feels that hierarchy within the 
courthouse when one visits the judges at the different levels. Among other things, 
the magistrate judges get paid less, have fewer assistants, and have smaller offices. 
They do not have life appointments, and, for the most part, their duties are as 
assigned by the district court. Magistrate judges perform a wide variety of tasks 
and the specific tasks vary by jurisdiction depending on the direction and needs of 
the district court. Magistrate judges have the statutory authority to make initial 
decisions on juvenile cases, misdemeanor cases, and certain motions. They also may 
make recommendations of fact, sign off on search warrants (our focus), and conduct 
certain civil trials with the consent of the parties and the permission of the district 
judge assigned to the case. Over the past few decades, as the caseloads of the federal 
courts have increased significantly and the number of Article III judges has not, 
 magistrate judges have become an increasingly important part of the trial court 
system, helping ease the burdens on district judges. In 1968, there were 83 full-time

---

15 The conference that this volume of papers comes out of is the exception. Among the handful of prior articles 
on the evolution of the magistrate judge system are, Philip M. Pro & Thomas N. Hnatowski, The Evolution and 
16 See Philip M. Pro, United States Magistrate Judges: Present But Unaccounted For (this volume); Thomas Davis & 
Douglas Lee, “Nothing Less Than Indispensable: The Expansion of Federal Magistrate Judge Authority and 
Utilization in the Past Quarter Century (this volume).
magistrates, 450 part-time magistrates, and 13 clerk-magistrates.\textsuperscript{17} With expansion, in 2011 there were 527 full-time magistrates, 41 part-time magistrates, and 3 clerk-magistrates.\textsuperscript{18}

While these magistrate judges perform many functions that are similar to those that a federal district judge might perform, they are selected via a different system. Magistrate judges are selected based on merit by committees formed by the Chief Judge of each federal judicial district, and officially appointed by the district judges of that district.\textsuperscript{19} Magistrates are required to be licensed practicing attorneys with five years of state bar membership in the state of appointment.\textsuperscript{20} The magistrate judge salary is fixed at 92\% of the district judge salary.\textsuperscript{21} Full-time magistrate judges are appointed for eight-year terms, which are renewable by a majority vote of the district judges in that district.\textsuperscript{22} Essentially, magistrate judges serve at the pleasure of the local district judges, although re-appointment appears to be the norm.

\textit{b. Narratives of Probability}

We began each of our interviews in roughly the same fashion: thanking the judges for making time to see us, describing our interest in understanding their process of evaluating and updating information about local context, and asking the judges if they might describe the warrant process for us. From that point on, we allowed the conversation to flow in the direction that the judge took it in. We did, however, have a set of topics that we ensured that we hit and the discussion that follows is organized as a function of those topics.

Almost all the judge-respondents appeared to have done some preparation in advance of our visits to their chambers in terms of thinking about aspects of the

\textsuperscript{18} These are the latest numbers reported by the association by federal magistrate judges, see http://www.fedjudge.org.
\textsuperscript{20} 28 USC 631(B)(1).
\textsuperscript{21} \textit{Timeline, supra} note 17.
warrant process that we might find interesting. We made certain to assure the judges that we would ensure their anonymity as respondents and would not press for answers to any questions that they felt were inappropriate (these were also the conditions under which we received approval from Duke’s Human Subjects Committee to conduct these interviews). We took pains to emphasize that our interest was in the general warrant approval and evaluation process, rather than any judge’s individual practices. In no case did we ask about individual cases that the judges had seen.

We begin with the descriptions of the warrant process that we heard.

_\textit{i. The Mysterious Role of the AUSA}_

The process that was described to us was essentially the same across the judges. Government law enforcement agents are supposed to call or email the judge’s chambers to inform them that a warrant request is forthcoming. Almost always, this contact is initiated by the U.S. Attorney’s office. A draft of the officer’s affidavit is then sent over to the judge’s assistant so that the judge can look at it ahead of time. A number of judges emphasized that they preferred to receive the draft affidavits ahead of time because they did not want to make the officers wait around in their offices while they were reading their statements. Every warrant, while formally submitted by an officer, came with the imprimatur of a federal prosecutor (an Assistant United States Attorney or AUSA). The informal practice that all of the judges followed was that the AUSA in question would have to either initial the application for a warrant or indicate in some other way to the judge who the responsible AUSA was.

The AUSA’s role in the warrant acquisition process was our first clue as to how ritualized the process was. When we asked about why the AUSAs were involved, the initial explanation we received on a number of occasions was that these AUSAs performed a “certification” or “gatekeeping” function.\textsuperscript{23} We found this

\begin{footnote}
\textsuperscript{23} The idea of lawyers playing gatekeeping roles in other contexts has been the subject of much discussion in the legal literature. \textit{E.g.}, Sung Hui Kim, \textit{The Banality of Fraud, Resituating the Outside Counsel as Gatekeeper}, 74 \textit{Fordham L. Rev.} 983 (2005).
\end{footnote}
interesting, at first, because it looked like the judges had figured out a system by which they could enhance the quality of the warrant submissions by threatening to impose reputational penalties on AUSAs who consistently provided low quality warrant applications either in their quality–clarity and specificity–or accuracy (something that presumably would be discovered using the returns that we discussed earlier).

It was clear from what we heard, we thought, that the officers had only secondary responsibility for the submission at least as to the assertion that “probable cause” to search now existed. In some cases, we heard that it was the AUSAs who even drafted the applications for the warrants. The officers would go to them with the relevant information and the AUSAs would draft the warrants in the appropriate format.

When we pushed the judges on this certification function performed by the AUSAs though, we did not get what we expected. The following is an example:

*Respondent:* The AUSAs perform something of a certification function.

*Us:* That is interesting. Does this mean that judges know where there are AUSAs who consistently turn in lower quality warrants? Do the applications of those AUSAs receive higher scrutiny?

*Judge:* No. That’s not it. There aren’t big discrepancies. We treat each warrant separately. And the agents are . . . good. Plus, there are many agents from different agencies.

As we went through the interviews, it began to dawn on us that what the judges understood to be the AUSA’s certification function was at odds with what we expected it to be. It was clear that the AUSAs were playing an important role, but what was it? What emerged from the interviews was that the AUSAs seemed to be certifying the *legal validity* of the warrant.

But this was puzzling. As a formal legal matter, it is the officer who has personal knowledge of the facts that he is claiming constitute the basis for the probable cause justifying a warrant. And, assuming that the submission satisfied the judge’s notion of probable cause, the officer (not the AUSA) would come in to formally swear to the underlying facts. There does not seem to be any role for AUSA
“translation” of the officers’ facts or the AUSA determining whether the warrant satisfied probably cause. What the judges were describing for us was a process where a key actor was the federal prosecutor. We heard statements along the following lines:

We require the AUSAs to read the warrants and think there is probable cause . . . my threshold requirement is that there is an intellectual investment by the AUSA . . .

Or

The AUSA . . . has made a quality determination.

At this point, one of us would typically interject and ask whether the point of the AUSA initial determination as to whether there was probable cause was something that was useful because the judge could defer to it. The response invariably (and, on occasion, with a dose of annoyance) was the equivalent of “of course not; we would never defer to an AUSA’s determination.” But the question that we were left with though was what role this AUSA was playing if it was neither certification in some reputational sense nor efficiency enhancing in terms of reducing the amount of scrutiny required by the judge.

We heard explanations along the lines of the AUSA evaluation being useful because AUSAs, as lawyers, had ethical duties as officers of the court or that AUSAs could help ferret out errors. But none of these explanations was satisfying. Judges rarely defer to lawyers because they think that the ethical obligations of those lawyers somehow overcome the lawyer’s advocacy role in an adversary system. As for the errors, the AUSAs were supposed to ferret out, how would this work? The AUSAs did not have first hand knowledge of the facts. We were left with the sense that the “certification” requirement fulfilled a variety of different functions in practice although there was no explicit agreement on the reasons for the requirement. We tried to push the judges on this issue. Their answer was that we were missing the point. The AUSAs were lawyers; they were certifying that the affidavit was of high quality as a legal matter; that is, in terms of meeting the relevant standards set down by circuit. To reiterate, what we were hearing was that probable cause was a legal standard and the judges wanted the warrants to meet the
legal requirements. At bottom, in terms of the function the AUSA involvement served, here is what we discerned:

- It ensured that the application would be in line with the legal standards; consistent with “circuit precedent”.
- It notified the judge whom to call if there were some problem with the warrant and it needed to be re-done;
- It put the attorney’s reputation on the line and might lead to the submission of an affidavit that was clearer, better written, more precise, and more persuasive of probable cause.

There was also a suggestion that the certification function might have originated in the judges’ desire that the law enforcement agents not come directly to the judge for a warrant without first getting the go-ahead from a prosecutor. Here the assumption was that in the days prior to certification many warrant applications were defective for a variety of reasons and that a prosecutor either could have fixed the defects or weeded out the bad warrants saving judicial effort in both cases. One judge (perhaps cynically) suggested that the AUSAs might be the only Bayesian updaters in the equation, to the extent that what they were doing was updating as function of their steadily improving knowledge of what sorts of evidence the judge was likely to sign off on and the format he or she liked to see the evidence in. To quote:

AUSAs get trained over time as to what the judges of a particular district find acceptable, and they do not want to be embarrassed by having their warrants routinely rejected and their investigations delayed or disrupted.24

At this stage, we generally shifted gears and moved to ask about the model of the judge’s interaction with the officer, in the event there was a deficiency with the warrant. Our interest was in the question of how officers were receiving feedback from the judges about deficiencies with their warrants and whether that process of

---

24 This is a quote we received from a judge who had seen an early draft of our paper; and therefore was aware of what our initial conclusions were (unlike the thirty primary respondents for this project).
giving feedback to the agents might result in the judges learning more about what the agents were seeing on the ground.

We heard a couple of interesting themes in response. First, the judges differed in terms of whether they saw themselves as having an interactive role with the agents or not. A number of judges viewed the interactive model as inappropriate. It would suggest, they explained, that the judges were part of the prosecution’s team if they were to be seen as helping the officers and the AUSA’s improve their warrants so as to have them pass muster. The responsibility for submitting a satisfactory warrant was that of the prosecution side. The judge had to be neutral.

Some of the judges expressed this bluntly:

*Us:* What was your interaction with the agents like when you had to reject an application?

*Judge:* I just said no. Nothing more. It is not my job to try and educate them.

Or

*Us:* Would you give reasons [for the rejection of a warrant request]?

*Judge:* I typically just say “no probable cause.”

Others maintained a distance, but expressed it in more nuanced terms:

Speaking only for me and my colleagues here, we take a middle ground approach. We make it clear that the judge is not on the prosecution team. It’s not their job to practice law. But, that said, it is not a game. If the information is stale, and here that is rare, but let us say, I get a stale application. I would probably ask “do you have any info more recent than last February?” It frequently occurs, I guess, that there is a small hole - just a clerical matter - I’ll point the AUSA to that. [However,] I seldom tell them how to fix the problem.”

Yet other judges viewed their relationships with the officers and their supervising AUSAs as more cooperative. One explained:

It serves no purpose for me to hide the ball. That just wastes everyone’s time. I try to tell them where I think their warrants are weak. I don’t tell them precisely what to say. But I don’t simply reject the warrant and expect them to read my mind.

One respondent explained that in some cases it was simply a matter of asking the officer questions about why they thought there was probable cause. When the
officers explained what was there, it would often turn out that important pieces of evidence that would have helped persuade the judge had simply not been included in the affidavit. This division among the magistrate judges suggests two sides or expectations of the judicial role: one, that the judge should be neutral, and two, that the judge must explain his or her actions. For us though, the underlying question we were interested in was whether there was learning going on for the judge via the process of interaction with the agent. There may have been some for the judges who were talking to the agents about what was deficient about their warrant applications—but none of the judges we spoke to suggested that any learning was going on (and it was especially unlikely to occur if—as was the case with many judges--the judge was providing his feedback to the agents via the AUSAs).

The other interesting aspect of the foregoing was that it did not appear that many warrant requests were rejected. We did not explicitly ask any of our respondents whether they had ever rejected warrants in a fashion such that it was clear to the officer that a resubmission would be futile. However, the impression we got was that the overwhelming majority of warrants were granted, perhaps after minor additions to the affidavit. Multiple judges told us that the sufficiency of warrants was simply not an issue in the federal context because of the high quality of the submissions by federal law enforcement officers.

For example, one judge emphasized the quality of the preliminary investigations:

If there are legal questions, I will ask the US attorney in charge, the AUSA. 95% of the time or more, it goes quickly. These are federal agents, they're pretty well trained... usually there is a whole lot of investigation. Generally, probable cause is not even close.

Another added that the low threshold for approval:

The US attorney produces high quality product. Probable cause is a low standard. Very... easy.

---

25 A number of the magistrate judges did appear to have been former federal prosecutors and it is possible that the difference noted above was a function of whether the judge had previously been on the other side of the fence in the warrant process. However, we did not collect this data.
On rare occasions, one judge suggested, there might be submissions by officers from divisions such as the Parks Services where the warrant was wanting in some respect. However, this was generally because those officers did not have as much experience with warrants as the typical FBI or ATF agent. Another judge explained:

“These federal officers are very good. They go through a great deal of training. They usually give us so much more than what is needed.

Juxtaposing their experience with federal officers against that with state officers, a point that was made to us again and again was that the issue of weak warrants really only came up when state agents were involved, as they sometimes were in joint federal-state task forces. In conjunction with the above point, a number of our respondents also added that, to the extent there were interesting issues and problems with judicial approvals of warrants, they were going to be in the state system. We were, the not-so-implicit message seemed to be, looking in the wrong place. At this point, there was usually some discussion of the fact that in the state system the “magistrates” authorizing warrants were not always lawyers. The theme that emerged, again and again, was that those on the federal side were clearly superior in quality and status to the state system. One judge expressed it as follows:

The state process is very different, very sloppy. Lawyers do not want to be embarrassed in front of federal judges. Much more shooting from the hip in state courts.

Another offered a more sympathetic interpretation:

I do see state court warrants—you know, frankly, there is a pretty big difference. The training and resources make a difference. They just do 500 cases for every one case we do. I have enormous respect for them.

We should note that overall, despite the lack of interest that the judges seemed to have in acting in a Bayesian fashion, we got little indication in our interviews that the judges did not take warrant submissions seriously or viewed their review as a chore. Instead, the judges seemed to spend considerable amounts of their scarce time tackling these warrants. As mentioned earlier, they would generally insist that

---

26 One former prosecutor did say that a possible exception might be the Manhattan DA’s office, where things would likely be run in a more professional fashion, akin to the federal system.
the draft affidavits were submitted ahead of time so that the judge could read them carefully. There seemed to be almost no delegation of the work on the warrants to law clerks. We asked about this on multiple occasions and the response was always that the task at hand was important; and important tasks did not get delegated to law clerks. By contrast, the drafting of judicial opinions, the judges seemed to be saying, could be delegated. As one judge put it, “No, no use of law clerks. We take search warrants seriously.”

The foregoing struck at least two of us as upside down. Granting routine warrants where almost every one gets approved (after all, the federal officers are so good at their jobs) gets direct attention from the judge. Yet, even though the judge is applying her scarce time to this task, there is little updating of information going on. By contrast, the drafting of judicial opinions to explain the resolution of a complex case--opinions are generally written only if the case has some complexity--can be delegated to law clerks.

Further, a number of the judges seemed to spend time interacting with the officers and AUSAs responsible for the warrants; they appeared to enjoy this interaction. Indeed, a few of the magistrate judges continue to personally “take” the return of the warrant from the law enforcement agent even though the federal rules of criminal procedure no longer require this personal interaction--and even though these very judges did not seem that interested in looking at what these returns actually said. It was the ritual that seemed to be important.

Overall then, the picture seems to be one of considerable judicial attention to warrants, and it is tempting to conclude that judges seem to work hard on even those aspects of their job where they face little risk of “discipline” from a higher court. But there is a different explanation as well. As we looked back over our interviews as a whole, the picture that we saw was one where many of the judges seem to have a high opinion of the federal officers and expect to approve their submissions. Particularly striking were the statements that we heard on more than a few occasions about how the federal warrants typically cleared the probable cause barrier by a wide margin, and that this was unsurprising given the high quality of the officers and their training. The judges see themselves as overworked; their
overwhelming caseload came up on multiple occasions. That meant, we assumed, that the judges were having to perform triage in terms of what cases and tasks to pay more or less attention to. In such a setting, it was only natural that the judges should choose to give less critical attention to the evaluative tasks where the decisions could be prejudged to be easy. And, based on what the judges were telling us, federal applications for warrants seemed to fall squarely within this category of easy decisions that should receive little critical attention—not in terms of time but more in terms of critical scrutiny. Yet, judges give these warrants considerable attention and treat the social dynamics of their interactions with the U.S. Attorney’s offices and the officers with great care.

Our take on this is that the judges see the rituals surrounding the warrant application and grant as having importance, independent of the substance of the probable cause determinations at hand. The whole process, with the swearing on the bible, the need for approvals by the AUSAs, the submissions of the returns, the judge’s personal interactions, the standard chit chat with the officers that would often occur between the judge and the office, all struck us as ritualistic and maybe even ceremonial.\(^{27}\) One judge explained that he wanted to interact with the officers before and after the search because a search is a significant intrusion on liberty. Might it be that the ritual around the granting of the warrant somehow reflects the judges’ continuing sense that the grant of a search warrant is an extraordinary exercise of judicial and executive branch power, even if the vast majority of applications will be granted easily?

\textit{ii. Returns}

As described at the outset, the primary motivating factor for our study was the fact that judicial decisions on search warrants constituted that rare instance where one could, in theory, meaningfully evaluate the quality of a judge’s decision. Or, at the very least, to assess the quality of the testimony and evidence from the officers and the informants that influenced the judge’s decision. If judges were

\(^{27}\) A couple of the judges had special bibles or tables or spaces in their chambers that were dedicated to the warrant interactions.
making probable cause determinations that were improving over time, the returns should demonstrate that. Such data from the returns would reinforce the trustworthiness of the officers and the informants on whom the judges rely. And if the trajectory of the quality of probable cause decisions was staying the same or going downwards, the returns should show that as well. And this negative trend in the data would presumably raise questions for the judge about the reliability of the relevant officers and informants. From this we assumed that judges who were interested in improving their probabilistic decision making would be interested in knowing how good or bad their probabilistic decisions were and what factors, such as the factual assertions of officers and informants, were associated with the unsuccessful searches.

Given that our plan to analyze the return data didn’t pan out, our primary question became one of asking whether magistrate judges were looking at the data on the returns; doing self-evaluations. As a distant second, we were curious about whether the returns factored into the evaluations of the magistrate judges that were performed when they were up for renewal.

On the first point, the overwhelming majority of the judges did not appear to pay any attention whatsoever to the data in the returns; at least, not in terms of reflecting about whether they were making appropriate probable cause decisions. One of us would typically ask at this stage: “Were you curious about what was in the returns? Would it help you learn to make better probable cause decisions if you were to know what fraction of your probabilistic calculations turned out to be correct?” The judges offered a range of answers, but the dominant perspective dismissed even the possibility that information about the returns could be useful.

First, some judges explained that they did not think it would be helpful to look at the returns; at least not in the way we were framing the questions. The returns typically just reported whether the officers found any of the contraband in question or not. It did not indicate how much was found. And to make a meaningful estimation of whether there was probable cause, it was explained to us, one really needed to compare the claim that was made in terms of the amount of relevant evidence that the officers were predicting they would find and the amount of such
evidence that was actually found. Second, some judges saw it as inappropriate to be examining the returns as a way of evaluating the quality of their prior decision. The decision on probable cause was a legal one, they explained, with circuit precedent clearly dictating what constituted probable cause. Whether or not the officers were finding the desired evidence was irrelevant to the decision on probable cause. This second answer perplexed us, since our understanding of the probable cause decision was that it was at best a mixed question of law and fact and mostly an estimation of the probability of finding what the officer was claiming he would find. Indeed, to reiterate a point we made at the outset, Justice Rehnquist explicitly said so in Illinois v. Gates three decades prior (Judge Posner reiterated it with greater clarity in U.S. v. McKinney). But our respondents appeared to see the matter differently, emphasizing that the standard for probable cause was a legal criterion established by precedent and not a practical criterion based on prior experience. One judge suggested:

As a legal matter, returns do not matter. I just signed them when they came in.
These are search warrants, not seizure warrants. It is legally irrelevant whether anything is found or not. I don’t see how it would help to know about the warrants in the returns. I don’t look at [returns].

Another said:

I’m not sure this would be information that would be useful. Probable cause is a zone; it is [a] rough [estimation]. I can’t imagine how information about an agent's success [in the past] would fit in.

Third, one judge said that it was a matter of respect for law enforcement officers. To be looking at the returns as a method of evaluating their submissions, he said, would demonstrate a lack of respect for the agents. Again, we found this perplexing. Wasn’t the judge supposed to be evaluating the officers’ probabilistic decisions; in other words, exercising skepticism about the information that the officers were submitting? Surely, it could not be the case, as a matter of law, that the judges were

---

28 On a couple of occasions we asked the follow up question of whether the returns should be improved so that the judges could get better or more usable information. The response to that was in every case something along the lines of: “No one looks at these returns anyway”.
29 See material cited in note 1, supra.
supposed to be taking all of the submissions by the police officers at face value as true? Finally, some of the judges thought that looking at the returns or inquiring of the officers as to the success or failure of the search warrant would show bias by the magistrate judge because it would suggest that the judge had some kind of interest in the outcome of the warrant and might suggest that the judge was “on the law enforcement team.”

The following exchange captures the dominant approach towards returns that we discovered:

*Us:* [Something along the lines of -- We are interested in the returns on search warrants. Do you ever look at the returns? Would knowing the information on the returns help you?]

*Judge:* Occasionally, I notice problems [with the warrants]. But nine out of ten times there are no problems. It is easy to meet the standard (probable cause). The returns are ministerial as far as we [judges] are concerned. That’s why we have changed the process [to one where there is no need to see the returns any more].

*Us:* [Might the information on returns not help to evaluate the quality of the AUSA?]

*Judge:* As for the returns helping us in the way you suggest [bursts into laughter], it is information. Typically, I see the effects [of granting a search warrant] in terms of arrests. [I ask] Did an arrest occur?

In contrast to the foregoing judges, who seemed quite certain that looking at the returns would serve no purpose, there were a couple of judges who affirmatively insisted on seeing the returns. Some background is useful here. Until a few years ago, as a matter of federal rule, the officers were required to come back to a magistrate judge (not necessarily the one from who they obtained the warrant authorization) and provide the judge with the returns from the search; that is, an inventory of what was found as a result of the search. The officer would then swear to the return in front of the judge. That rule has changed. Now, the returns do not need to come back to the judge. Instead, the returns are filed electronically. In the old system, because the officer had to come in and deliver the return to the judge, the judges were forced, as a structural matter, to see the returns. And an opportunity was created to discuss the actual search. Under the new rule though,
that element of the process had been removed. In effect, the current system requires judges who have an interest in the returns to expend effort to see them and there is no informal opportunity to find out “what happened.” Given what we had been hearing in our initial interviews, we assumed that no judges would be asking for the returns. But there were a couple of judges who were still following the old rule. These judges continue to see the physical return and ask the agent to swear to it. They typically discuss the execution of the warrant and whether there were any problems or surprises in the execution.

When asked why they were interested in the returns though, neither of these judges suggested that they needed to see the returns in order to evaluate the officers or AUSAs or themselves. One of the judges explained that he thought it was important that the judges be seen by the officers as actively involved with oversight of the entire system, so that issuance and execution of warrants did not become mechanical. Another explained that this was simply the practice that his colleagues in the district had adopted and that they found that it worked for them. In other words, both the judges who were asking for the returns disclaimed that they were in any way using the returns to evaluate the officers and AUSA again or test their own judgments.

iii. What if We Provided the Data?

Our final question at every judge-interview was a hypothetical. In the hypothetical, we had examined the data on returns across all of the AUSAs or officers who were submitting warrants to this particular judge, and had prepared a set of summary tables that provided information on what the rate of success of any given officer or AUSA was in terms of the percentage of items that he found (as compared to what he claimed he would find) and on what sets of factors correlated with what levels of findings of contraband.

The judges were willing to consider the hypothetical, but did not, overwhelmingly, think the information would be useful to them. One said that he might have found the information useful in his first couple of years on the bench, when he was still feeling his way around and trying to learn about the officers and
AUSAs, but spending the time on looking at our summary table or regression results would not be worthwhile for him these days. He had already learned all that he needed to about search warrant evaluation. Another explained rather forcefully that providing statistical information on what factors correlated with findings of contraband was not useful for him because his determinations of probable cause were dictated by the law of the circuit, by its case law and not by some computer generated table. Only two judges were even willing to consider the possibility that our table might help them. And only one specifically said that he might revise his views regarding a particular AUSA if it turned out that, over time, this AUSA was always bringing in officers whose warrants found very little. This judge, however, then went on to explain that the AUSAs typically did not last that long in their particular rotations and that it was hard to develop a good sense of what an individual AUSA was going to do (which begged the question of what kind of certification function the AUSAs were performing anyway).

We asked the hypothetical because of our initial assumption—which turned out to be faulty—that one of the reasons for requiring the data on returns was in order for the judges and court administrators to evaluate how effective probable cause determinations were and not simply to make sure that the agents were held accountable for precisely what was taken from the searched location. Whatever the underlying rationale for the rule, however, the federal rules committee appears to have decided a few years ago that it was inefficient to require officers to hand-deliver the returns to a judge; filing them in a central system would be more than adequate.

While there does not appear to be much of an attempt to either evaluate the return data or have it reported in a form that would make its use easier, data on magistrate judges is collected and utilized in their periodic evaluations. The data that is used, more than a couple of our respondents said, was on the raw numbers of warrants a magistrate judge might hear. There is no attempt to analyze the quality of decision making on probable cause determinations (or any others).

At bottom, we asked our hypothetical to make sure that the judges were not ignoring the return data either because of the form in which the data was reported
or because the court administrators were not providing summaries and analyses of the data. With a few exceptions, the judges don’t believe they can learn anything from the data. Their instincts, intuitions, and prior experiences (often as prosecutors) have provided them with enough tools to make the necessary analyses. More data would not help.

Conclusion

How Bayesian are judges? Best we can tell, and in terms of the updating part of the Bayesian equation: Not at all. If anything, the judges are anti-Bayesian in that, for the most part, even if we did the tabulation and statistical analyses of their returns for them, they apparently would not find the information useful to their calculations.

The judges whom we interviewed did not see the probable cause decision as a probabilistic decision about facts but rather as a formalistic decision applying a pre-determined legal criteria. If the petition addresses all of the features of the existing legal criteria of probable cause, then it will be deemed sufficient. For them it appears to be more of an accounting task than a factual determination. While the judges do not seem to care about improving the quality of any conceivable probabilistic estimates, they do care about legal form and ritual. Both in terms of their use of the AUSAs as a screening mechanism and their concern about the legal dictates of the circuit, they make the probable cause determination legalistic and ritualistic even though they are not required to do so.

---

30 We had a glimmer of hope for the Bayesian model during a single interview where about half way through an interview that had been seeming to frustrate our respondent, he said:

I think I...[get] what you are interested in...I do indeed update my views in one context; that of Rufus and Max.

Max used to have pinpoint accuracy in predicting the presence of drugs and I could always trust his judgment. But then he retired and I began to find that Rufus was nowhere near as accurate...[H]e made a lot of mistakes. So, I was less willing to trust him.

Rufus and Max, it turned out, were drug sniffing dogs (we have changed the names of the dogs). To our surprise, two separate judges to whom we sent our draft to for comments remarked on the fact that they had seen discussions of the relevance of past performance of an agent in determinations of probable cause, but only when the agent was a dog. And one judge even pointed us to a judicial opinion on the topic. U.S. v. Anderson, 367 Fed. Appx. 30, 32-33 (11th Cir. 2010).

31 We are not the only researchers to be puzzled by decisions on probable cause. Rachlinski, Guthrie and Wistrich, in studying the impact of the hindsight bias on judicial decisions, found that outcome information did not seem to impact probable cause determinations. See Jeffrey J. Rachlinski, Chris Guthrie & Andrew J. Wistrich, Probable Cause, Probability, and Hindsight, 8 J. EMPIRICAL LEGAL STUD. 572, 597 (2011) (“it is not so much their sense of the likely outcome of the search that influences judges, as their ability to recall a case that would support ruling one way or the other in the case before them.”).
Two concluding questions emerge from our analysis. First, why do judges treat these cases differently than they would other witnesses in other kinds of cases where the trustworthiness and reliability of the witnesses would be relevant to their decisions? And what does this tell us about modeling judges, beyond the observation that they do not seem particularly Bayesian? And here we are puzzled. Acting Bayesian should come quite naturally to all of us, including these judges. So, in a sense, the question is whether the judicial role is somehow constraining them from what we thought would be basic curiosity—wanting to know how one’s decisions turn out. There is something that these judges see about their judicial role that seems to constrain them from seeking out information about the quality of their decisions. And if so why and how? Answering those questions may help us get to the next stage of asking whether this anti-Bayesianism shows up in other forms of judicial decision making beyond probable cause determinations regarding warrants.

A number of caveats apply to our observations of course. They include that: (i) we have a small sample of respondents; they are primarily magistrate judges (arguably different from other judges, in terms of the pressures and constraints they face); (ii) we only know what we were told is done (which might be different from what is actually done); and (iii) the warrant evaluation process might be idiosyncratic and not representative of the ways in which judges behave more generally.

On the first caveat, we should note that our findings had a remarkable degree of consistency. Bottom line: Not a single judge was consistently looking at the returns; nor did anyone think that there was much to be gained by doing so (a couple had looked at them on occasion, but there was nothing systematic). On the second, the magistrate judges didn’t strike us as particularly different from the other judges we have researched in the past (or the handful of district judges we spoke to for this project). But this is ultimately an empirical question.\footnote{The paper by Christy Boyd for this volume suggests that magistrate judges, in terms of decisions, are quite similar to the Article III district judges. See Boyd, \textit{supra} note \_ (this volume). But the research in this area is preliminary. And, as the Tracey George and Albert Yoon paper for this conference tells us, there are systematic difference in, for example, the backgrounds of magistrate judges and district judges. [insert cite to George \\& Yoon paper, this volume]}

32
On the third caveat, we think there is a need for more research. But as a preliminary note, we did present our findings at two workshops where the audiences contained a large number of judges (more than twenty-five at the two put together). At neither session did a single judge say that there were other contexts in which judicial updating was occurring and that this was a special context because of X or Y reason. This is not to say that they didn’t think our research was defective for other reasons (they did) and we note some of those below.

At the first workshop, we presented our results to a group of eighteen judges from a wide array of courts (state, federal, trial, appellate), all of who were enrolled in the Duke Judicial LLM program. To a person, we were not able to convince them that there was a puzzle here. And to the extent they were puzzled, it was as to why we were studying the Bayesian aspect of judging in the first place. At the second workshop, which was in the context of a conference on the topic of magistrate judges, there were ten judges in the audience, nine of who had been federal magistrate judges and one of whom was a state judge at roughly the same level. Here, the widespread reaction from the judges was to say that none of them looked at returns either, and nor did they see any reason to begin doing so after reading our paper.33 They also seemed puzzled as to what the point of our study was.

A somewhat sarcastic comment we received at the end of the first workshop might illustrate the reason for why the judges were puzzled. It was (roughly): “Academics concoct utterly unrealistic theories about how judges behave and then they test the theories against actual behavior and discover that the theories don’t hold up. Wow.”34 There is no doubt something to this point; although we’d respond that the theory of Bayesian judging is highly plausible since it is how most of us behave in daily life. But it also misses the mark in that the theory of judging that we set out to test is one that we would argue underlies the reason why trial judges are given deference on their findings of fact. This is particularly so in the case of warrants where not only is there high deference under Illinois v. Gates, but there is

---

33 After the session was over, one judge came up to us to say that he did look at the returns regularly and was puzzled that the others did not.

34 This was not said in an interview context, so we did not get the exact wording.
almost no incentive for the decisions to be appealed as a result of the good faith exception from *U.S. v. Leon*.

And this leads to our second concluding question: what does the approach that magistrate judges take to probable cause petitions imply for the evolution of the law of probable cause over time? Our judicial system is grounded in the idea that the law evolves to meet changing circumstances through an ongoing process of judicial analysis and reassessment. The set of facts that would constitute sufficient probable cause to justify a constitutionally acceptable search would presumably change over time as the nature of criminal activity changes.\(^{35}\) If the trial judges are not updating their criteria, this evolution in the area of probable cause will be left to the appellate courts. But if the probable cause decisions of the trial judges are, for all intents and purposes, no longer open to rigorous appellate review, the criteria will not be reviewed or reassessed by the appellate judges. This produces an awkward institutional dilemma: with no readily available institutional mechanisms for review, the opportunity for the law of probable cause to change with the times is significantly diminished.

To conclude, we turn back to Richard Posner’s modeling of judges as Bayesians in his book *How Judges Think* that we began with.\(^ {36}\) Have we shown that he was wrong? A key element of being Bayesian is that one updates. And the judges we spoke to, in the warrant evaluation context, do not appear to be updating. In that sense, they are not being Bayesian. Posner’s model though, to be fair, emphasizes a rather different part of the Bayesian equation all through his book. And that is the part about how judges are inevitably going to be making decisions as a function of their priors – that is the understandings of the world that they come to the bench with (some might call these biases). What we are showing, possibly, is that these priors or biases are not getting updated or corrected with new information that the judge receives on the job.

\(^{35}\) For example, consider how changes in technology can affect the nature of criminal acts and thus raise new questions about probable cause.

\(^{36}\) See Posner, *supra* note 5.