

Draft: July 1, 2013

IN THE ABSENCE OF SCRUTINY: NARRATIVES OF PROBABLE CAUSE

Mitu Gulati, Jack Knight and David Levi*

Abstract

This Article reports on a set of roughly thirty interviews with federal magistrate judges. The focus of the interviews was the impact of the Supreme Court case, *United States v. Leon*, on the behavior of magistrate judges in making probable cause determinations. *Leon*, famously, put in place the “good faith” exception for faulty warrants that were obtained by the officers in good faith from neutral judicial officers. The creation of this exception diminished significantly the incentive for defendants to challenge problematic warrant grants, particularly those where the warrant allegedly lacked probable cause. That effect, in turn, could have diminished or enhanced the incentive for magistrate judge scrutiny of the warrants at the front end of the process. We do not find any indication of diminished or increased scrutiny. What we do find, however, is a highly ritualized and formalistic process for the evaluation of warrants where calculations of probabilities are viewed through a legalistic rather than a pragmatic lens.

I. INTRODUCTION

A central question in the design of judicial systems is where to set the level of accountability. Judges can be made accountable in a variety of ways, such as by making their decisions subject to scrutiny and reversal by higher courts, by making their jobs renewable by some external agent such as the electorate, or by having their job performance factor into

* Faculty at the Duke Law School. We are grateful for the comments we received from the seventeen judges who were part of the judicial masters program at Duke Law during the 2012-13 years. For research assistance, thanks to Kim Lott.

decisions regarding possible promotions to higher levels of the judiciary. There are also softer reputational measures that bring to bear professional or public pressure, including academic criticism, reporting of statistics (such as pending motions), and media coverage. The upside of accountability is that judges will be conscious of satisfying those with power over their futures or reputations and will be less inclined to make decisions according to personal values or interests. The downside is the danger that judges will focus all their attention on satisfying those who have power over them or who affect their reputations and standing, diminishing the independence considered crucial for an effective judicial system.

Although the debate over the optimal level of judicial accountability is a perennial one, much of that debate has operated at the level of rhetoric. A small body of research, however, attempts to examine the question empirically. Much of this literature tackles the question in the context of the U.S. state court system. In that system, the fifty states set their accountability levels in at least three different ways (using partisan elections, non partisan elections and appointments). Taking advantage of the variation in systems, scholars have attempted to compare performances of the different state judiciaries to draw inferences about the comparative effectiveness of the various systems.¹ In the context of the federal system, where Article III judges all have life appointments, examining the accountability versus independence balance has proved more difficult. Scholars have, however, attempted to examine the impact of the fear of reversal by a higher court on the behavior of the lower court judges.² Studies in both the state and federal court contexts find that judges seem to adjust their behavior to satisfy the preferences of those who have power over them. That is, the ones

¹ *E.g.*, Stephen J. Choi et al., *Professionals or Politicians: The Uncertain Empirical Case For an Appointed Rather Than Elected Judiciary*, 26 J. L. ECON. & ORG. 290 (2010); Alexander Tarrabok & Eric Helland, *Court Politics: The Political Economy of Tort Awards*, 42 J. L. & ECON. 157 (1999); Andrew F. Hanssen, *The Effect of Judicial Institutions on Uncertainty and the Rate of Litigation: The Election versus Appointment of State Judges*, 28 J. LEGAL STUD. 205 (1999).

² *E.g.*, Kirk A. Randazzo, *Strategic Anticipation and the Hierarchy of Justice in the U.S. District Courts* (Feb. 1, 2008) (unpublished manuscript, available at <http://papers.ssrn.com/>); Joseph L. Smith, *Patterns and Consequences of Judicial Reversals: Theoretical Considerations and Data from a District Court*, 27 JUSTICE SYSTEM J. 29 (2006); Stephen J. Choi et al., *What do Federal District Court Judges Want?*, ___ J. L. ECON. & ORG. ___ (2011); *but see* Christina L. Boyd, & James F. Spriggs II, *An Examination of Strategic Anticipation of Appellate Court Preferences by Federal District Court Judges*, 29 WASH. U. J. L. & POL'Y 37 (2009); David E. Klein & Robert J. Hume, *Fear of Reversal as an Explanation for Lower Court Compliance*, 37 L. & SOC. REV. 579 (2003).

deciding judicial reappointments, in the state context and the ones deciding reversals, in the federal context.³

There is also a broader literature on the impact of accountability on decisions that examines decision makers across the spectrum of society. The literature is too vast for us to meaningfully summarize. However, researchers have found that accountability often makes a difference, not only in the types of decisions, but how decisions makers think about them and justify them. Among the effects that have been found are that more accountability can force greater reflection and self-criticism on the part of decision makers, attenuate biases and reduce over emphasis on irrelevant information.⁴

In this article, we examine a group of judges who have largely been ignored in the research on judicial institutions: United States Magistrate Judges. A Supreme Court decision from 1984, *United States v. Leon*, presented us with something of a natural experiment.⁵ *Leon* is an important criminal procedure case because it established the “good faith exception” to the warrant requirement under the Fourth Amendment. Under this exception, law enforcement agents who obtained a warrant from a magistrate judge in good faith are protected from having that warrant being challenged on the ground that it lacked probable cause. The Court’s rationale was that so long as the government agents were adequately incentivized to ask for warrants, and were doing so in “good faith”, it would advance no Fourth Amendment values to suppress evidence on the grounds that the magistrate judge had erred in granting the warrant.

³ E.g., Joanna Shepherd, *The Influence of Retention Politics on Judges’ Decisions*, 38 J. LEGAL STUD. 169 (2009); Joanna Shepherd, *Money Politics and Impartial Justice*, 58 DUKE L. J. 623 (2009); Joanna Shepherd, *Are Appointed Judges Strategic Too?* 58 DUKE L. J. 1589 (2009).

⁴ E.g., Philip E. Tetlock, *Accountability and the Complexity of Thought*, 45 J. PERSONALITY & SOC. PSYCH. 74 (1983); Philip E. Tetlock, *Accountability: A Social Check on the fundamental Attribution Error* 48 SOC. PSYCH. Q. 227 (1985); Philip E. Tetlock et al., *Social and Cognitive Strategies for Coping with Accountability Conformity, Complexity and Bolstering*. 57 J. PERSONALITY & SOC. PSYCH. 57: 632 (1989); Max Bazerman et al., *Why Good Accountants Make Bad Decisions*, 80 HARV. BUS. REV. 97 (2002).

⁵ 468 U.S. 897 (1984). There was a sister case decided during the same term, *Massachusetts v. Shepherd*, 468 U.S. 981 (1984) that confirmed the adoption of the good faith exception.

The obvious question, and one that was raised in the debate among the justices deciding *Leon*, as revealed by their separate opinions in the case, was the matter of the incentives of the magistrate judges. Putting in place the good-faith exception would effectively reduce to zero the likelihood of reversals (and, therefore, appeals) of the magistrate's decisions on probable cause in granting the warrants. This is because criminal defendants and their lawyers would recognize that, even if the district court judge reviewing the matter decided that the magistrate judge had erred, the evidence would not be suppressed so long as the officers had asked for the warrant in good faith. Magistrate judges, in turn, recognizing that their decisions on this subset of matters were not going to be reviewed, but having numerous other matters where they were going to be subject to review, might respond to this state of affairs either by reducing the amount of effort they put in to reviewing warrants or by giving the benefit of the doubt to law enforcement by finding probable cause where they might not have done so had there been the threat of review. If so, the end result of the good-faith exception might be one where magistrate judges effectively become rubber stamps for whatever warrants were put before them.⁶

The justices writing the majority opinion in *Leon* dismissed this possibility. In their view, the behavior of magistrate judges was unlikely to be negatively affected by the removal of the threats of appeal and reversal. The dissenting judges were not as sanguine, expressing concern about the alternative scenario. And Justice Blackmun, in his concurrence split the difference by, in effect, asking that the new rule be viewed as an experiment. He said that the Court might have to revisit its decision if it turned out, at some point in the future, that the assumptions about judicial behavior made by the majority did not hold up to the evidence. As Blackmun's concurrence implicitly suggests, neither side, however had much of a basis for their predictions regarding magistrate judge behavior. Twenty-five years later, the good-faith exception is an established element of modern criminal procedure. But, contrary to Blackmun's expressed

⁶ Another possibility is that magistrate judges would become more careful and exacting knowing that they were the "last line of defense."

hope, no one seems to have a better understanding of which of the assumptions relating to judicial behavior turned out to be correct.

Three aspects of the good-faith exception in *Leon*, in particular, got us interested in studying its impact. First, we had a Supreme Court case where the justices had explicitly based their decision on assumptions regarding the behavior of lower court judges. One of the criticisms often made of the growing literature studying judicial behavior by more traditional legal scholars and judges is that these judicial behavior scholars study aspects of judicial behavior that are largely irrelevant to anything “law” related. Here was a case and a legal rule where the relevance of understanding judicial behavior seemed obvious.

Second, even if one does accept that a better understanding of judicial behavior is important and “law” related, it is generally difficult to study the impact of changes in institutional factors on the qualitative aspects of judicial behavior (for example, whether a judge becomes a rubber stamp in the absence of any risk of reversal). Here, an unusual feature of the rules relating to warrants made it possible to examine the question. Officers seeking a warrant are required, after they execute the warrant, to report on precisely what they found (this is called the “return”). Using data on returns, before and after *Leon*, one should in principle be able to tell whether the quality of probable cause determinations had improved or deteriorated. If officers, in anticipation of magistrate judge scrutiny, were making better probable cause determinations ex ante, then one should find them returning more relevant evidence. Conversely, if magistrate judge scrutiny after *Leon* had reduced significantly, then the warrants should end up returning a lot less after *Leon* than before.

Third, the judges in question were magistrate judges. These judges are unusual within the federal system in that they go through periodic reviews, at which time a decision is made as to whether to reappoint them. Even if the threat of reversal of their decisions were eliminated, it was possible, we thought, that the moment of accountability might get shifted to the point in the process where the magistrate judges faced their periodic evaluations. Maybe the committees doing these periodic reviews, would examine the performance of the

magistrate judges on the grants of warrants? After all, the re-appointment is made by the active district court judges of the district who, in the pre-*Leon* era, would have been evaluating the quality of the magistrate judges' probable cause determinations because of appeals to the district court.

Our optimism regarding the quality of the data on returns that was available was unwarranted. While we were able to obtain some data from the pre and post *Leon* eras for three separate districts, it proved difficult to determine exactly how effective the searches had been from what was reported on the return forms.⁷ As background to the data collection project, however, we had talked to a number of judges. Even as we found that the numerical data that was supposed to be our primary source of insight was unyielding, the narratives of the judges proved to be an unexpected source of insight into judicial thinking (judicial talking, to be precise) about probabilities and data. That was not the direction we had planned for our project. But what we found surprised and intrigued us enough to report on this. In what follows, we report on what we found in our thirty-plus conversations with judges about *Leon* and the determination of probable cause.

Some brief background on magistrate judges and search warrants may help provide context.

II. THE RELEVANCE OF *UNITED STATES V. LEON*

Judges, particularly at the appellate level, make decisions in a manner that are supposed to set up the rules for future judges to follow. How those rules will be applied in actual disputes will be a function of the behavior of trial court judges. Predictions of the appellate court judges regarding how their lower court brethren will behave in response to a rule, therefore, are important in determining how at least some sets of rules are formulated. *United*

⁷ As we explain later, the returns that we examined did not neatly divide into successful and unsuccessful searches. For example, some of the search warrants involved controlled deliveries. This circumstance would seem unaffected by any change in probable cause review. Many of the warrants involved searches for drugs and evidence of possession or drug dealing. It would require a high degree of subjective judgment to declare a search a success or a failure. Probably one could devise such a project, according to certain assumptions or definitions, but it would require a large data set and would be a large undertaking.

States v. Leon illustrates this point. It provides an example of how assumptions that appellate judges make about the behavior of lower courts impact their decisions. Of interest to us here was that the justices in *Leon* were arguing about how a subset of lower court judges, magistrate judges, would react to the removal of an accountability mechanism, the threat of reversal by a higher court.

THE CASE AND SOME CONTEXT

Leon is one of a line of Supreme Court cases expanding and contracting what is referred to as the “exclusionary rule”. The exclusionary rule derives out of the Court’s decision in a 1914 decision, *United States v Weeks*, where evidence collected in violation of the Fourth Amendment was excluded by the Court.⁸ There is language in *Weeks* that suggests a flavor of concern on the part of the Supreme Court justices about judges seeing themselves as part of the government apparatus, with a responsibility to punish criminals.⁹ The exclusionary rule provided a protection against that, by penalizing the government team for improperly collected evidence. As the exclusionary rule developed in the case law though, the focus was on constraining the behavior of law enforcement – judges began to be viewed more as neutral actors. In Justice Cardozo’s words, “the criminal is to go free because the constable has blundered”.¹⁰ A series of exceptions to the exclusionary rule have developed over the years in areas where the Court has felt that admission of the tainted evidence would not have any deterrence effect on the proverbial constable.¹¹

The question in *Leon* was whether evidence obtained in a search should be excluded if the officers had asked for a warrant in good faith, but the judge had erred in granting the warrant in circumstances where there was no probable cause. Given the focus on deterring

⁸ 232 U.S. 383 (1914).

⁹ *Weeks*, 232 U.S. at 392-393. The Court said at one point: “The efforts of the courts and their officials to bring the guilty to punishment, praiseworthy as they are, are not to be aided by the sacrifice of those great principles established by years of endeavor and suffering which have resulted in their embodiment in the fundamental law of the land.” 232 U.S. at 393.

¹⁰ *People v. Defore*, 150 N.E. 585, 587 (N.Y. 1926).

¹¹ See Susan A. Bandes, *The Roberts Court and the Future of the Exclusionary Rule* (American Constitution Society, 2009 draft) (available at <http://www.nlg-npap.org/html/documents/Bandes--ExclusionaryRule.pdf>).

misbehavior by law enforcement officers, the government argued that, so long as the officers had acted in good faith in asking for the warrant, there was no deterrence value in excluding the evidence because the error in question was by the judge and not the officer. The incentives of the judge were supposedly irrelevant, since the judge was a neutral actor whose errors did not need to be deterred. There were, however, a couple of counter arguments. First, the judges might have biases. If it turned out that the judges in question all came from law enforcement backgrounds (for example, they tended to be former prosecutors), that might produce a bias. Second, putting in place a “good faith” exception to the exclusionary rule had the potential to affect appeals since there was no longer much of an incentive to challenge a search as lacking in probable cause and then seek appeal on the same ground. Given the lack of a threat of reversal, judges might begin taking less care with their decisions regarding warrants, and the police, responding rationally to that, might begin submitting warrants of lower quality. What impact the good faith exception was likely to have, therefore, was squarely a function of how this particular group of judges was likely to behave.

The Court, in a 6-3 decision, accepted the government’s argument that the sanction of exclusion of evidence was an extreme one to apply, where the officers had behaved in good faith.¹² The Court explained that so long as the warrant had been evaluated by a “neutral and detached” magistrate, there was no deterrence value in excluding the evidence.¹³ Not only did the majority opinion take the view that the behavior of the magistrate judges, and the level of scrutiny they applied to warrants, was unlikely to be diminished as a result of the good faith exception, it suggested that the scrutiny level might even increase. According to the majority, lower level judges, in response to a loss of scrutiny by a higher court, might take on the responsibility of self monitoring and, therefore, apply more care to their job of reviewing proposed search warrants. Justice White wrote that “[l]imiting the application of the

¹² 468 U.S. at 926.

¹³ *Id.* at 919.

exclusionary sanction may well *increase the care* with which magistrates scrutinize warrant applications.”¹⁴

Justice Brennan’s dissent made a different prediction regarding the likely behavior of the magistrate judges. His view was that magistrates would get the message that “their decisions to issue warrants are now insulated from subsequent judicial review.” The impact of that would be that “the care and attention devoted to such an inconsequential chore will dwindle.”¹⁵ Brennan also made a prediction regarding what the reviewing judges would focus on (and this logic would apply to judges at higher levels reviewing an appeal, in the rare instances that they might see one). Those reviewing judges, faced with the option of evaluating the probable cause determination of the magistrate or skipping ahead to the easier question of whether there had been good faith would take the second option. If so, probable cause was essentially irrelevant. The end result would be that “the good faith exception will encourage police to provide only the bare minimum of information in future warrant applications.”¹⁶

The majority’s view of magistrate judges was that of unbiased public servants, driven by intrinsic motivations; willing to work harder in the absence of scrutiny from the court above, so as to make up for the lack of the monitoring. The dissent took a realist perspective: Already-busy judges, when they see the removal of appellate scrutiny, engage in triage by working harder in the places where there is oversight, and perhaps the work is more interesting, and exerting less effort where there is little or no monitoring and the work is a “chore”. A third position was that taken by Justice Blackmun, who, in presumably referring to the lack of evidence backing up either the majority’s or the dissent’s predictions, said that the Court might be willing to revisit the good faith rule if evidence showed that some of the empirical assumptions turned out to be false.¹⁷

¹⁴ *Id.* at 917, n. 18 (emphasis added).

¹⁵ *Id.* at 956.

¹⁶ 468 U.S. at 957.

¹⁷ There was little empirical evidence regarding the search warrant process that was available at the time. However, there was one study in a draft form that the opinion in Leon cited. That study, published soon after, was

LEON AND THE LEGAL CANON

Leon is a part of the criminal procedure canon. Every one of the major criminal procedure casebooks contains a discussion of the case; usually a substantial excerpt.¹⁸ As for academic articles and cases, the citations to *Leon* are in the thousands.¹⁹ It seems safe to say that the good faith exception that *Leon* introduced was considered a major legal development.

The basic theme surrounding the discussions of *Leon* in the literature is that it represented a major cutback in the protections provided by the exclusionary rule. It is not unusual to see *Leon* and the good faith exception discussed in the context of assertions that the exclusionary rule is either hanging by a thread, being gutted or in danger of being eviscerated.²⁰ There has, however, been little examination of what impact the good faith exception really had, particularly on magistrate judge behavior. The one exception is a study that was published in 1988, a couple of years after *Leon*, in a criminology journal.²¹ This article found that *Leon* had almost no impact on the ground level behavior of actors such as the

RICHARD VAN DUIZEND ET AL., NAT'L CTR. FOR STATE COURTS, THE SEARCH WARRANT PROCESS: PRECONCEPTIONS, PERCEPTIONS, AND PRACTICES 21 (1984).

¹⁸ Among the casebooks were, RUSSELL L. WEAVER ET AL., CRIMINAL PROCEDURE 596-608 (2007) (redacted portions of the majority opinion and the dissent); RONALD N. BOYCE ET AL., CRIMINAL LAW AND PROCEDURE 1251-1256 (2010) (redacted majority opinion); JOSEPH G. COOK & PAUL MARCUS, CRIMINAL PROCEDURE 708-720 (2009) (redacted portions of the majority opinion, concurrence and dissent); ARNOLD LOEWY, CRIMINAL PROCEDURE 567-582 (2010) (redacted majority opinion with portions of the concurrence and two dissents); STEPHEN A. SALTZBURG & DANIEL J. CAPRA, AMERICAN CRIMINAL PROCEDURE 511-520 (2010) (redacted majority opinion, along with portions of the concurrence and two dissents); YALE KAMISAR ET AL., MODERN CRIMINAL PROCEDURE 114-124 (2002) (redacted portions of majority opinion and portions of the concurrence and two dissents).

¹⁹ A cite count on WestLaw in June 2013 showed that *Leon* has been cited extensively by courts since 1984, with roughly 3,000 federal court opinions citing the decision, and over 2,000 state court decisions citing the decision. It also has more than 1500 law review cites. We have not done a systematic comparison of the citations to *Leon* versus other Supreme Court case. But based on what we know of other court citation studies, these are exceptionally high numbers.

²⁰ Steven Duke, *Making Leon Worse*, 95 YALE L. J. 1405, 1422 (1986) (describing the "gutting" of the exclusionary rule); Robert M. Bloom, *United States v. Leon and its Ramifications*, 56 U. COLO. L. REV. 247, 252 (1984); Wayne R. LaFave, *"The Seductive Call of Expediency": United States v. Leon, Its Rationale and Ramifications*, 1984 U. ILL. L. REV. 895, 930 (1984); Reginald Gordon, *United States v. Leon: The Good Faith Evisceration of the Exclusionary Rule*, 28 HOW. L. J. 1051 (1985); Timothy P. O'Neil, *Exclusionary Rule Hanging by a Thread*, CHICAGO DAILY LAW BULLETIN, February 13 (2009).

²¹ Craig D. Uchida et al., *Acting in Good Faith: The Effects of Leon on the Police and the Courts*, 30 ARIZ. L. REV. 467 (1988).

police.²² Given our focus on federal magistrate judges, the article is somewhat tangential since its focus was on the state system and, that too, only within the first few years after *Leon* was released. Still, the findings are interesting, given the rhetoric about that case and the exception to the exclusionary rule that it created.

III. METHODOLOGY AND DATA

To examine how *Leon*'s introduction of the good faith exception translated into the behavior of actors on the ground; in particular, that of magistrate judges, our initial goal was to look at two sources of information. First, data on warrant applications and returns from three different districts spanning the pre and post *Leon* periods and interviews with roughly thirty federal magistrate and district judges.

In terms of the first source, we were able to obtain data on warrants and returns for three districts (all in the same state) for the period 1974-2002. However, the information on the forms yielded little useable information. Typically, the description of items found on the forms was too vague to assert with confidence that the search had succeeded or had failed. Our second source though, the judges who were reviewing both the warrant applications and the returns and talking to the officers in person, turned out to be a richer source of information than we had anticipated. These interviews with judges, therefore, form the basis of our article. We mention the data collection part of our project here because that was the context within which we conducted our interviews. We would tell our respondents about our data collection project and they would give us their views on the project and what they thought we might get out of it.

The usual caveats with interview studies apply. These interviews represent what our respondents say they do; they do not represent what these same actors actually do. The stories that we report are useful as a source of insight into how a subset of legal actors talk about their jobs. And, in particular, how they talk about data and probabilistic determinations.

²² As of January 29, 2011, the article had received no more than a handful of cites in the law review literature after 20 plus years.

THE DATA ON WARRANTS AND RETURNS

Our research on *Leon* began as a result of a seminar that three of us were teaching in 2009-10 titled “The Study of Judicial Behavior”. The students in the seminar were required to produce research papers at the end of the year-long course and a number of them were having difficulty both coming up with research questions and finding data to answer the questions. We decided to give them an illustration of how easily a project could both be conceptualized and implemented. The basic idea was to take an important Supreme Court case and show the students how empirical assumptions and predictions about how judicial actors will react to changes in law can often be the basis for Court decisions, but that those testable assumptions are rarely examined.

Leon seemed to be the perfect case because it had worked a clear change in the law and expressly rested upon a set of predictions about how actors within the legal system would react. The results, if the foregoing model of judicial behavior held (judges reacting to the reduction in review levels for a particular task by lowering the amount of effort exerted on that task) should show up in the data on returns. That is, warrant applications, as a result of the rule change, should become less successful in terms of the officers finding less relevant evidence after *Leon* than before (assuming other things remained constant, of course).

The data on returns while, in theory, public, turned out to be difficult to obtain. In the districts where we checked, it often was not clear where the old files were (nothing was coded in a form that we could access directly). We did find one court clerk, however, who not only appeared to find the question we were asking interesting, but had kept carefully organized boxes of warrant applications and returns going back to the pre-*Leon* period. We would have to tabulate the data ourselves, but we potentially had access to close to 1000 warrant applications and their corresponding returns from three districts. And, with corrections for documents that had either been misplaced due to clerical error or judicially sealed (mostly juvenile files, we were told), we had the population of warrant applications from three separate districts. It should be relatively easily, we hoped to show our students, how one could simply

estimate the difference in returns between pre and post-*Leon* warrants, control for factors such as differences in judges and subject areas, and draw some inferences as to a largely-ignored subset of federal judges, the magistrates.

We failed for two reasons. First, while we had a number of boxes of data from the pre-*Leon* period, there were not very many warrants there. Worse, there were pre-*Leon* warrants only for one of our districts. Still, we had close to 50 boxes, so there was still the possibility we could say something from the data. The second problem was more substantial. The warrant applications and the returns were too vague for us to do much with. Yes, we had the explanations for why the officers believed there was probable cause, but they rarely made specific predictions as to what they were going to find. Worse, the returns were generally even vaguer, mostly suggesting that something was found, but not giving us enough specifics so as to be able to estimate the success of the probable cause prediction. It would have helped, for example, if, for a warrant that predicted that based on factors A,B, and C, 8 kilos of X—quality cocaine would be found, six kilos of Y-quality cocaine had been found. With the help of a research assistant, we ended up coding 800-plus warrant applications and warrants, but the only meaningful conclusion we can draw from our results are that if one defines success or failure in terms of whether the search yields anything, then over 90% of the searches seem to find something. But that did not seem to be a satisfactory base on which to mount a comparison of pre and post *Leon* probable cause determinations. To provide an illustration, there are a handful of arson investigations in our data where the officers are looking for evidence and ask to collect soil samples that, of course, they succeed in collecting. But it is not clear from the return whether the soil sample subsequently yielded anything relevant to the case.

One other aspect of the data bears mention (whose importance will become evident later). The possibility had been suggested to us that one of the ways in which judges had instituted a system of scrutiny of the warrant applications, to substitute for the lost scrutiny after *Leon*, was to require the federal prosecutor's office to certify the quality and veracity of each warrant; sometimes, by asking the responsible prosecutor to initial the warrant or vouch

for the warrant in some other way as by accompanying the agent to chambers or making the appointment for the agent.²³

Given that we had explained to our students that part of the learning process was to figure out creative solutions to answer one's question, once the obvious data source yielded nothing, we had to come up with an alternative method of getting at our question. The solution suggested itself to us in class since we had been having a series of guests come to our seminar all through the year to discuss the *Leon* case and its implications. Those guests included defense lawyers, prosecutors, appeals court judges, justice department officials, and magistrate judges. The magistrate judges who came to class had been looking at hundreds of these warrants over the years. And the ones who came to our class seemed well informed about what was in the warrants and how they had evolved (unsurprising, in hindsight, given that a substantial part of the job of the judge is to evaluate warrants). A second-best solution might be to simply ask the magistrate judges how the system had changed over time, particularly as a result of *Leon*.

That led to our interviews. To reiterate, we had expected the interviews to be at best a supplement to our quantitative data.

There were two main questions that we were hoping to gain insight about. Embedded in those questions were assumptions regarding what we expected to find.

First, what was the impact of the good faith exception in *Leon* on judicial behavior? Given the importance of this case in the law school canon and the explicit disagreement among the justices about the likely impact of the *Leon's* good faith exception on judicial behavior, we expected this to be a topic that judges would have strong opinions about.

Second, how were the judges processing the information on the quality of their probable cause determinations? This information, on returns, would, in theory, tell them whether their

²³ This is not to suggest that *Leon* had anything to do with causing federal prosecutors to take on a role in scrutinizing warrant applications. That practice likely existed even prior to *Leon*. Rather, the point we heard made was that perhaps, after *Leon*, the role of the federal prosecutors as gatekeepers was enhanced.

probable cause determinations were improving or worsening? As noted earlier, warrant applications for probable cause are one of the rare instances in the judicial system where the judge clearly gets to see the impact of her probabilistic calculations. This is because the officers have to bring the judge a report on what they found (the return). Judges, we assumed, would be constantly updating their prior understandings of how best to calculate probabilities with this information about returns.²⁴ The suppression law that has developed gives quite a bit of deference to the probable cause determination of the magistrate judge based on that judge's experience and knowledge of the local community. It is unusual that judges have the opportunity repeatedly to make a certain kind of factual judgment – here, probable cause – and then see if they got it right. We were interested in whether the magistrate judges earned the deference they have been accorded by keeping their own score and refining their judgments going forward.

THE INTERVIEWS

During the 2011-13 period, we conducted roughly 30 interviews with magistrate judges and district judges in two states in the southern part of the United States. The goal was to talk to a majority of district and magistrate judge in these two states. We also spoke to a handful of prosecutors, defense lawyers, higher court judges, state judges and court administrators. In total, just in terms of the judges, we spoke to 30 judges; 27 of whom had been federal magistrate judges at some point in their careers.

We identified our judges from the federal court websites for the states in question. For the lawyers to whom we spoke, we tried to speak to respected senior lawyers within the system.

One of us is both a former federal judge and a former U.S. Attorney. That, however, may have created a bias in what we were told since our respondents knew that at least one of us had many years of experience tackling the same issues that they were talking about (albeit, in a different context). Perhaps it has affected our own outlook as well.

²⁴ Cf. RICHARD A. POSNER, *HOW JUDGES THINK* (2010) (articulating a model of judge as Bayesian updater).

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, 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. Much of what we found, however, surprised 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 were attempting 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 warrant application process.²⁵ 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.²⁶

In Section IV, we report on the themes from the narratives, as they pertain to the question we started out being interested in – the effect of a change in accountability on judicial behavior. Our impressions are necessarily subjective. What we can primarily report on is the ways in which judges from the trial courts talk about a particular aspect of their job and the impact of a Supreme Court case that is said to be of great importance by the academic and legal policy communities.²⁷

MAGISTRATE JUDGES

Magistrate judges are an understudied group of judges; the literature on judging largely ignores them.²⁸ Their importance within the federal system, however, has been consistently

²⁵ We formally began each interview with a statement about the subject's rights of confidentiality and anonymity.

²⁶ The judge interviews were almost all conducted in the judges' offices.

²⁷ For discussions of methodology similar to ours, see, e.g., John M. Conley, *Tales of Diversity: Lawyers Narratives of Racial Equity in Law Firms*, 31 LAW & SOC. INQUIRY 831 (2006) (using the ethnographical method to study law firm diversity); Lissa Broome et al., *Dangerous Categories: Narratives of Corporate Board Diversity* 89 N.C. L. REV. 720 (2011).

²⁸ Among the handful of articles on the evolution of the magistrate judge system are, Philip M. Pro & Thomas N. Hnatowski, *The Evolution and Administration of the Federal Magistrate Judges System*, 44 AM. U. L. REV. 1503 (1995); Jack B. Streepy, *The Developing Role of the Magistrate in Federal Courts*, 21 CLEVELAND ST. L. REV. 81 (1980).

increasing over recent years as the workload of the federal trial courts has increased. 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, 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 magistrates, 450 part-time magistrates, and 13 clerk-magistrates.²⁹ With expansion, in 2009 there were 521 full-time magistrates, 48 part-time magistrates, and 3 clerk-magistrates.³⁰

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

²⁹ *A Timeline of the U.S. Magistrate Judge System*. The Third Branch, October 2008. http://www.uscourts.gov/News/TheThirdBranch/08-10-01/A_Time_Line_of_the_U_S_Magistrate_Judge_System.aspx.

³⁰ <http://www.fedjudge.org> (website for the Federal Magistrate Judges Association).

judicial district, and officially appointed by the district judges of that district.³¹ Magistrates are required to be licensed, practicing attorneys with five years of state bar membership in the state of appointment.³² The magistrate judge salary is fixed at 92% of the district judge salary.³³ Full-time magistrate judges are appointed for eight-year terms which are renewable by a majority vote of the district judges in that district.³⁴ Essentially, magistrate judges serve at the pleasure of the local district judges, although repeated re-appointment appears to be the norm.

Our original interest, as mentioned, was in doing a quantitative study of the impact of a reduction in judicial accountability in the probable cause area. The *Leon* case was our centerpiece because it arguably reduced the level of judicial accountability. But that goal did not materialize. What we collected were a set of narratives of how judges talk about the use of data and the making of probabilistic decisions.

IV. 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 the search warrant process in the wake of *Leon*, describing the data that we were in the process of collecting, 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 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

³¹ Tim A. Baker, *The Expanding Role of Magistrate Judges in the Federal Courts*, 39 VAL. U.L. REV. 661 (2005).

³² 28 USC 631(B)(1).

³³ *Timeline*, *supra* note ____.

³⁴ Lawrence O. Anderson, *United States Magistrate Judge: The Utility Fielder of the Federal Courts*, ARIZONA ATTORNEY (January 2007), 10, 11.

inappropriate. We also 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.

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 US 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 overall 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.³⁵ We found this 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).

³⁵ The idea of lawyers playing gatekeeping roles in other contexts has been the subject of much discussion in the legal literature. *E.g.*, Sung Hui Kim, *The Banality of Fraud, Resituating the Outside Counsel as Gatekeeper*, 74 *FORDHAM L. REV.* 983 (2005).

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 probable cause. To take this back to *Leon*, in all of the discussion of the incentives of the various parties in *Leon*, there was no mention of the role played by

federal prosecutors and their incentives. Yet, 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 particularly 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: 1. It notified the judge who to call if there were some problem with the warrant and it needed to be re-done; 2. 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; and 3. It put the judge on notice that the warrant application

might need closer scrutiny if the assistant had a reputation for sloppy work. 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.

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, in the wake of *Leon*, officers were receiving feedback from the judges about deficiencies with their warrants. After all, one of the impacts of *Leon* was likely to have been that there would be very few appeals attacking the grants of warrants. And that in turn meant that there were unlikely to be judicial opinions performing a teaching function with regards to what constituted a satisfactory warrant.

We heard a couple of interesting themes in response. First, the judges radically differed in terms of whether they saw themselves as having a teaching or explanatory function; that is, a role where they would help explain to the officers/AUSAs how their warrants should be improved. Some judges viewed this kind of interaction as inappropriate. It would suggest 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: In the first part of your time, would you interact with the agents?

Judge: No. I just said no.

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, I 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. 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.³⁶ 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. The tension arises here because the warrant application process is necessarily *ex parte*. That is, the defendant isn’t there to argue about whether the evidence is sufficient or not.

³⁶ 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.

The second 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 (that, in some cases, the judge might have suggested). 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 (and, perhaps, this relates back to the involvement of the AUSA).

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.

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 “judges” 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.³⁷

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.

One question we were specifically interested in was whether, in the wake of *Leon* and the good faith exception, magistrate judges had begun rubber-stamping warrants. 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 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.”

³⁷ 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 foregoing struck 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) get direct attention from the judge. 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. 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 seemed to 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 these rituals surrounding the warrant application and grant have taken on importance for the judges, quite independently of the substantive importance or difficulty of the probable cause determination. These interactions with law enforcement officers and AUSA are important to the judges. The whole process struck us as having a strong whiff of ritual, with the swearing on the bible, the need for approvals by the AUSAs, the submissions of the returns, the judge's personal interactions, the chit chat with the officers that would often occur between the judge and the officer. 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?

RETURNS

One of the motivating factors 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. If judges were making good probable cause determinations, the returns should demonstrate that. And if judges were making questionable probable cause decisions, the returns should show that as well.

Our question for the judges was once again, a function of the accountability question. In the absence of an appeal process where a district judge would be deciding on whether the magistrate judges had made appropriate decisions on probable cause, we had speculated as to the possibility that the data on returns might be a source of information on the same question. Indeed, over a large enough set of decisions, the data on returns for any individual judge might well be a better gauge of performance. In theory, the use of this data could potentially occur in two ways. First, magistrate judges might be looking at the data themselves; doing a sort of self-evaluation. Second, the returns might factor into the evaluations of the magistrate judges that were performed when they were up for renewal.

On the first point, we found that a 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 discounted the usefulness of the return information.

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 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. But these judges saw it differently, emphasizing that the standard for probable cause was a legal criterion established by precedent and not a practical criterion based on prior experience. For example, one judge suggested

As a legal matter, returns do not matter. I just signed them. 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 warrants..... I'm not sure this would be information that would be useful. Probable cause is a zone-it is rough. I can't imagine how information about an agent's success 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. 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 supposed to be taking all of the submissions by the police officers 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 there are no problems. It is easy to meet the standard (probably cause). The returns are ministerial as far as we [judges] are concerned. That's why we have changed the process [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 other 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. To our surprise, there were a handful 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, none of the 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. All of these 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.

Even those few judges who saw the AUSAS’ involvement in the warrant application process as providing a certification function did not appear to be using the returns to evaluate the quality of probable cause determinations. At least, not in a probabilistic fashion. We tried to push the judges on this issue. Their answer seemed to be 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. Once again, what we were hearing was that probable cause was a legal standard and the judges wanted the warrants to meet the legal requirements.

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, for the most part, 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 would not be worthwhile for him these days. 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 a couple of 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.

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.

THE IMPACT OF *LEON*

Ultimately, perhaps the most puzzling finding from our research has been the utter lack of perception of any significant impact of *Leon* by any of the actors in the system. As one judge put it, "I have yet to come across any agents hanging their hats on *Leon*." In addition to the judges, who were the primary focus of our study, we also spoke to a number of federal prosecutors and defense lawyers. All of them struggled mightily to find any plausible impact of *Leon*. We say "struggled mightily" because we spoke to a number of these lawyers in the context of a seminar that we ran over a year ago that was centered around the impact of *Leon* and at least a few of our guests mentioned (after the class) that they had felt embarrassed about coming to our class and saying, in front of our students, that the case that we had picked to center our course around had been utterly irrelevant. One guest, a former federal

prosecutor, suggested that it might be the case that *Leon* was particularly important in states where the state itself had not accepted the good faith exception. There, this former prosecutor said, there was the possibility of the silver platter phenomenon – where the state prosecutors, because they had obtained their search warrant via a submission that lacked probable cause (and were, therefore, likely to have evidence excluded because they had no good faith exception) would walk over to the federal prosecutor’s office to deliver their case on the proverbial silver platter. This scenario, as it was narrated, sounded right out of an episode of *Law & Order* and got one of our seminar students excited enough to do an empirical paper investigating the silver platter phenomenon. According to his respondents, they had never heard of the scenario occurring – and they emphasized to him that it was unlikely that the federal prosecutors would ever want a state case with poorly obtained evidence.

We think there is something quite puzzling about the purported irrelevance of *Leon* though. For our respondents, as best they either could recollect or they had heard from others, *Leon* had largely been irrelevant right from the beginning. And the reason was that the warrant submissions of federal officers consistently came in well above the probable cause bar. Apparently, there were other factors driving the consideration of when a warrant request would be submitted and federal officers, being time constrained and having lots of demands on their time, were unlikely to ask for a warrant unless they thought it was highly likely they would find something helpful to their case.

In addition, there was the matter of the judges requiring that the officers obtain the approval of an AUSA before making a submission. As one judge, who was also a former prosecutor, explained it, the fact that the officers had no choice but to go through the AUSA gave the AUSA some leverage over the officer to push her to do more work on the case than might be needed just to satisfy probable cause. The officers, it was explained to us, largely washed their hands of the case, once the search was completed, since that was generally the final stage before the case got put into the hands of the prosecution for the indictment/litigation (or, more likely, plea agreement) phase. The AUSAs, therefore, had their

own incentives to withhold their signatures until they had a good case not so much for probable cause but for the prosecution that likely would follow upon the search.

Our own examination of the warrants themselves bears out the foregoing. While at least two of us are no experts on probable cause, and it is not even clear what probability standard probable cause is supposed to represent, it was hard to find more than a warrant or two among the many hundreds of warrants that we looked at where there did not seem more than enough indication that there was some contraband. And, in the vast majority of cases, there was at least some relevant evidence.

So, the judges were probably right. *Leon* does not appear to have produced a lowering of the quality of search warrants submitted. And, while we cannot tell for sure, we suspect that the judges are still exercising just as much scrutiny of the warrants as they were pre *Leon* (no one claimed that the certification by the AUSAs was a function of *Leon* and AUSAs were likely certifying warrants even prior to *Leon*).³⁸ The question then is: Why was *Leon* considered such a big deal at the time it was litigated and soon after? To look at the debate in the literature over the good faith exception, to talk to those who were prosecutors, and to read the opinions of the justices, all suggest that *Leon* was supposed to be a very significant case. On the one hand some claimed that it was eviscerating the Fourth Amendment and others argued that it would be a great boon to law enforcement. Instead, not much of anything appears to have happened as a result of the case.

Thus, one additional question that interests us is why it was in the interests of so many of those involved in both the decision itself (after all, a choice had to be made to utilize resources in arguing the case all the way to the Supreme Court in order to obtain the good faith rule) and in the subsequent debate to characterize the case as either a major cutback in Fourth Amendment rights or a major readjustment of the suppression rule in favor of federal law enforcement. In reality, neither occurred, nor does there appear to have been much reason to

³⁸ The Uchida et al. study of the state courts mentioned earlier finds, in that context, that there was variation across jurisdictions in terms of whether certification by the district attorneys was required by the magistrate judges. See *supra* note ____.

expect either to. The answer, we suspect, was that it was in the interests of some sets of interests on both sides to either posit that a problem had been solved (improved law enforcement) or that one had been created (constitutional rights being eviscerated)

V. IMPLICATIONS

In this study, we sought to address two related questions. First, what was the long term impact of *Leon* on the criminal process in the United States? Second, what has been the effect of a change in the institutional mechanisms of accountability on the decision-making of magistrate judges? Our research allows us to make some preliminary comments about these two issues. And it also allows us to say something about the complexity of the task of undertaking empirical research on judicial impact.

To assess the impact of *Leon* on the criminal process, one must have a clear understanding of how to conceptualize impact and then of how to empirically measure it. If we define impact in terms of institutional changes in the formal legal process, then *Leon* caused an institutional change in the process by which probable cause is assessed and evidence is obtained. Prior to *Leon* magistrate judges played a central role in assessing the likelihood of probable cause and district judges were commonly asked to review those assessments in suppression hearings initiated by motions brought by defense attorneys. These district court reviews served as a mechanism of accountability for magistrate judges. Post *Leon* the institutional process appears to have changed in two significant ways, although the first change is not attributable to *Leon* so far as we can tell. On the one hand, now, while the formal authority to approve the warrant remains with the magistrate judges, much of the practical authority has been ceded to the U.S. Attorney's office which vets warrant applications before they are presented to a judge. Without acknowledging a rule of deference, the importance that the magistrate judges place on the certification process for warrant petitions comes through loud and clear. Given everything that we were able to discern from our interviews, the fact that an AUSA has certified the petition before final submission to the judge goes a long way towards satisfying the criterion for approval both because that review probably makes a

difference to the adequacy and quality of the affidavit and because of an unstated rule of a kind of deference. On the other hand, the moment of accountability for magistrate judges through the review of the probable cause decision by the district court has essentially been eliminated. This follows from the fact that defense lawyers no longer see a suppression motion for the evidentiary search as a viable strategy due to the *Leon* good faith exception.

As an alternative, if we define impact in terms of the effects of the *Leon* decision on the substantive outcomes of the criminal process, then we must turn our attention to the effects of these institutional changes on the actual decisions that the judges make. And this puts the focus directly on the effects of the changes in accountability on the quality of the decision-making of the magistrate judges. Our interviews suggest that, despite the importance that the magistrate judges place on the certification by the AUSA's, they continue to take the warrant petition process seriously and spend time reviewing the details of the petitions. In one sense, this tends to undermine the idea that the elimination of an institutional mechanism of accountability would significantly diminish the amount of time and energy that magistrate judges give to the warrant process.

But any comprehensive assessment of the substantive effects of *Leon* on the process would need to delve beyond the nature of the decision-making process and investigate the substance and quality of those decisions. One measure of the substantive effects of *Leon* might involve an assessment of changes in the rate of convictions that occurred after the introduction of the good faith exception. If we were able to control for other potential causal factors, we might get an assessment of the impact of the revised standard for evidentiary exclusion on criminal prosecutions.

A more refined and normatively significant measure of impact might be the effect of the revised procedures and standard on the rate of convictions that were based on evidence gathered without probable cause. On the one hand, such a measure would, in principle, capture our concern with the protections of the Fourth Amendment and allow us to assess the incentive

effects of the introduction of the good faith exception. On the other hand, it is not obvious how we would construct the necessary counterfactuals to develop such a specific measure.

Furthermore, each of these two proposed measures would place the actual focus on the decisions of the district judge and only very indirectly on the decisions of the magistrate judges. Our primary focus is on the magistrate judges and the quality of the probable cause decisions themselves. This is what we had hoped to capture in our analysis of the warrant applications and returns.

Our analysis of the warrant applications was intended to assess the effects of the *Leon* changes on the quality of probable cause decisions by the magistrate judges. Our plan was to use the following measure of quality, did the warrant actually return the anticipated evidence that was used to justify the application? We hoped to assess the potential rates of change in success before and after the *Leon* changes in the process. We hoped to answer the question that Justice Blackmun left hanging. But, as previously reported, the information included in the returns was not sufficiently detailed to conduct this kind of analysis.

And once we had interviewed the judges, it became clear that they had a different conception of the nature of the probable cause decision-making process, a difference that called into question our original measure of quality. That is, our original measure is based on the view that the probable cause assessment is a practical decision, do the existing circumstances submitted in the warrant application support a reasonable belief that the police will in fact find the evidence that they claim exists? By analyzing the success rate of the returns, we thought that we could assess whether the magistrate judges did as good a job of making this assessment after *Leon* with the elimination of the existing mechanisms of accountability as they did before *Leon*.

But our interviews revealed that most of the judges conceive of the probable cause assessment as purely a legal decision, not a pragmatic or practical one. From their perspective, the probable cause decision answers the question, have the prosecutors included in the warrant application sufficient information to satisfy the criteria set out in prior judicial

decisions? For this question, the actual returns are irrelevant. Nothing about the ex post fruits of the search are relevant to an assessment of the quality of the probable cause decision from the magistrate judges' perspective. The more relevant measure of the impact of *Leon* on their legal conception would involve the success rate at which the judges properly applied the probable cause criteria established by previous decisions to their review of the warrant applications.

This is an approach to the question of the judicial impact of *Leon* that we had not considered in our analysis of the data. It is but one instance of the various complex questions that we confront when we undertake empirical analyses of the courts. It is not yet clear to us which is the more appropriate measure of substantive impact to employ. As with many areas of empirical analysis, it may depend in the end on the particular question we intend to ask. But it will be nonetheless necessary to resolve the conceptual question in order to give an adequate account of the actual impact of what so many in the broad legal community considered an important Supreme Court decision.