The Inadmissibility of Guilt Rate Evidence

By Gideon Yaffe

1. Introduction

The question of what evidence attorneys ought to be permitted and prohibited from showing to fact-finders—the normative issue needing to be addressed to rationally evaluate a body of evidential admissibility law—is intimately entwined with the further question of whether a piece of evidence ought to be considered by the fact-finder. Setting aside some complexities[1], if evidence ought not to be considered—if something significantly bad takes place when the evidence is weighed—it ought not to be admissible. If attorneys are allowed to show something to the fact-finder, some will, and the fear is that anything the fact-finder might see, they might

[1] The complexity derives from the fact that 'ought' implies 'can'. Fact-finders have shirked no obligation in failing to consider evidence never presented to them. Since it does not follow from the fact that evidence is *permissibly* presented that it is, *in fact* presented, it does not follow from the fact that evidence is admissible that the fact-finder labors under an obligation to consider it. Or, put equivalently, if evidence ought not to be considered because it was never presented, it still might have been permissible to present it. So it is not strictly true that evidence ought to be admissible only if the fact finder ought to consider it. Still, what does seem to be true is that evidence ought to be admissible only if nothing will have gone wrong were it both presented to and considered by the fact-finder.

weigh. Since we have no idea how often the evidence will be presented to fact-finders, and so no idea how often the bad state of affairs realized by their weighing it will come to pass, the potential problem that is generated could be arbitrarily large. This problem is exacerbated in circumstances in which either the defense or the prosecution will benefit from the bad state of affairs, and so has an incentive to present the evidence to the fact-finder. So, if fact-finders' weighing of the evidence would be significantly objectionable, attorneys should not be allowed to show it to them.

This thought underlies some well-entrenched rules of American admissibility law. For instance, under the law, if a piece of evidence's probative value is outweighed by its prejudicial influence, then it is inadmissible.[2] To be admissible, the evidence must be suited to play a greater role in moving a bias-free fact-finder's judgment of the chances of guilt or innocence than it is suited to prompt such judgments in people thanks to typical biases. The gory photo of the body must speak more to the question of whether the defendant did it than viewing the photo leads a typical person, subject to biases as we all are, to desperately want someone, anyone, to pay for the crime. The problem with permitting the introduction of evidence that is more prejudicial than probative is that, in the typical case, something will have gone terribly wrong if it is presented to the fact-finder and weighed: there is a significant risk that the resulting verdict will be more reflective of bias than rationality.

This same normative theory, under which it is impermissible to present a piece of evidence to a jury if something would go seriously wrong were it to consider it, underlies the bar on admitting evidence that incentivizes unacceptable government methods for collecting

[2] [[ref to FRE]]

evidence.[3] A plausible confession, induced by torture, is inadmissible even if far more probative than prejudicial, for instance, since admitting the evidence encourages the government to collect evidence through torture. The problem here is that were the confession presented to the fact-finder and weighed, something will go wrong: the government will be incentivized to keep torturing people for confessions. Admit tortured confessions, and there is no principled limit to how strongly such government behavior will be incentivized.

Acceptance of explanations of legal rules of admissibility of the sort just rehearsed depends upon applying a standard of justificatory explanation suitable, primarily, for public policies. What makes it a good idea to institute a particular rule as a public policy is that *in the typical case*, following the rule avoids a problem without any significant loss. Of course, for any set of rules, we can think of examples in which following the rule does not avoid the problem, and examples in which there are significant losses to following it, as well as examples in which both things are true. But, still, the thought is that these are not typical cases and what we want from public policies are good ways of handling the typical cases. (This kind of reasoning is closely linked to rule consequentialist approaches according to which people do the right thing just in case they follow rules that pass the following test: following them will result in best consequences in typical cases.) So, the reason that it is good to have the rule that bars admission of confessions obtained through torture is that, in the typical case, by excluding such evidence we do not give the government incentive to inflict torture and we are still able to achieve an accurate verdict. Of course, we can construct untypical cases: those in which the government is incentivized to torture by the evidence's exclusion, those in which verdict accuracy is lost by

[3] [[ref to FRE]]

excluding the confession, and those in which both things are true. But since these cases are untypical, the explanation is a good one: it identifies a good reason in favor of instituting the rule as public policy, as we do when we make it law.[4]

It is universally recognized that information about the guilt rates among the accused is inadmissible and should be. Say criminologists have determined that 90% of those arrested are guilty of the crimes for which they are tried, or, alternatively, that criminologists have found that only 10% are. The prosecution in the one case, the defense in the other, would benefit were the jury to see this evidence. Should it be admissible? Of course not. But why not? It is clearly probative and it does not appear to be prejudicial. It is true that, in the typical case, it would increase the fact-finder's credence regarding the defendant's guilt or innocence. But it doesn't seem likely, at least not in the typical case, to do so to a greater extent than is warranted by careful, rational Bayesian calculations, given certainty that the defendant was, indeed, arrested. Nor would the use of such evidence by the fact-finder incentivize bad government behavior; in fact, the reverse is true: admit information about the rate of guilt among the accused and you would encourage the government to improve their techniques so as to accuse a higher percentage of people who turn out to be guilty. And yet in our guts we know that something would be going

[4] Note that one can take this approach to justifying public policies without being consequentialist. One might hold, for instance, that the reason it is bad for the government to torture people is, simply, that torture is wrong, even if it results in good consequences. The problem, then, that is avoided by the inadmissibility of coerced confessions is that the government is not given incentive to do something wrongful, even if doing it has good consequences.

terribly wrong if, in the typical case, we were to convict even guilty people on the strength of nothing but the fact they were arrested together with evidence showing that the police get it right so frequently that any doubt about that would be well beyond reasonable. What would be going wrong? The primary goal of this paper is to answer this question.

What I am here calling "guilt rate evidence" is evidence about the rate of guilt in some group to which the defendant clearly belongs. In the example just given, the relevant group is all those who are arrested. But this paper also concerns evidence about the guilt rate in any other group to which the defendant belongs. Sometimes at least one good reason for the evidence's inadmissibility derives from the social significance of the group. Part of the reason why the jury should not be shown evidence that the vast majority of crimes are committed by men, for instance, is that such evidence exacerbates gender-driven inequality in government treatment. However, such a line of thought will not help to explain why the jury should not be shown evidence about the rate of guilt among other groups of which the defendant is a member that are not historically victims of oppression or granted unjustified entitlements. Our target here is the inadmissibility of evidence of the rate of guilt in *any* group, even if the group itself is of no independent social importance.

Some other reasons for the inadmissibility of particular types of guilt rate evidence similarly cannot be generalized. For instance, one reason for barring the admission of the rate of guilt among *the arrested*, in particular, is that to weigh such evidence is for the fact-finder to defer more than she should to the police's judgment about the defendant's guilt. We might think that fact-finders should reach their own independent judgment if they are to play their legitimate political role. They are to serve as a check on the police and so cannot use the fact that police frequently get it right in reaching their conclusions. This explanation, however, fails to explain

why the rate of guilt among bald people, if the defendant is bald (to choose a random group) should be inadmissible. In addition, it fails to explain why fact-finders often do just as they ought when they defer to experts in reaching their conclusions; forensic experts, whose judgment juries routinely weigh, are not "the people's" representatives in the sense in which the jury is supposed to be.

Other lines of thought turn out, on reflection, to be excessively general; they exclude much evidence that should be admissible. Consider, for instance, a line of thought that starts with the observation that guilt rate evidence is generally collected by first collecting evidence favoring the guilt of individuals and then collating it with evidence of those individuals' membership in a particular group. We see evidence that person is x is guilty, separate evidence that y is guilty, and only then see evidence that x and y have something in common and so are members of the same group. Guilt rate evidence, then, is what is sometimes called "higher order evidence": it is evidence to the effect that there is frequently evidence favoring the guilt of members of a particular group. Perhaps there is good reason to make higher order evidence of all sorts inadmissible. After all, there are some (albeit a minority) of epistemologists who do not think that higher order "evidence" is even evidence.[5] However, such a line of thought suggests that it should count *against* the admissibility of a piece of scientific evidence that it is widely accepted by the scientific community when, at least intuitively, this counts strongly for it. In fact, this intuition is captured by current admissibility law: acceptance by the scientific community is a necessary condition for the admission of a piece of scientific evidence.[6] But,

[5] [[refs to epistemology lit]]

[6] [[ref]]

acceptance by the scientific community is itself higher order evidence: it is evidence that there is evidence for it that experts recognize to be sufficient for acceptance. In general, to disallow all higher order evidence is to bar anyone with expertise from testifying about their professional conclusions without fully explaining their evidence-based reasoning to the fact-finder. But why should such testimony be barred? Often nothing has gone wrong when a reason a person has to believe something is that people who know what they are talking about also believe it.

Further, the rationale proposed here for the inadmissibility of guilt-rate evidence is very different from a rationale rooted in the valid observation that there is something unfair about directing government power against people on the basis of problematic qualities possessed by *other* people, who happen to share their traits. The thought is that to conclude that someone is guilty solely because many people like him (e.g. those who were also arrested or those who are also male) are guilty, is to judge him based on what is known about those others (e.g. a lot of them are guilty). That can seem unfair. We find the same problem, however, with this line of thought: it favors the inadmissibility of much that should be, and is, perfectly admissible. After all, every conclusion based on evidence involves information about other people besides the defendant. A fingerprint matching the defendant's is found at the scene. If we reach the conclusion that the defendant was there at some point, it is only because a very large percentage of people who, like the defendant, have fingers with prints on them, were once in the locations where similar prints were found. To conclude that someone has a quality (e.g. having been at the scene sometime in the past) that is only contingently, but regularly, found with another we know her to have (e.g. having a skin pattern on her finger that matches one found at the scene) is to reach conclusions about her on the basis of a fact about how those qualities are conjoined in

others. But nobody has been treated unfairly when verdicts are reached on the basis of fingerprint evidence.

The problem with guilt rate evidence, in short, is not that, were we to use it, we would be failing to treat the defendant "like an individual", whatever, exactly, that means. Still, as we will see by the end, the answer to be offered and defended here takes us some way, although not the whole way, towards solving an old, long-standing problem in legal theory that springs from the intuition that reaching conclusions about individuals on the basis of facts about their groups is problematic: what is sometimes called "the problem of individual and statistical evidence". The problem is illustrated by various famous hypotheticals. In Judith Thomson's, for instance, V is injured by a weaving car that drives off into the night. She could tell it was a cab from the light on top, but because of the darkness she could not tell what color it was. It turns out that the Red Cab company has 60% market-share; 60% of the cabs on the road are theirs, and the rest belong to Green Cab. If that's the only evidence introduced to establish that Red Cab owns the cab that injured V, is it permissible for the government to force Red Cab to pay for V's injuries? Of course not. But why not? The problem is made hard by the fact that the government would be allowed to force Red Cab to pay on the strength of, for instance, a tire mark found at the scene, given some lab analysis showing there to be a 60% chance that the mark was made both by one of Red Cab's fleet and by the car that struck V. The evidence in the two cases supports equal levels of credence in guilt (modeled as .6), appears to be non-prejudicial, and there is no government wrongdoing in either case, but it appears intuitively that the government is authorized to force Red Cab to pay for V's injuries in the tire track case but not in the market-share case. A natural way to re-state this intuition is by stating an intuition about admissibility:

intuitively, V's attorneys should be permitted to present the tire track evidence to the fact-finder, but not the market-share evidence.

In the Red Cab/Green Cab hypothetical, the statistical information that we should not admit into trial does not concern the distribution of guilt in the offender's group but, instead, the distribution of company ownership. The cab that hit V is one of the many that were on the street that night and their ownership is distributed unevenly between the Red and Green cab companies. As we will see, the story to be told here identifies a problem with any system that admits evidence about the distribution of *guilt*, in particular, in the group to which the defendant belongs; the very same problem does not attend a system that allows the introduction of information about market-share. But, still, as we will see, what is to be offered here is highly suggestive, pointing the way to a direction of inquiry about the statistical-individual evidence problem that may very well be fruitful.

What emerges from the answer offered here is that the inadmissibility of guilt rate evidence is like other forms of inadmissibility: were guilt rate evidence presented to and weighed by the fact-finder, something will have gone terribly wrong. In particular, it is suggested, what has gone wrong is that the fact-finder will have followed procedures that are non-optimal: following those procedures often results in fewer accurate verdicts than alternative procedures which do not require the use of guilt rate evidence, and would, in fact, be likely to be interfered with by the use of it. Or, put another way, if the fact-finder is given guilt rate evidence across the full range of typical cases, the fact-finder will reach fewer accurate verdicts across those cases than she would have had the guilt rate evidence been hidden from her. The rationale to be offered here for the inadmissibility of guilt rate evidence is, broadly-speaking, rule consequentialist. It is easy to generate particular, singular examples of uptypical cases in which

the fact-finder would have been more likely to reach a correct conclusion if offered guilt rate evidence. But still, the choice to permit the presentation of guilt rate evidence to fact-finders results in a less good track record of verdicts across typical cases than one that can be achieved by banning its presentation.

The explanation to be offered here of the inadmissibility of guilt rate evidence, such as the rate of guilt among the arrested, has several steps that are not obviously linked to the issue. The initial steps involve a short trip into a relatively obscure corner of theoretical biology: the corner concerned with the modeling of ideal foraging behavior by, for instance, bees that choose among flowers from which to harvest pollen, or predators that choose among hilly and forested areas for harvesting prey. As we will see, contributors to that literature have identified a distinctive strategy that will allow foragers to maximize the rewards they seek when faced with an environment not uncommon in the biological world: an environment of diminishing returns in which the amount of effort required to rake in a reward increases the more efforts have been successful. If, for instance, prey do not reproduce as quickly as they are eaten, then the more that are successfully hunted, the fewer there are, and so the greater the amount of work is required to capture them. What we learn from the modeling of ideal forager behavior is that a peculiar strategy will lead the forager to maximize rewards when facing such a situation.

As we will see, what follows from reflection on this peculiar strategy is that information about the proportion of targets that are rewarding to capture—for instance, information about the percentage of flowers in the field that actually contain pollen—is at best useless and can easily mislead the forager away from its best strategy for maximizing reward. Foragers often do better if such information is hidden from them.

Conceived of in a certain light, and this is the next step, we aim in our legal systems to "harvest" guilty people—our aim is to accurately identify them and single them out for distinctive treatment, such as punishment. So conceived, this suggests that our legal processes for singling out the guilty ought to mirror that of an ideal forager. In particular, information about the percentage of those who are eligible to be convicted who are actually guilty can mislead us; it can lead us away from maximizing our convictions of the guilty just as information about the proportion of flowers that contain pollen can mislead the bee and lead her to harvest less pollen than she would harvest if that information were hidden from her. There is reason, then, to prefer a system in which such information is not information on which we can draw in determining our verdicts. The best way to assure it is hidden from us is to make it inadmissible evidence. The criminologist's information about the degree to which arrest is a signal of guilt—which is, without a doubt, probative—is, nonetheless misleading: if we are armed with that information when reaching verdicts, then, across typical cases, we will not convict the guilty and acquit the innocent in as high numbers as we would if that information were hidden from us. At least, so it is argued here. What is supplied, that is, is a pragmatic justification of our bar on the admissibility of evidence of guilt rates: it is inadmissible because we improve our verdict accuracy in typical cases if it is hidden from our view. Given the emphasis in the justification of public policies on proper performance of a rule in typical cases, this justifies our adoption in law of the inadmissibility of guilt rate evidence.

2. Foraging for Guilty People

Although I will not return for some time to this example, it is important to see that there is a deep similarity between the problem faced by juries—convict or acquit?—and the problem faced by foragers like bees. The bee can land, let's say, on n flowers in a day—not one more. Let's imagine that each time it picks a flower it has a choice between a yellow and a red. It cannot tell until it has put effort into landing on a flower and poking its head through the petals how much pollen, if any, the flower has and which the bee can harvest. If the flower has pollen, the bee is rewarded—and the more it has the more the reward—while if it does not, the bee has wasted her effort there. Given imperfect information about her situation, what strategy should the bee adopt for allocating its n choices between red and yellow? What strategy should she follow so as to maximize the chance that she harvests the maximum amount of pollen in n tries? Should she always choose red flowers? Or alternate between yellow and red? Or should she stick with a color provided that her last choice of that color was rewarding? Or alternate for awhile and then stick with the color that has been most rewarding? Or what?

Similarly, there are n American trials. Like the bee who has a choice between red flowers and yellow, The People have a choice between two options in each trial: acquit or convict. The People are rewarded by convicting the guilty and they are rewarded in a different way by acquitting the innocent. This difference is mirrored in the bee's situation when the yellow and red flowers yield different rewards, as when yellow flowers deliver, on average, a different quantity of pollen than red flowers, or perhaps a different kind of pollen. Just as the bee elects a sequence of the form red-yellow-yellow-yellow-red-red etc.—a sequence of token efforts, allocated between two types—the jury elects a sequence of the form acquit-convict-convict-convict-acquit-acquit, etc. The bee's total reward is the sum of the sequence of quantities of pollen of each type found in the sequence of flowers on which she expends effort.

The jury's total reward is determined, also, by adding up the rewards achieved through each act of acquittal or conviction in the sequence. It is rewarded each time it convicts a guilty person or acquits an innocent person, and perhaps it is rewarded more by convicting a guilty murderer than a guilty thief, and perhaps it is less harmed by acquitting a guilty person than convicting an innocent. However exactly we are to measure the individual rewards of true verdicts and losses in false verdicts, they are to be summed up in tallying the jury's overall reward from its sequence of verdicts just as the bee's overall reward is tallied by summing up the rewards it reaps from each of the flowers in the sequence of visits.

The parallels do not end here. Say the bee is harvesting over a set amount of time: from sunrise to sunset, for instance. During that period of time, there is some finite amount of pollen distributed among the red flowers and some distinct finite quantity distributed among yellow. These may be all present at sunrise or they might all appear at sunset or they might appear in a pattern throughout the day; who knows. However they are distributed across the day, each time a bee harvests pollen from a red flower, the total amount available from red flowers diminishes and similarly for yellow flowers. What this implies is that effort of both types—both effort expended by landing on red flowers and effort expended by landing on yellow—has diminishing marginal returns: the more the bee has successfully harvested a unit of pollen from red flowers, the more red flowers need to be searched in order to harvest the next unit of pollen from red flowers; similarly for yellow. (Or, more carefully, that is true provided that the harvesting of pollen does not increase the number of pollen-holding flowers in the field; more about that possibility in a moment.) The jury faces a similar situation: over a given period of time—from January 1 to December 31, for instance—there are a finite number of people brought to trial. This pool is divided among the innocent and the guilty. Each time the jury convicts a guilty

person, there are fewer guilty people in the pool; and similarly for acquittal of the innocent. Therefore, each accurate guilty verdict slightly decreases the chances that the next guilty verdict will be accurate; similarly, each accurate acquittal slightly decreases the chances that the next acquittal will also be of an innocent person. (Or, more carefully, that is true provided that reaching verdicts does not increase the number of trials; more about that possibility in a moment.)

Notice, however, that there is an important difference between the bee's situation and the jury's. Bees have a strong reason to favor foraging strategies that support flower populations. In fact, they have adopted such strategies. As basic botany teaches, by foraging flowers, the bees contribute to flowers' genetic variation which, in turn, positively affects flower populations. There are more flowers, and so more pollen worth harvesting, thanks to the fact that bees harvest pollen. This is a happy result for bee populations. Further, if the bee could harvest more pollen using a strategy that would drive down flower populations, while it would be smart for the *particular* bee to adopt that strategy (assuming that it will not live long enough to encounter the flower population decline) it would be a mistake for bees *generally* to adopt that strategy. Bees as a species are better off using foraging behaviors that allow flowers to populate even if they could adopt alternative strategies that would be better for individual bees but would reduce the flower population over time.

In the case of juries, by contrast, there is no difference between the interests of one jury and the interests of the totality of all juries; the jury's interests our "our" interests, which are, by nature, collective. This is not the only difference between the bee's situation and the jury's. Trials are analogous to flowers: the more there are, the more that juries can potentially reap the rewards of accurate verdicts. But by contrast with bees, if juries had a way of distributing

verdicts that drove down the number of trials, that would not be, by itself, a reason to avoid such behavior; in fact, it might be a reason to engage in it. To know whether juries should behave in a way which drives down the number of trials, we would need to know why the behavior has that effect. This could happen for a variety of reasons, some of which are in our interests to discourage others to encourage. Perhaps jury behavior drives down trials because it discourages people from becoming prosecutors or because it makes it difficult for the police to arrest people or because it prompts legislatures to cut funding for trials; all of these are reasons for the jury to avoid such behavior. Or maybe, alternatively, jury behavior reduces the number of trials because it drives down crime; this is a strong reason to engage in such behavior. How jury behavior reduces the number of trials matters to the question of whether juries should adopt such behavior. But if there were a jury strategy that drives down trials by driving down crime, the jury should adopt it while the bee should not adopt a strategy that drives down flower populations by, for instance, driving down seed production. Similarly, while the bees have a reason to adopt a strategy that happens to increase flower populations, jury behavior that drives trials up by driving up crime should be avoided. Following the acquittal of the officers who beat Rodney King, there was an increase in crime as many people looted and vandalized in anger and in protest. But this was nothing to congratulate ourselves on. So, this is an important difference that gives pause to the suggestion that juries should adopt the strategies that bees should adopt. Until we know how bee behavior favors bee interests thanks to its impact on flower populations, we do not know whether it should be emulated by the jury.

However, a response is available to this line of thought. To see it, start with the observation that verdict accuracy is lexically prior to lowering crime. Say we had a choice between the jury following an approach that had perfect trial accuracy and no effect on the

frequency of crime, on the one hand, and, on the other, an approach that drove down crime to a certain degree but with the cost of false convictions of the innocent with some degree of regularity. The latter approach cannot be justifiably preferred over the former by appeal to the lowering of crime that it produces. Lowering crime is not a reason to convict the innocent. (Plausibly, this is because convicting the innocent in order to achieve any good result, including the lowering of crime, is to objectionably use the innocent as means.) To justify the approach that way is to overlook the lexical priority of verdict accuracy over lowering of crime. Both are good things, but the former silences the latter in the justificatory contest. We can only take lowering of a crime as reason in favor of one strategy over another if they have equivalent accuracy rates. What this lexical priority point demonstrates is not that jury behavior has no effect on crime; this is false and it would be naïve to suggest it. Rather, what it demonstrates is that in assessing public policy, such as a law of evidential admissibility, a finding that it contributes positively to verdict accuracy silences any question that can be raised about how following the policy, in typical cases, affects the number of trials. What this implies is that we can, in our context, safely set aside questions about the effect of admissibility rules on crime rates if we learn that they improve our verdict accuracy. If following an admissibility rule improves verdict accuracy, it should be law, regardless of how following it effects the quantity of trials or the quantitiy of crimes. Because this is the strategy that is undertaken here, the disanalogies with the bees' case need not concern us.

So, we should expect that general principles that ought to be followed by the bee, in order to maximize the quantity of pollen it harvests, ought to be followed, also, by the jury in order to maximize the quantity of true verdicts it reaches in trials. When this idea is taken seriously, as we will see, we learn that the jury ought not to be shown information about guilt rates among

groups to which the accused belongs (e.g. the arrested). The reason is the same as the reason that the bee should not be shown information about the distribution of pollen among red or yellow flowers. This information will, at the least, be useless to the bee—she will get no closer to maximizing her reward given that information than she would if she didn't have it—and the information might be genuinely misleading; she might actually adopt a strategy for distributing her efforts between red and yellow that is less good given this information. For the same reason, the jury should not be shown guilt-rate information. At the least, it will be useless—the jury will get no closer to maximizing true verdicts given that information than if they didn't have it. And it might be genuinely misleading: guilt rate information might lead the jury to adopt a strategy for allocating its efforts between conviction and acquittal that results in fewer accurate verdicts than an alternative strategy, that it might have been adopted, if the information was hidden from it. The result: guilt rate information should be inadmissible.

The argument here focusses on qualities possessed by the *sequence* of efforts—the sequence of flower colors or the sequence of verdicts—when that sequence has the highest possible probability of maximizing reward. If the bee alternates between red and yellow flowers, or juries alternate between guilt and innocence, those are simple qualities of the sequence. As we will see, if the bee maximizes its pollen harvest, the sequence of flowers it chooses will have a peculiar, technical quality—it involves what will be called "match"—that it will lack if the bee sees information about the expected quantity of pollen to be harvested in a single effort. The bee will therefore fail to maximize if it sees this information. This is true, even though it is very far from clear how having that information interferes with the bee's capacity to choose the most promising flower on each effort, which appears to be a sufficient strategy for maximizing the pollen harvest. Similarly, the jury will not end up with the best sequence of verdicts if it sees

guilt rate information. The sequence of verdicts will not "match", and every optimal sequence of verdicts—every sequence that has the highest probability of having the most accurate verdicts we can expect—has this quality. This is true, even though it is very far from being clear how having guilt rate information interferes with the jury's capacity to issue the verdict that appears most likely to be true, which would seem to be a sufficient strategy for generating the best sequence of verdicts. Exactly how this all works should become clear by the end.

3. The Matching Law Explained

To understand how harvesters, like bees, ought to solve their foraging problems, we need to take a long side trip to consider the Matching Law, proposed originally by Richard Herrnstein.[7] The Matching Law is descriptive: it specifies a regularity in the behavior of foragers that is claimed to be found, in fact, in their behavior. According to the Matching Law, over the long haul *the proportion of a creature's behavior that is of a particular type equals the proportion of her overall reward rate that is associated with actions of that type*. The overall reward rate is calculated by adding up the frequencies with which each relevant type of act yields rewards. So, for instance, imagine that one type of act has yielded rewards one out of every three times in which it was tokened; its reward rate is 1/3. And imagine that another has yielded

[7] What is in fact discussed here is what is now referred to as the Strict Matching Law, which contrasts with the Generalized Matching Law. The latter attempts to describe behavior by also modelling both bias and the organism's capacity to learn what the probability of reward is. For our purposes, it will suffice to confine ourselves to the Strict Matching Law. [[ref]]

two rewards for every token act, and so has a reward rate of 2/1. If these are the only two types of action that the agent can has tokened, the overall reward rate is 1/3 + 2 = 7/3. The first type of act, then, accounts for (1/3)/(7/3) = $1/7^{th}$ of the overall reward rate. The Matching Law predicts, then, that the creature tokens the first type of act $1/7^{th}$ of the time and the second $6/7^{th}$ of the time. When a sequence of efforts conforms to the Matching Law, we will say that the sequence has the quality of "match".

For instance, imagine that you are given a choice between drawing a piece of paper from bag #1 or from bag #2. In each bag are some ten dollar bills and many blank slips of paper that are shaped just like bills. You are given the choice between the two bags over and over again. Let's say that you draw 1000 times, with the draws divided between the two bags. Over these 1000 draws, let's imagine, you pulled a ten dollar bill from bag #1 once in every ten draws from that bag, and pulled a ten dollar bill from bag #2 nine times out of every ten draws from it. So your total reward rate is 1/10 + 9/10 = 1. The Matching Law predicts that you drew from bag #1 a tenth of the time (.1/1), and from bag #2 nine tenths of the time (.9/1). Or, in other words, given that you drew 1000 times total, the Matching Law predicts that you drew 100 times from bag #1 and 900 times from bag #2. Given the reward rates from these two actions, it follows that you reaped 10 ten dollar bills from bag #1 (one tenth of 100) and 810 (nine tenths of 900) from bag #2 for a total of $8200. If, indeed, the sequence of efforts—represented as a sequence of the form bag #1-bag #1-bag #2, etc—conforms to the Matching Law, it has the quality of match.

As these examples illustrate, if an organism's behavior does indeed conform to the Matching Law, that can seem inexplicable. After all, the only reason that the organism forages at all is to reap rewards, and behavior conforming to the Matching Law sometimes yields substantially lower rewards than a different pattern of behavior would have yielded. There's no

clear conceptual link between matching and maximizing. Consider the example in which one type of act yields a reward only every third time tokened, while the other yields two rewards every time it is tokened. If you assume that the reward rate remains the same no matter how the token efforts are distributed between the two types, the organism would do best by putting all of its efforts into the second type of act.[8] In general, why would you allocate any of your efforts to a type of action that is less rewarding than another? Had you drawn from bag #2 every time, in the above example, instead of nine tenths of the time, then if the rate of return remained stable, you would have made $9000 instead of $8200.

This can make it seem like a good thing that the Matching Law is, in itself, purely descriptive. It can seem like a good thing that it does not say that creatures *should* conform to it, only that they do, in fact. Support for the normative assertion that creatures should conform to the Matching Law would be provided by some conceptually visible link between matching and maximizing; but there isn't one. As a descriptive matter, however, the Matching Law has a lot of empirical support, although much of its support comes from studies of non-human vertebrates, like pigeons, in controlled laboratory conditions. In a famous experiment, for instance, caged pigeons have two levers that they can peck. A certain percentage of pecks to the one lever release food and a different percentage of pecks to the other lever do so. The experimenters control the rates at which food is delivered by the two levers and so control the total reward rate. For instance, if the first lever releases food half the time, and the second one a third of the time, the total reward rate is 1/2 + 1/3 = 5/6, meaning that the first lever is responsible for ((1/2)/(5/6)

[8] If the organism tokens the first type of act n times, it could have generated 6 times as many rewards in those n efforts by tokening the second type of act all of those times instead.

= 3/5 of the reward rate. The researchers then keep count of how the pigeon allocates its pecks—how many to the first lever, how many to the second. And what they find is that the two ratios match. In this example, 3/5th of the pigeon's pecks are of the first lever. Change the payoff rate of pecks to the two levers and the two ratios change, but in lock-step, preserving match.[9] Similar results are found in other animals, and in humans too.[10] Recently, it has been shown that the behavior of some bodies of neurons known to be involved in animal and human decision-making processes behave as predicted by the Matching Law.[11] Of course, often a pattern of neural behavior does not align, even imperfectly, with the patterns we find in the organism's macro behavior. But, still, neuroscientific results of this kind are suggestive. They give us some reason to expect that the behavior driven by those neural mechanisms will, also, conform to the Matching Law.

4. Maximizing by Matching

Conforming one's behavior to the Matching Law has been shown to lead to maximal rewards, but only under certain conditions. It does so, in particular, when one faces a diminishing returns situation: one in which the reward expected from tokening a particular type of act reduces in response to having successfully been rewarded by tokens of that type of act in the past. Continuing the example, imagine that bag #1 contains 1000 slips of paper 100 of which

[9] [[ref]]

[10] [[ref]]

[11] [[ref to Newsome monkey paper]]

are ten dollar bills. So, each time that you draw a bill from that bag, you reduce the probability that the next slip you draw will be a ten dollar bill; the more success you have in drawing from that pile, the less return you can expect from future efforts expended that way. In such conditions, you will maximize your returns by conforming to the Matching Law.

While it has been mathematically demonstrated that conforming to the Matching Law maximizes rewards where all the options for conduct have diminishing returns, this can seem puzzling. As a first step towards appreciating it intuitively, consider a simple example. As before, you have a choice between drawing slips of paper from bags #1 and #2. Some of these slips are blank, others are ten dollar bills. Imagine that blanks drawn are returned to the bags, while you keep bills that you draw. This is a diminishing returns situation: the ratio of bills to blanks in each bag decreases when you are rewarded by drawing bills, and since blanks are returned to the bag when drawn, the ratio does not change when your efforts are not rewarded. So the more bills you have drawn from a particular bag, the lower the expected value of a new draw. Now imagine that the two bags begin with exactly the same contents: they have the same total number of slips of paper and the same percentage of those are ten dollar bills.

Now, one way in which the probability of maximizing reward in a sequence of draws is the highest one can hope for—one way, that is, to give yourself the best chance of doing as well as can be done in a sequence of draws—is for each draw in the sequence to be made from the bag with the highest expected reward.[12] It is, in fact, conceptually visible how choosing the

[12] Here, as is standard, the "expected reward" from an act is the sum of the rewards involved in each possible outcome of the act multiplied, respectively, by the probability of achieving that

option with the highest expected reward will maximize your chances of reaping maximum rewards. A good strategy, then, is to calculate the expected reward of each option and put your effort into the one with the highest expected reward. It is instructive to consider the sequence of acts performed by an agent who follows this strategy in our example. Since initially the expected reward from the two bags is the same—after all, they have the same contents—the person aiming to follow this strategy might flip a coin between bags #1 and #2. Say the coin is heads and so he ends up drawing from bag #1. That draw is either a ten dollar bill or it's not. If it is, then the expected reward from drawing from bag #1 has reduced slightly—there is one fewer reward in that bag and just as many blanks slips of paper. So: if draw #1 yielded reward, the agent should shift to bag #2 for the second draw. If, by contrast, draw #1 did not yield a bill, then he returns the blank slip that he drew to that bag and the two bags are, again, identical. So, if draw #1 did not yield a bill, the agent can once again flip a coin to determine which bag he draws from in his second draw.

Assume for a moment that draw #1 yielded a reward. So draw #2 is from bag #2. If that draw yields a reward then, again, the bags are the same and so draw #3 can be either from bag #1 or bag #2; the agent should flip a coin again. If, however, draw #2 yields a blank, then the third draw should also be from bag #2: it has a slightly higher expected reward than drawing from bag #1 has.

It should be clear that the agent who knows all the facts—he knows that the two bags are, at first, identical and he knows whether he has been rewarded by each of his draws—will

---

outcome. If you win a dollar for a head and two dollars for a tail, flipping the fair coin has an expected reward of ½*$1 + ½*$2 = $1.50.

eventually end up dividing his draws evenly between bags #1 and #2. He is certain not to do so in a single draw—one bag or the other will be the one from which he draws. And he might not do so in two draws: it is possible that the coin directs him to bag #1 for draw #1, he draws a blank from that bag and so flips again, and the coin directs him to bag #1 again in draw #2. So both of his two draws might be from the same bag. But the more he draws, the closer the probability approaches 1 that he divides his time evenly between the two bags. Notice, and this is the crucial point, that this is exactly what is predicted by the Matching Law: Divide your time evenly between two identical bags and your reward rate, also, will be divided evenly between them—or, rather, the probability that both your time and your reward rate are divided evenly approaches 1 as the number of draws in the sequence increases. So, what this shows is that an agent who is doggedly and successfully pursuing maximization of rewards in this example by invariably choosing the bag with this highest expected reward also conforms his behavior to the Matching Law.[13]

---

[13] Notice that we can take significant steps towards illustrating this same point for cases in which the two bags are not initially identical, where drawing from one bag has a higher expected reward than drawing from the other. Imagine that bag #1 has the higher expected reward: a larger percentage of the slips in bag #1 are ten dollar bills than those in bag #2. The agent aiming at maximization, then, who pursues it by choosing the bag with the highest expected reward, draws from bag #1 continuously until he has reduced bag #1's expected reward so far that it is no longer greater than bag #2's. If, for instance, bag #1 has 13 slips in it 10 of which are ten dollar bills while bag #2 has 10 slips 7 of which are bills, then the agent draws from bag #1 until he has removed 3 ten dollar bills from it, making its contents identical to those in bag #2.

It is important to note a further complexity. In the example just given, the bags have an important feature in common: their returns diminish at the same rates. This need not be the case since they might have the same expected reward while differing in their contents. For instance,

---

He then flips a coin to determine which bag he will next draw from. That is, after an initial focus on bag #1, the situation comes to be identical to the situation in the previous example in which the two bags have the same contents. The result is that initially the agent will devote himself to drawing from bag #1, but he will then divide his efforts evenly between the two bags once they come to have the same expected reward. So, overall, he will expend more effort on bag #1, but he will also draw a higher number of his rewards from bag #1. This does not show, without doubt, that his behavior will necessarily conform to the Matching Law. To show that in full rigor, we would have to show that if you added his initial efforts on bag #1 to those that he put into that bag after it had come to have the same expected reward as bag #2, the numbers would work out: the more draws, the closer to 1 would be the probability that the share of his reward rate coming from bag #1 would be equal to the share of his effort devoted to that bag. But even in the absence of rigorously showing this—although it has been rigorously shown by others—one can see how intuitive doubts that maximizing behavior would conform to the Matching Law are unfounded. After all, the intuition was that maximizers put all their efforts into the option with the highest expected reward. But what has been shown is that in a diminishing returns situation, putting one's efforts towards the highest expected reward requires splitting one's efforts among one's options. And there is no reason to be skeptical that one would split them in exactly the way that the Matching Law predicts.

imagine that bag #1 contains 10 slips 7 of which are are ten dollar bills while bag #2 contains 100 slips 70 of which are ten dollar bills. So drawing from each has the same expected reward: the expected reward on a single draw from either bag is $7. But a successful draw from bag #1 reduces that bag's expected reward to $10 * 6/9 = $6.67, while such a draw from bag #2 reduces it to $10 * 69/99 = $6.97. Bag #1 has approximately ten times the rate of diminution in expected reward as bag #2. The result is that in order to maximize, the agent will need to spend much more time drawing from bag #2 than bag #1. If he draws one bill from bag #1(reducing the bills from 7 to 6), not until he draws ten bills from bag #2 (reducing the bills there from 70 to 60) will the expected rewards of each bag be equal again. But this is intuitively in line with the prediction of the Matching Law. After all, of the 77 ten dollar bills that are, in theory, within the agent's reach in this example, 70 of them are in bag #2. So it is therefore no surprise the he would spend more of his time drawing from that bag than from the other, and also that a larger percentage of his overall reward rate would come from there.

There is darkness around the link between matching and choosing the option with the highest expected reward. It is visible why choosing the option with the highest expected reward results in maximization, but it is obscure what it is about realizing match in one's sequence of efforts that does so. But, despite this darkness, we know that interfering with conformity with the Matching Law will interfere with maximizing reward.

5. Maximizing Given Imperfect Information

If you are facing a diminishing returns situation, you can repeatedly choose that which has the highest expected reward only if you know both (1) what your chances of reward are on

the next draw and also (2) by how much that chance would diminish if your next effort is successful. That is, to follow the strategy of pursuing maximal expected reward, consistently, you need to know both the chance that an effort will yield reward and also the rate of diminution in expected reward in response to expenditures of effort.

In the examples given so far, the agent chooses that which has the highest expected reward and ends up conforming to the Matching Law. However, given imperfect information, it can be impossible to confidently choose the highest expected reward, because you cannot calculate it, while entirely possible to conform to the Matching Law. Imagine that you have no idea how many slips of paper are in either bag #1 or bag #2, nor do you have any idea how many ten dollar bills there are in either bag. All you know is that each bag contains some slips, some of which are bills, but you have no idea how many of either kind are in either bag. You therefore lack sufficient information for calculating the expected reward from either bag and also lack enough information to calculate the rates at which the expected rewards will diminish in response to successful, rewarding acts. So you cannot determine what the expected reward will be from the next effort, and nor can you determine by how much your successes on the next effort would diminish your expected reward on the efforts after that one. You would therefore be powerless to follow the advice, "Choose that from which you expect the greatest reward!" But even in this informationally impoverished situation, you could have enough information to conform your behavior to the Matching Law, or successfully follow the advice "Choose that which moves you towards conformity with the Matching Law!"[14] To conform to the Matching

[14] How can you conform your conduct to the Matching Law even while lacking the ability to assess the expected reward of your options? Consider the example in which you have no idea

how many bills or how many slips are in either bag. Initially you have done 0 draws and have 0 rewards. You are therefore in trivial conformity with the Matching Law. When that's true, but there are still rewards to be reaped, you should choose arbitrarily from among your options. So, your first move is to flip a coin between bags #1 and #2. Say the coin directs you to choose from bag #1. Either you are rewarded or you are not. Say you are. So you know that you have received 100% of your reward rate from bag #1 and have also expended 100% of your efforts there. You therefore conform to the Matching Law, which is your goal, and so flip a coin again to determine where your next effort should be expended. Say that, again, the coin directs you to choose from bag #1, but this time you draw a blank. You are still in conformity with the Matching Law: 100% of your draws have been from bag #1 and 100% of your reward rate (100% of the reward rate of ½ comes from draws from bag #1) is also from there. So, again, you flip a coin. Imagine that it sends you to bag #2 this time and you draw a ten dollar bill. Now you are not in conformity with the Matching Law. Your overall reward rate is ½ + 1 = 1.5: you have drawn twice from bag #1 and been rewarded once and you have drawn once from bag #2 and have been rewarded once. .5/1.5 = 1/3 of the reward rate comes from bag #1, but 2/3 of your efforts have been expended there. Similarly, 1/1.5 = 2/3 of the reward rate came from bag #2 but you have drawn from there only one out of three times. So, to move one in the direction of conformity with the Matching Law, draw #4 should be from bag #2—whether it results in a reward or not, the choice of bag #2 brings you a tiny step closer to conformity again with the Matching Law. If you draw a blank from bag #2 on draw #4, then your reward rate from each bag will be the same (1/2) and your efforts will have also been divided evenly between the two; you will therefore be in conformity with the Matching Law and will need to flip a coin to

Law you need to know two things: (1) how your efforts have been allocated among your options, and (2) how often you have been rewarded by the efforts allocated to each option. Given this information, you can determine when your sequence of efforts so far matches or fails to match and why it fails when it does. If you are in a state of matching, you can flip a coin between your options. If you fail to be in a state of matching thanks to the fact that one of your options has occupied a higher percentage of your efforts than its portion of the reward rate, you have a reason to choose the other option. In fact, all of the possible grounds for failing to match dictate which option you should next choose. In short, you only need information about your own behavior. You do not need information about the probability of reward, or information about the distribution of rewards among your options, nor do you need information about how the expected reward of your options is affected by your efforts. The result is that you can lack

---

determine from which bag you choose for draw #5. If, instead, draw #4 from bag #2 is rewarding, then 50% (2/4) of your efforts will have been expended on bag #2, which is the source of 2/3 of the reward rate. (Your overall reward rate from the two bags is ½ + 2/2 = 1.5. The portion of that reward rate that comes from bag #2 is 1/1.5 = 2/3.) You still know all you need to know to deduce what you should do in order to bring you closer to conforming with the Matching Law: choose yet again from bag #2. And so on. Follow your nose and, given enough draws, your probability of conforming to the Matching Law will get closer and closer to 1. Given that there is alignment between maximizing behavior and conduct conforming to the Matching Law in a situation of diminishing returns, it follows that you have enough information to maximize your rewards even though you lack enough information to calculate the expected reward of your efforts.

information essential to choose the option with the highest expected reward and still have the information you need to match.

What all this shows is that there are two strategies that one can adopt to maximize rewards in diminishing returns situations, and the two have quite different informational burdens: either repeatedly choose the option with the highest expected reward, which requires that you are able to calculate that, or choose the option that moves you closer to conformity with the Matching Law, which requires that you keep track of some information about your own behavior so far in the task. Either way, your behavior will tend towards reward maximization.[15]

[15] More carefully, both strategies lead to reward maximization in the long term assuming that there are not reward-subtracting costs associated with their respective informational burdens. Imagine, for instance, that you could pay money to learn how many slips of paper, and how many bills are in each bag. Imagine, at the same time, that information stored in your memory is not free: you would have to pay money to know how many times you have chosen from each bag and how many of those choices yielded ten dollar bills. It would then follow that either strategy *could* be pursued—you can get all the information you need to either pursue expected reward or to pursue conformity with the Matching Law—but which approach maximizes your money would depend on the size of these respective fees. If it costs more to consult your memory about your track record of draws than it costs to learn how many bills are in each bag—if the costs required to get the information needed to follow the matching strategy are higher than the respective costs for following the expected reward strategy—then it might follow that only the expected reward strategy leads to overall maximization of monetary reward. Or, if the fees are reversed in their comparative size, then the reverse.

6. Estimating an Effort's Reward

There is a very important way in which the cases so far discussed differ from the situation faced by juries. Juries do not know, independently, whether their verdicts are accurate and so they do not know if they have been rewarded or disappointed by their efforts. It is not as though the lawyers in the case tell them, at the end, whether the defendant was really guilty. So, to get closer to modeling the jury's situation, consider the distinctive problem faced by a person aiming to maximize who cannot know, even after completing his act, whether or not he has reaped a reward from it. A blind person, for instance, might not know whether the slip of paper just pulled from a bag is a blank or a ten dollar bill. Imagine that despite this disability, the person does know, right before pulling the slip, what the probability is that the slip will be a ten dollar bill, whichever bag he draws from; imagine, for instance, that the blind person knows how many slips are in each bag and what percentage of them are bills. Given this information, the agent's inability to determine if he was rewarded by draw #1 does not interfere with his ability to pull from the bag, on the first draw, that offers the greatest expected reward: he knows exactly what the expected reward from each bag is. But he cannot know the expected reward on draw #2 from choosing again from bag #1. In fact, he cannot know this even if he knows the rate of diminution of return from bag #1. He might know that his expected rewards from bag #1 will reduce by 10¢ *if draw #1 is a bill rather than a blank*. But since he doesn't know what he drew

on draw #1, he is at a loss to calculate the expected reward from bag #1 for draw #2; it is either the same, or 10¢ less, and he doesn't know which.[16]

Notice, however, that in this example, the agent has a useful strategy for *estimating* the expected reward from bag #1 on draw #2: to calculate it, reduce the expected reward on draw #1 by the amount that it would reduce if the draw was rewarding *corrected by the probability that draw #1 was rewarding*. So, say, for instance, that the blind person knows that his chances of drawing a bill are 1/2. So, his expected reward from bag #1 is $5. And imagine he knows that if he draws a bill, then a second draw from the bag will reduce expected reward by 10¢ to $4.90. So one way for him to estimate the expected reward from draw #2 is to subtract 10¢ * ½ = 5¢ from $5 and so to treat draw #2 from bag #1 as having an expected reward of $4.95. To do this is to reduce the expected reward by the *expected* amount it will reduce, rather than by the actual amount. He either drew a bill on draw #1 or he did not, so there is a 50% chance that, as a matter of fact, draw #2 from bag #1 has an expected reward of $4.90 and a 50% chance that it has the same $5 expected reward that it had on the first draw. Split the difference and you determine what to do by treating the expected reward as $4.95. It will take longer for this strategy to work, but given enough draws by the blind person, these estimates of the expected reward of each draw will get arbitrarily close to their actual values.

What has just been described is a way of following the strategy of choosing the option with the greatest expected reward, even when one lacks some of the information which is

[16] It could be that, either way, the expected reward from drawing from bag #1 is higher than drawing from bag #2. But one cannot expect this forever: after enough draws, the chance that the gap between the expected rewards of the two options has disappeared will approach 1.

required for calculating the expected reward values. To make that calculation accurately, one needs all three of the following pieces of information: (1) the expected reward for the next effort of each option, (2) the rate of diminution: the impact on the expected reward of one's efforts being rewarded, and not being rewarded, and (3) the rewardingness or non-rewardingness (and to what degree) of each effort. The blind person in the example just discussed has (1) and (2), but not (3). What has been suggested is that he can nonetheless follow the strategy by estimating the quantity of reward reaped by each effort and so, in turn, to use that information to estimate the impact of each effort on the expected reward of the next effort. He estimates the value of that slip of paper in his hand by equating it with the expected reward of that slip. And estimates its effect on the expected reward of the next effort by imagining that it has the expected effect. As important as (3) is in diminishing returns situations, you can still pursue the strategy of choosing maximal expected reward without it.

Notice, however, that this all falls apart if the agent missing (3) also does not have (2). If he does not know the rate at which expected rewards diminish, he cannot even *estimate* (by imagining that he got exactly what he expected) how his expected rewards have changed in response to his choices.[17] However, and this is the crucial point, the person who has none of

[17] Imagine, for instance, that each bag contains only one piece of paper and it is a ten dollar bill, and while the blind agent is informed that he can expect $10 from the next draw, and so knows that all the slips in the bag are bills, he does not know how many there are. So, he does not know that the expected reward of his second draw will reduce to $0. For all he knows, the expected reward of another draw from that bag will be the same as the first draw: $10. He is therefore entirely incapable of following the strategy of choosing that which has the highest expected

these pieces of information—he has no idea what the expected reward of any option is, no idea by how much those expected rewards diminish in response to rewarded efforts, and is not in a position to ascertain with certainty, even, whether he has been rewarded by any given effort—can nonetheless conform his behavior to the Matching Law provided that he can *estimate* the degree to which he has been rewarded by a given effort. Imagine, for instance, that bills feel slightly different to the blind person than other slips of paper. But he's not perfect: his credence that a piece of paper that feels like a bill is a bill, let's imagine, is .7. This blind person can use touch, as imperfect as it is, for estimating his reward rate: he calculates the reward reaped by drawing a slip that feels like a bill to be $7. Say, for instance, that he has drawn 10 times from bag #1 and half the slips pulled felt like bills. So, he estimates the rewards drawn to be equal to $7 * 5 = $35. Since he has drawn ten times from bag #1, he estimates the reward rate reaped from drawing from there to be $35/10 = $3.50. Say using the same estimation method he estimates the reward rate associated with draws from bag #2 to be $7: all the slips he has drawn from there felt like bills. Since 1/3 of the overall estimated reward rate ($3.50/ ($3.50 + $7 = $10.50)) comes from drawing from bag #1, to conform to the Matching Law, 1/3 of his overall draws should come from there. He can engineer this, and so, given enough draws, can approach a probability of 1 that his conduct will conform to the Matching Law. Using this method for estimating the degree to which his efforts have been rewarding, the blind person can conform his conduct to the Matching Law, and so maximize his rewards. The point is that one can conform

---

reward; there is always the possibility that the expected reward of the bag has dropped below that of the other bag, as when each bag contains only one slip of paper, and also the possibility that it has not.

one's conduct to the Matching Law, and so maximize reward, even given highly impoverished information *if* one has a way of estimating the degree to which one's efforts have been rewarding.

7. Bad Estimation Methods

Not every tool for estimating the degree to which one's efforts have been rewarded can serve to help the agent to follow the advice of conforming his conduct to the Matching Law. It is attractive, for instance, to estimate the reward reaped through an act to be equal to the expected reward from that act. So, if you think that 70% of the slips in the bag are bills, you would estimate the reward from a single draw to be $7. You act as if you received what you expected. The problem, however, is that if you know the expected reward of a single effort, but cannot follow the strategy of selecting the option with the highest expected reward, it must be because you do not know the rate of diminution, or how quickly your efforts will reduce the expected reward of your options. But if you are in the dark about that, then your tool for estimating the degree to which an effort has been rewarding will get worse and worse the more efforts you expend. But the result will therefore be that you do not approach conformity to the Matching Law in your behavior while using this estimation method.

The problem here is that to move towards conformity with the Matching Law, you need to use the same method over and over for estimating the reward earned by each effort. You cannot estimate the first effort's reward using the expected reward and then give up on using it to estimate the second effort's reward, for instance. Were you to do that, then you could not use these estimates to productively estimate the reward rate. Your estimates of the reward harvested

by one type of act could not be meaningfully compared to your estimates of another, and so you would not know how they compared to the amount of effort expended one way rather than another. You would therefore be in the dark about what you need to do to move towards conformity with the Matching Law.

Like the agent who estimates the reward of a particular piece of paper by feel, the agent needs a way of estimating the size of a reward reaped through a particular effort the degree of accuracy of which does not reduce the more effort he has expended. The sighted person, for instance, knows whether he just drew a ten dollar bill by looking at the slip of paper in his hand. That method for determining if the draw was rewarding is in no way dependent on his past efforts: it will be just as reliable a way of figuring out if he was rewarded and by how much no matter how many past efforts he has invested. If the situation is one of diminishing returns, his past efforts will decrease the likelihood that the paper in his hand will look like a bill, but they do not reduce the degree to which estimations based on looking are reliable. The same point can be made in favor of the blind person's method of estimating the value of a slip of paper in his hand by touch.

We can distinguish different ways of estimating the degree to which an effort has been rewarded. We always use the degree of credence in light of the evidence that the act was rewarding (as when one's evidence is what the slip of paper looks like or feels like, or as when one's evidence is the percentage of slips of paper in the bag that are bills). But some bodies of evidence support a given level of credence only given information about the probability that a token of the type of act is rewarding, and others do not require that information. Estimating the quantity of the reward reaped by equating it with the expected reward of the act is of the former variety; estimating it based on how the slip of paper feels in one's hand is of the latter variety.

What has been shown here is that in situations of diminishing returns, an agent with impoverished information—so impoverished as to make it impossible for him to calculate the expected reward of his alternative actions in repeated efforts, and impossible for him to know with certainty the degree to which his efforts have been rewarded—can still, even in that case, conform his conduct to the Matching Law (or, rather, the probability that it will conform will approach 1 the more efforts he expends). To do so, he has to use evidence for estimating the quantity of reward that his efforts have reaped *the evidential import of which does not depend on the probability that his effort has been rewarding*. The problem with expected reward is that it falls into the problematic category; it is of the sort that the agent cannot use for making the needed estimate.

What this implies is that an agent with the kind of impoverished information that we have been discussing must have access to some evidence that his efforts are rewarding other than information about the distribution of rewards among the set of possible efforts. If the only way to estimate whether the slip he just pulled is a bill is to consult information about the proportion of slips in the bag that are bills, then the agent can only conform to the Matching Law if he could also just choose the bag with the highest expected reward. By contrast he could be completely in the dark about the contents of the bag and still conform to the Matching Law provided that he can assess—by look or by feel, for instance—the chances that a slip he pulls is a bill.

Say we had a choice about what information we can pass to the agent in our example who has his sight but has no idea how many bills or slips are in either bag. Imagine that we are in possession of the following two pieces of information and are wondering if we should pass them to him: (1) the percentage of slips in the bag that are bills, and (2) the percentage of slips that look like bills but are actually counterfeit. (1) by itself will allow the agent to effectively follow

the strategy of choosing the act with the highest expected reward on the first draw, but will not help him to follow that strategy for any subsequent draw. To help direct draw #2 towards the bag with the highest expected reward, he would need to know, in addition to (1), how the percentage is affected by successfully drawing bills from the bag. Further, (1) will not allow him to conform his conduct to the Matching Law: lacking information about how the percentage of bills is affected by successfully drawing bills from the bag, estimates of the expected reward reaped by his efforts will become hopelessly error-ridden after several draws. Since the same estimation method would need to be repeatedly employed to help him to bring his conduct in conformity with the Matching Law, (1) is not of any use. (2) is of no use at all for helping the agent to choose that which has the highest expected reward. Since he doesn't know how many slips are in the bag, he cannot use (2) to calculate expected reward. But (2) will allow him to conform to the Matching Law (assuming that he is not blind). The reason is that he can estimate the value of that which he draws from the bag like so: If it does not look like a bill, it is estimated to be worth $0; if it does look like a bill, it is estimated to be worth $10 * the probability that the bill is not counterfeit. Using this estimation method, the agent can estimate the reward rates associated with each bag repeatedly and so can determine whether the percentage of his draws from the bag is higher or lower than the reward rate. Given this, he can make smart choices about how to distribute his choices so as to conform to the Matching Law.

What has just been offered is an argument in favor of giving the agent (2), rather than (1). But, in fact, it's important to see that it might be positively harmful to give the agent (1). If the agent is likely to overlook the fact that he will know nothing, after even the first draw from a bag, what the probability is that future draws will yield rewards, then the agent is likely to think that the probability is the same, or close to the same, no matter how many rewards he has drawn

from that bag. He would therefore be very likely to be deluded about his chances of being rewarded by a draw. That is, given (1), the agent can calculate the expected reward of draw #1. But if he overlooks the fact that he knows nothing at all about the expected reward of draw #2—since he knows nothing about how the probability of reward is affected by prior efforts—he is likely to rely on some faulty judgment. He might, for instance, just assume that the chances of pulling a bill on draw #2 are the same as that for draw #1, even though they have reduced by who-knows how much—and so he would over-estimate both the overall reward rate and the reward rate due to drawing from this bag. Information about the probability of being rewarded by the first effort is prejudicial: it tells the agent nothing of use about what to do beyond the first effort, but is likely to be considered by the agent to contain useful information about that. Given this thought, there is not merely reason to pass along (2)—it makes it possible for the agent to conform her conduct to the Matching Law and so maximize reward—but also a reason to hide (1) from the agent: it encourages the agent to mistakenly think that she can tell where the greatest expected rewards lie and pursue them.

8. The Inadmissibility of Guilt Rate Evidence

What has just been offered is the driving engine of the justification for the inadmissibility of guilt rate evidence. To see this, let's walk through the argument just offered but applied to the foraging problem of interest: the problem of foraging for guilty people. It helps to think of the entity doing the foraging as a collective entity, what I will call "the people". The entity that decides on the verdict, in any given trial—the jury, or the judge in bench trials—acts on behalf of the people, or exerts the people's efforts. But it is the people, and not the jury, who are rewarded

by true verdicts. The people allocates its efforts between two ways of getting rewarded by accurate verdicts, namely conviction and acquittal. While the jury is deciding a single case in the face of evidence presented to it, policies such as admissibility laws are aimed at crafting jury behavior generally, across all cases, in a way that maximizes accurate verdicts.

Those who are deciding, in each case, how a particular effort will be expended—the actual jury members who are settling on a verdict for the case in front of them—know nothing of the track record of past verdicts. They, therefore, cannot direct their current efforts so as to either continue or break a pattern found in past verdicts; they have no idea what patterns there have been. So, a well-designed system will set rules (e.g. "Vote to convict only if convinced beyond a reasonable doubt", or "Don't present evidence acquired through government-inflicted torture") which will result in the jury's collective behavior serving to maximize accurate verdicts. What needs to be shown, then, is that the inadmissibility of guilt rate evidence plays a positive role: it increases the chances that the collective behavior of juries will maximize reward for the people.

For reasons explained in section 2, in crafting policy aimed at maximizing the rewards of true verdicts, we must assume that both conviction and acquittal have diminishing returns. Assuming that verdicts do not generate new trials by generating crime—which must be assumed if the justice system is to be worth crafting—there are a finite number of guilty people and a finite number of innocent people and true verdicts remove them from the potential pools from which reward can be gained from a true verdict. I will return to the question shortly of whether the people can maximize simply by instructing the jury to elect the verdict with the highest expected reward. But let's assume, for now, that they cannot but must, instead, aim to conform their conduct to the Matching Law.

Importantly, given that goal, nothing allows juries to assess whether they have succeeded in being rewarded by past verdicts; there is no way to tell how often they have gotten it right. In this regard, the jury is like the blind person who cannot know for certain whether the slip that he just drew is a bill or a blank. Similarly, juries cannot know for certain whether their verdicts are accurate. But, there is a strategy the jury can adopt: *estimate* the degree to which each verdict was rewarding and use that information to guide the sequence of verdicts towards conforming to the Matching Law. How should the jury go about making this estimate? It should estimate how rewarding each verdict was by equating it with the rational credence in guilt or innocence given the evidence. This works well, *provided that the evidence in question does not concern guilt rate*.

There are two serious problems with using degree of credence supported by guilt rate to estimate the reward the people earn from a verdict. First, it is of no use: it becomes less and less accurate the more it is used, and it must be used repeatedly if it is to help the people to conform its sequence of verdicts to the Matching Law. The reason it becomes less and less accurate the more it is used is that we have no idea how its accuracy changes in response to new verdicts. Say that what is known is that 80% of past arrestees have been guilty. But perhaps convictions encourage police to arrest innocent people, thus driving down the guilt rate of arrestees; perhaps, for instance, the rate was 90% before the most recent 10,000 verdicts but was driven down ten points by those verdicts. If so, then appealing to the 80% guilt rate is slightly less useful for estimating the rewardingness of the second verdict than the first. How much less useful will depend on how responsive police behavior is to verdicts, which is anyone's guess. There is no way to know how either past verdicts or future verdicts will affect the expected rate of rewarding true verdicts. This means that the jury cannot use guilt rate evidence repeatedly to estimate the

degree to which the verdicts it reaches have been accurate, as it must in order to use is to estimate the reward rates of conviction and acquittal, and which it must estimate to conform its conduct to the Matching Law.

In fact, and this is the second problem, guilt rate evidence is not just useless, it is also misleading: it is extremely tempting to estimate a given conviction, for instance, to be exactly as rewarding as the guilt rate. This is so despite the fact that that will be a problematic estimate for any future conviction, and will become more and more problematic as the number of verdicts mount. Our tendency, that is, is to overlook the facts about guilt rate that make it limited in its evidential value and to treat it as though it is not just useful, but a decisive estimate of the likelihood that a conviction would be of a guilty person.

Taken together, what these two problems show is that the people have no chance of conforming their conduct to the Matching Law if they estimate the quantity of reward reaped by jury verdicts through appeal to guilt rate evidence. But if the jury is presented with a piece of evidence, sometimes it will consider it. We will thus be at a loss to determine, from the verdict, a plausible range of rational credence supported by the body of evidence excluding the guilt rate evidence. And without an ability to estimate that, we will lack the information we need to estimate the degree to which our verdicts are rewarding. The result: in a system in which guilt rate evidence is inadmissible, the people meet a necessary condition for maximizing verdict accuracy.

Showing, however, that a rule allows the people to meet a necessary condition for maximizing verdict accuracy only serves to justify such a rule if there is reason to think that the people could hope to meet a sufficient condition. Only if there's a way for the people to bring its conduct in line with the Matching Law is there a reason to remove an obstacle to that goal. In

our examples so far of agents adopting the strategy of conforming to the Matching Law, their reasoning began by assessing whether their efforts so far were in disconformity. The agent would ask himself, for instance, whether the share of his reward rate that came from expending his efforts in one particular way (drawing from bag #1, for instance) mismatched with the share of his efforts expended that way. If there was mismatch, then the agent would need to expend his next effort on that of his alternatives which would get him closer to match. For the reasons just explained, the people have a tool needed for estimating the reward rate of past verdicts—the degree of credence rationally supported by the evidence that excludes guilt rate evidence. But that tool is clearly insufficient for meeting the informational burdens of conforming behavior with the Matching Law: the people do not need, only, to repeatedly estimate in the same way how rewarding a verdict is, but also needs to estimate what proportion of verdicts are convictions and acquittals. This is, of course, public information, but what reason is there to think that decisions about verdicts are in any way sensitive to it? What reason is there to think, for instance, that we have arranged things so as to make it more likely that the jury will acquit when the portion of the overall true verdict rate associated with acquittal is very high but the portion of overall verdicts that are acquittals is very low? It would be disingenuous to suggest that I do not have space to answer this question here, as though I have a lengthy answer to it elsewhere; I do not. Still, something of use can be said.

It is clear, first, that jury behavior is sensitive to defense and prosecutorial practices. And it is clear, second, that rates of conviction and acquittal influence such practices. When acquittal rates feel high, prosecutors change their behavior in ways that they hope will drive them down, for instance. It is possible, as a conceptual matter, that attorneys change their practices in ways that increase convictions (or increase acquittals) when convictions (or acquittals) are what the

people need in order to move their behavior towards conformity with the Matching Law. This is not what consciously drives such behavioral changes. But, then, conformity with the Matching Law is not what *consciously* drives bee foraging behavior, or pigeon pecking behavior, that has been shown to change so as to produce it. It is possible, that is, that collective behavior by legal actors moves the jury towards conformity with the Matching Law, and so towards the maximization of accurate verdicts, even though none of the legal actors involved consciously represent that as their goal. If there are such entrenched mechanisms in the collective behavior of legal actors, then there is a condition sufficient to make jury behavior likely to conform to the Matching Law: the right kind of behavior by legal actors. So the rationale for the inadmissibility of guilt rate evidence offered here depends on the thought that the collective of legal actors, whether they know it or not, are drawn towards behavior that moves the jury towards conformity with the Matching Law.

Return to the question of whether the people ought to maximize reward simply by moving the jury to choose the verdict with the highest expected reward, rather than by aiming to conform its behavior to the Matching Law. In earlier examples in which this was not possible and so aiming at conformity with the Matching Law was the agent's only plausible way to maximize—e.g. the example in which the blind man has no idea how many slips, or how many bills, are in either bag—the reason was that the agent had no way of determining the expected rewards of the options; his information was impoverished. But the jury does have a way of doing so. The jury is presented with evidence of guilt; that evidence supports some degree of rational credence in guilt, which aligns with the likelihood that a guilty verdict would be rewarding. So, the expected reward from a conviction (acquittal) is the level of credence in guilt (or innocence) * the magnitude of the reward from a true conviction (or true acquittal). The

result is that the jury does as well as it can hope to do either by choosing the highest expected reward or by trying to conform its behavior to the Matching Law; either strategy is available and would result in maximization.

However, what we know is that if guilt rate evidence were admissible, this would interfere with the people's ability to conform to the Matching Law. Since it is demonstrable that maximizing behavior does conform, what we know is that the admissibility of guilt rate evidence would prevent maximization. Since it is demonstrable, also, that the repeated choice of maximum expected reward also maximizes reward—there is a conceptual visible link between such an approach and maximization—the admissibility of guilt rate evidence must somehow interfere, also, with correctly following that strategy. Tell the jury what percentage of arrestees are guilty and in the long haul they will not choose verdicts with the highest expected probability of being accurate. What is unclear is why the admissibility of guilt rate evidence would have this effect. Its presentation would interfere with the jury's estimate of the probability of guilt. But how? What our discussion of the Matching Law, and the strategy of maximizing by conforming conduct to it, makes clear is what is bad about guilt rate evidence: it is useless for a task that requires repeated use, like calculating reward rate, and it is dangerous. So it must be both useless and dangerous to the project of choosing maximum expected reward. But what is clear is that it is useful for calculating the expected reward *of the next verdict*, and seems safely used for that purpose; it is probative. True, it gets less useful the more verdicts have been issued since the collection of the data on which it is based, but that's hardly a reason to ban its presentation to the jury. Much evidence that the jury should be allowed to see is more valuable if little has happened since it was collected; memories fade, for instance, but testimony about what is remembered is and should be admissible. Because we do not need to use it over and over in

order to choose the verdicts that are most likely to be true, and so have the highest expected reward, it does not appear to interfere with our efforts to do that.

The position we are in, then, as theorists, is peculiar. We know that guilt rate evidence does interfere with our effort to choose verdicts that are most likely to be accurate—because, after all, it interferes with another strategy for maximizing reward, namely aiming at conformity with the Matching Law—but we are at a loss to say how. This is an objection to the view offered here only if the following is true: we can justify guilt rate evidence's inadmissibility only by showing how its admissibility interferes with the strategy of choosing verdicts that are most likely to be true. But this is not so: we can justify guilt rate evidence's inadmissibility by showing that verdicts in a system in which guilt rate evidence is admissible are less often accurate, no matter what ordinarily effective strategy the people are following to maximize accuracy. The argument here establishes that and so serves to justify the inadmissibility of guilt rate evidence. The most intuitively appealing view of how the people reap this reward is this: the jury issues verdicts that align with its best guesses about guilt. But if we frame the justificatory question—namely what justifies making guilt rate evidence inadmissible?—by starting with this intuitively appealing view, we cannot answer it. To answer it, we need to see that what the maximally effective jury is actually doing, even if it does not know it, is issuing verdicts that conform to the Matching Law.

9. Conclusion

Return, finally, to the so called problem of individual and statistical evidence illustrated by Thomson's Red Cab/Green Cab hypothetical. What has been established here is that we

ought not allow juries to see a particular type of statistical evidence: statistical evidence regarding the distribution of guilt in the group of which the defendant is a member. But market share evidence, such as the proportion of cabs owned by the Red Cab company, is not quite of this sort. Such a statistic, instead, concerns the distribution of corporate relationships in the class of people to whom the offender belongs.

However, there is sufficient affinity between market share evidence and guilt rate evidence that it can seem as though the problem with the latter might be found, in a different way, in the former. The problem with guilt rate evidence, if the argument offered here succeeds, is the following: it provides one with a helpful estimate of the rewardingness of one's verdicts, and so allows one to conform one's conduct to the Matching Law, only if one knows how the probability of guilt changes in response to convictions and acquittals, something which no jury knows or could hope to know. Could the same problem attend market-share evidence?

Imagine that the response to a guilty verdict is reduction of the offending party's share in the market. Sat that if a cab driver is found to have struck a pedestrian, the response is to strip the driver of his license, or in some other way prevent him from returning to his job, thereby reducing the market-share of the cab company for which he works. So understood, market share is not independent from conviction: ceteris paribus, conviction reduces market share. It would therefore be a mistake to estimate the reward rate of conviction by repeatedly estimating the individual reward of verdicts about Red Cab by using the company's market share. It would be a mistake, that is, to estimate the reward of convicting Red Cab as .6, rather than 0 or 1, because Red Cab has 60% market share. Such an estimate for the value of conviction in the hypothetical would be fine, assuming that there is no change in market share between the date of the estimate and the date of the verdict, but it is only useful as an estimate of the rewardingness of verdicts

directed to Red Cab in other cases, and so only useful for calculating the reward rate of Red Cab convictions, if one has some idea of how market share is affected by conviction. Estimating the reward rate of convictions of Red Cab, which one must to conform to the Matching Law, requires estimating in the same way the reward reaped through individual convictions. But, unless one knows how market share is affected by conviction, repeated usage of this tool for estimating will lead one away from, rather than closer to, an accurate estimate of the reward rate.

Still, this is merely suggestive. Whether there is a problem here that would warrant making market share evidence inadmissible depends on several factors. It depends, at least, on the size of the impact of convictions on market share, the frequency of convictions that have such an effect, and the frequency of cases in which rational degree of credence in guilt is a function of market share. If it takes thousands of convictions of Red Cab to have any effect on its market share, then our ignorance about the matter is not a reason to make market-share evidence inadmissible. Similarly, if it would take thousands of convictions and we find that there are at most one per year. And there would similarly be no reason to exclude the evidence if one is virtually always in position to estimate the likelihood of guilt without its aid. Further, the uneasiness which we feel in the Red Cab/Green Cab hypothetical does not diminish when we are told, as is true in the vast majority of trials of businesses, that neither conviction nor acquittal will make any difference to market share. Even then it feels like something has gone wrong in admitting the market share evidence. This suggestive line of inquiry about the individual and statistical evidence problem is, then, at most merely suggestive.

While it is possible that the puzzle raised by hypotheticals like Red Cab/Green Cab is best resolved using consequentialist tools of the sort employed here, it is also possible that it is not. What has been shown in this paper is that we can learn something of importance about how

jury decisions ought to be influenced by studying how foragers like bees ought to go about harvesting. The low-level mechanisms that contribute to biological fitness of the sort for which evolution ought to select ought also to structure a corner of our evidence law. How widely this lesson applies, and even how far up the food-chain we must in general travel to find ideal behaviors that justify our legal practices, is anybody's guess. It is possible that bees would harvest more pollen if they knew what percentage of nutrients were gobbled by red flowers, while we would harvest fewer guilty people if informed of the percentage of fares gobbled by Red Cab. Future work, I hope, will determine if this is true.

Let's end with a premise-conclusion presentation of the derivation, offered here, of the inadmissibility of guilt rate evidence:

(1) If the people's allocation of verdicts will not have the highest probability of maximizing verdict accuracy if the jury is presented with a type of evidence, then evidence of that type should be inadmissible.

(2) If the criminal justice system is of use, then verdicts do not increase the number of trials.

(3) If verdicts do not increase the number of trials, then both conviction and acquittal have diminishing returns: the more true convictions there have been the lower the probability is of a conviction being true; similarly for acquittals.

(4) If all the options for expending effort have diminishing returns, then any sequence of efforts, distributed among the options, that has the highest possible probability of yielding maximal reward will conform to the Matching Law.

(5) One can successfully follow the strategy of conforming one's sequence of efforts to the Matching Law only if one can estimate how rewarding one's efforts have been without appeal to information about the probability that an effort of a particular type was rewarding.

(6) So, the people cannot conform to the Matching Law if they must estimate the truth of past verdicts by appeal to guilt rate evidence.

(7) The people must use evidence presented to the jury to estimate the truth of past verdicts, which is an estimate of how rewarding a verdict was.

(8) So, the people cannot conform to the Matching Law if the jury is presented with guilt rate evidence.

(9) So, the people's allocation of verdicts will not have the highest probability of maximizing verdict accuracy if the jury is presented with guilt rate evidence.

(10) So, if the criminal justice system is of use, guilt rate evidence should be inadmissible.

In constructing rules for governing our trial processes and other aspects of the criminal justice system, how else can legislators proceed except by starting with the assumption that the criminal justice system is of use? Perhaps this assumption is false; perhaps the criminal justice system produces more crime than it responds to. But those of us who hope this is not the case have a decisive reason to bar attorneys from presenting the jury with guilt rate evidence. To admit such evidence is to undermine our best strategy for maximizing verdict accuracy, given our impoverished information: our only chance is to pursue match.