Eyewitness Identification: Bayesian Information Gain, Base-Rate Effect–Equivalency Curves, and Reasonable Suspicion

Gary L. Wells, Yueran Yang, and Laura Smalarz
Iowa State University

We provide a novel Bayesian treatment of the eyewitness identification problem as it relates to various system variables, such as instruction effects, lineup presentation format, lineup-filler similarity, lineup administrator influence, and show-ups versus lineups. We describe why eyewitness identification is a natural Bayesian problem and how numerous important observations require careful consideration of base rates. Moreover, we argue that the base rate in eyewitness identification should be construed as a system variable (under the control of the justice system). We then use prior-by-posterior curves and information–gain curves to examine data obtained from a large number of published experiments. Next, we show how information–gain curves are moderated by system variables and by witness confidence and we note how information–gain curves reveal that lineups are consistently more proficient at incriminating the guilty than they are at exonerating the innocent. We then introduce a new type of analysis that we developed called base-rate effect–equivalency (BREE) curves. BREE curves display how much change in the base rate is required to match the impact of any given system variable. The results indicate that even relatively modest changes to the base rate can have more impact on the reliability of eyewitness identification evidence than do the traditional system variables that have received so much attention in the literature. We note how this Bayesian analysis of eyewitness identification has implications for the question of whether there ought to be a reasonable-suspicion criterion for placing a person into the jeopardy of an identification procedure.

Keywords: base rate, eyewitness confidence, eyewitness identification, lineup, system variables

The vagaries of eyewitness identification are well-known; the annals of criminal law are rife with instances of mistaken identification.
—The United States Supreme Court (U.S. v. Wade, 1967)

Writing for the majority in U.S. v. Wade (1967), U.S. Supreme Court Justice William Brennan’s observations about eyewitness identification preceded the use of scientific approaches to examining the reliability of eyewitness identification evidence. In the mid to late 1970s, however, psychological scientists were able to summarize emerging programmatic research experiments on eyewitness memory for events (Loftus, 1979) and on the ability of eyewitnesses to identify culprits from lineups (Wells, 1978). Research on eyewitness identification flourished throughout the 1980s with experiments showing that under many conditions eyewitness identification can be highly unreliable. But it was not until the advent of forensic DNA testing in the 1990s that significant numbers of mistaken identifications began to reveal themselves outside of the psychology lab (Wells et al., 1998). The use of forensic DNA testing in the United States to exonerate people who were convicted of serious crimes shows that 75% of these DNA-based exonerations were cases involving mistaken eyewitness identification (Innocence Project, 2014). Moreover, studies of actual police lineups in the United States and England show that an average of one in every three eyewitnesses who make an identification from a lineup picks a known-innocent filler (e.g., Behrman & Davey, 2001; Horry, Halford, Brewer, Milne, & Bull, 2014; Memon, Havard, Clifford, Gabbert, & Watt, 2011; Wells, Steblay, & Dysart, 2015).

A rich experimental literature by psychological scientists has developed on the accuracy of eyewitness identification from lineups. Research programs have focused on factors influencing acquisition during the witnessed event such as weapon focus (see Pickel, 2007), distance (e.g., Lampinen, Erickson, Moore, & Hittson, 2014), culprit and witness race (e.g., Brigham, Bennett, Meissner, & Mitchell, 2007), and stress (see Deffenbacher, Bornstein, Penrod, & McGorty, 2004). Other research programs have focused on the lineup itself, such as pre-lineup instructions (see Steblay, 2013), how fillers are selected for lineups (see Malpass, Tredoux, & McQuiston-Surrett, 2007), and suggestive behaviors of lineup administrators (e.g., Greathouse & Kovera, 2009). Still other research programs have focused on computational models of the witness decision process (e.g., Clark, 2003), confidence-accuracy relations in eyewitness identification (see Leippe & Eisenstadt, 2007), and how postidentification feedback influences eyewitness identification confidence (see Steblay, Wells, & Douglass, 2014). There are several broad reviews of the eyewitness identification literature (e.g., Lindsay, Ross, Read, & Toglia, 2007; Wells, Memon, & Penrod, 2006).

One of the types of variables that has played a strong role in the eyewitness identification literature is called system variables (Wells, 1978). System variables are those that affect the reliability of eyewitness identification over which the justice system has (or
can have) control. Pre-lineup instructions or the degree of similarity between lineup members, for example, are system variables whereas exposure duration during witnessing or whether the witness and culprit are the same race are not system variables. System variables have received considerable attention because they can help inform the justice system about ways to reduce mistaken identifications. The current article operates in the tradition of this system-variable approach.

Overview

This article articulates a Bayesian approach to the experimental eyewitness identification literature. Two previous Bayesian treatments of certain aspects of the eyewitness identification literature have appeared previously in the literature (Wells & Lindsay, 1980; Wells & Turtle, 1986). For the most part, however, Bayesian treatments have not been given much attention in the eyewitness identification literature on lineups beyond the observation that posterior probabilities of guilt depend not only on the accuracy of the eyewitness but also the base rate of the guilty party being in the lineup (e.g., Clark & Wells, 2008; Wells & Olson, 2002). Moreover, prior Bayesian treatments of the eyewitness identification issue have tended to characterize the base-rate variable as if there were some base-rate number (55%? 85%?) that applies to real-world lineups. According to this view, if we just knew the actual base-rate number that applies to real-world cases, then we could make more precise use of Bayes’s theorem to calculate posterior probabilities from our experiments. Numerous articles in the eyewitness identification literature have mused that “we don’t know what the actual base rate is in the real world” and “it would be useful to know that the base rate is in actual cases.” But, as we will note, the base rate is actually a system variable and we can expect it to vary from one department to another (and even between detectives in the same department) because the justice system does not have policies that attempt to regulate it. Our Bayesian treatment of eyewitness identification in this article is more formal and novel than prior treatments and its importance rests strongly on the fact that the base rate is a system variable.

We begin by describing how eyewitness identification is a “natural Bayesian problem” in the sense that the conditional probabilities of interest to the legal system naturally map into Bayesian formulations. Next, we use aggregated data from published lab studies to show how powerful the base-rate variable is in terms of the important outcomes resulting from eyewitness identification procedures. We note how the base-rate variable is something that the legal system can influence and, hence, is actually a system variable. We then use data from the eyewitness identification literature to display various Bayesian curves. Each type of curve contains different information about how the base rate affects conclusions about eyewitness identification evidence. Our approach uses not only standard prior-by-posterior curves, but also information–gain curves (Wells & Lindsay, 1980). We then introduce a new type of curve that we call the base-rate equivalency (BREE) curves. These BREE curves index how much change in the base rate is needed to match the impact of other types of variables. We apply the BREE curves to lab data published by various researchers on five prominent eyewitness identification system variables: lineup instructions, presentation format, lineup administrator influence, lineup filler similarity, and show-ups versus lineups. From these curves we conclude that even modest increases in the base rate would have a greater positive impact on the reliability of eyewitness identification evidence than any of the five system variables. Finally, we discuss some possible reasons why the legal system has ignored the powerful base-rate variable in eyewitness identification and how a reasonable-suspicion policy might raise the reliability of eyewitness identification evidence.

The Basic Eyewitness Identification Experiment and the Standard Data Pattern

In an eyewitness identification experiment, participants view a controlled event (e.g., staged crime) and are then randomly assigned to view a lineup that contains the culprit embedded among fillers or to view a lineup in which the culprit has been replaced with an innocent person (innocent suspect) embedded among those same fillers. Witnesses are given the opportunity to either identify someone or make no identification. The resulting data are commonly reported as probabilities, frequencies, or percentages like those shown in Table 1.

In a lineup, fillers are known-innocent lineup members who are put in the lineup to “fill it out” to make it fair to the suspect. Some eyewitness researchers call lineup fillers “foils” or “distractors” but we prefer the term fillers because that is the common name that is used by police. Notice in Table 1 (modeled after Wells & Lindsay, 1980) that the identification of a filler is not called a mistaken identification. For purposes of this article, the term mistaken identification is reserved for instances in which an eyewitness identifies the innocent suspect. Of course, at one level the identification of a filler is a mistaken identification. But in the forensic sense, the identification of a filler is very different from the identification of an innocent suspect. When an eyewitness identifies a filler in an actual case it is immediately known to be an error and no charges will be laid against that filler. Hence, throughout this article when we refer to a mistaken identification we are

<table>
<thead>
<tr>
<th>Culprit presence</th>
<th>Identification of suspect</th>
<th>Identification of filler</th>
<th>No identification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Culprit present (guilty)</td>
<td>Accurate identification 46.1%</td>
<td>Filler error Type I 21.2%</td>
<td>Incorrect rejection</td>
</tr>
<tr>
<td>Culprit absent (innocent)</td>
<td>Mistaken identification 13.4%</td>
<td>Filler error Type II 34.5%</td>
<td>Correct rejection 52.0%</td>
</tr>
</tbody>
</table>

**Note.** These data are from Clark et al.’s (2008) quantitative summary of 94 eyewitness identification studies.
pattern and it makes good sense as long as eyewitness identification. How witnesses’ views were and other memory-related factors. How study to the next depending on many factors such as how good the culprit was more probable than Type I filler identifications (filler culprit was more likely to be identified than the average filler in a probable than mistaken identifications; in 94 of the 94 studies the example, in 94 of the 94 studies accurate identifications were more to the suspect, the exact wording of prelineup instructions, and so inTable 1 varied considerably in a number of ways, such as how good of a view the eyewitnesses had, how similar the fillers were to the suspect, the exact wording of prelineup instructions, and so on. Nevertheless, these aggregate data capture the kind of patterns that emerge consistently in eyewitness identification data. For example, in 94 of the 94 studies accurate identifications were more probable than mistaken identifications; in 94 of the 94 studies the culprit was more likely to be identified than the average filler in a culprit-present lineup; in 82 of the 94 studies Type II filler identifications (identification of fillers when the suspect is not the culprit) were more probable than Type I filler identifications (filler identifications when the suspect is the culprit); and in 85 of the 94 studies correct rejections were more probable than incorrect rejections. Of course, the absolute levels of performance vary from one study to the next depending on many factors such as how good the witnesses’ views were and other memory-related factors. However, the pattern observed in Table 1 is the typical or expected pattern and it makes good sense as long as eyewitness identification performance is above chance.

Eyewitness Identification Is a Natural Bayesian Problem

The standard eyewitness identification experiment yields data that can be represented as six conditional probabilities. In Table 1, for example, the 46.1% is a .461 probability that the suspect was identified given that the suspect was the culprit. For the purpose of summarizing the results of an eyewitness identification experiment, displays of percentages or probabilities like those in in Table 1 are useful. Suppose, however, you are a police investigator, prosecutor, judge, or juror who wants to use information of the type displayed in Table 1 to estimate the probability that a suspect who is identified is in fact the culprit (i.e., suspect is guilty rather than innocent). The tendency might be to assume that the odds are .461 to .134 because those are the probabilities that a guilty suspect is identified and an innocent suspect is identified, respectively. That would mean that the probability that an identified suspect is guilty is .461/.775 or .665. In fact, however, the probability is .775 only if we can assume that the base rate for the culprit being in the lineup is 50/50. And there is no reason to make that assumption. Experiments on eyewitness identification (like all those that contributed to the data in Table 1) tend to use a 50/50 base rate as a matter of experimental design because eyewitness researchers are typically interested having stable estimates for both the culprit-present and culprit-absent conditions. But what if the base rate is 35/65 or 75/25? This requires a Bayesian approach to the problem.

Our Bayesian approach to the eyewitness identification problem emerges from articulating two contrasting conceptualizations about the purpose of a lineup. These contrasting conceptualizations of the purpose of a lineup have never been fully articulated in the literature, but they are important to understanding the value of our Bayesian approach to the lineup. At a recent talk by the lead author of the current article, the audience (composed mostly of lab-based psychological researchers) was asked “What is the purpose of conducting a lineup in an eyewitness identification case?” The dominant answer was “to test the reliability of the witness” or some close variant of that answer (e.g., “to test the witness’s memory”). In actual cases, however, the purpose of a lineup is not to test the reliability of the witness. Instead, the purpose of a lineup in the real world is to test the hypothesis that the suspect is the culprit. These two purposes are very different. To see how and why they are different, we have constructed a mental experiment. Suppose that you are a detective and you think that Saul might be the perpetrator. So, you construct a lineup and prepare to show the lineup to the eyewitness. However, moments before you show the lineup to the eyewitness, an omniscient God appears and goes on record to announce that the eyewitness’s memory in this case is 100% reliable. If the purpose of a lineup is to test the reliability of the eyewitness, then you can simply call off the lineup. There is no need to do the lineup because the question of whether the witness is reliable has already been answered. But, of course, you still want to do the lineup. Why? Because, testing the reliability of the witness was not the purpose of the lineup. Instead, the purpose of the lineup was to test the hypothesis that your suspect is the culprit.

Now, suppose that the same situation exists (you have a lineup that you are about to present to the eyewitness) and the omniscient God appears. This time, however, the God goes on record to say that your suspect, Saul, is in fact the culprit. Would you still need to do the lineup? If the purpose of the lineup was to test the reliability of the witness, you would still need to do the lineup. But you no longer need to do the lineup. Why? Because testing the reliability of the witness was not the purpose of the lineup. Instead, the purpose of the lineup was to test the hypothesis that the suspect is the culprit. [Of course, in this example you might need to do the lineup anyway so that you have something to present to the jury in case they are skeptical that you got a direct message from God.]

Obviously, we do not have the luxury of omniscient Gods in answering these questions, but this mental experiment is useful in drawing a sharp contrast between these two alternative conceptu-
alizations about the purpose of a lineup. As this mental experiment helps make clear, the purpose of a lineup in criminal investigations is not to test the reliability of the witness but instead is to test the hypothesis that the suspect is the culprit. Of course, the reliability of the witness is one component of the test that the suspect is the culprit. After all, a totally unreliable eyewitness makes for a very poor test of the hypothesis that the suspect is the culprit. In other words, witness reliability matters, but the purpose of the lineup and the reason for its use in actual criminal investigations is to shed light on the hypothesis that the suspect is the culprit.

The reason that eyewitness researchers might have a tendency to view the purpose of a lineup as a test of the eyewitness instead of a test of the hypothesis that the suspect is the culprit is because lab experiments on eyewitness identification are in fact designed to test eyewitnesses, not to test the hypothesis that the suspect is the culprit. In lab studies, researchers already know whether or not the suspect is the culprit because they randomly assign witnesses to view lineups in which the suspect is or is not the culprit. Therefore, data analyses from eyewitness identification experiments take the form of examining the probability that the witness will identify the suspect given that he is the culprit versus identify the suspect given that he is not the culprit. Indeed, this is the appropriate analysis for the types of questions that eyewitness researchers are interested in answering (e.g., Under what conditions is eyewitness identification accuracy improved/reduced? What factors affect the certainty-accuracy relation in eyewitness identification?). In the real world of criminal investigations, however, the critical question is the obverse of the conditional probability that is estimated in experiments. Specifically, the question in the real world is: What is the probability that the suspect is the culprit given that he was identified by the witness? That is the question that of interest to police, prosecutors, judges, and jurors.

The direction of a conditional probability is, of course, extremely important. Consider, for example, the huge difference in your estimate for these two questions: What is the probability that a person is a male given that the person is taller than six feet six inches? versus What is the probability that a person is taller than six feet six inches given that the person is a male? The former probability is extremely high (>99%) whereas the latter is extremely low (<1%). Likewise, the probability that a witness will identify the suspect given that he is the culprit is not the same as the probability that the suspect is the culprit given that the witness identified him. Experimental data, like those displayed in Table 1, tell us about the probability that the witness will identify the suspect given that the suspect is or is not the culprit. But how do we express the probability that the suspect is the culprit given the behavior of the witnesses? Fortunately, Bayes’s theorem specifies the exact relation between a conditional probability (e.g., probability of B given C) and its obverse (probability of C given B) by taking into account the base rate for the truth of the proposition (C). So, if an experimental result tells us the probability of a behavior (B) given that the suspect is the culprit (C) we can calculate the probability that the suspect is the culprit (C) given a behavior (B) as long as we know the base rate for the suspect being the culprit.

We do not need to actually manipulate the base rate in the experiment to calculate the impact of base-rate changes. Instead, we can mathematically simulate any base rate (or all possible base rates) regardless of the base rate used in the experiment. We can imagine someone questioning this claim on the argument that the base rate needs to be experimentally manipulated in order to control for its effect on eyewitnesses’ behaviors. But that would be an incorrect construal of the eyewitness identification experience. In a standard eyewitness identification experiment, the eyewitness views one lineup that either contains the culprit or contains an innocent substitute for the culprit. There is no opportunity for an eyewitness to experience a base rate. For other types of studies that use multiple trials, such as face memory studies in which there are multiple previously viewed (old) faces among foils (new faces), participants could very well be influenced by the base rate (what proportion are old vs. new) across trials. In multiple-trial studies, manipulations of the base rate could influence the decision criteria of witnesses as they experience the test stimuli and therefore need to be controlled for experimentally. But in a single-culprit/single-lineup eyewitness identification experiment, manipulations of the base rate cannot have such effects. Therefore, the question of how the base rate affects eyewitness identification performance can be answered through mathematical simulations.

What we will show in this article is that the base rate has a large impact on the trust that we can have in an identification and that we can quantitatively compare the impact of the base rate to the impact of other variables that have received a great deal of attention in the literature. We start with some simple Bayesian equations and point estimates for base-rate examples using data from Table 1. Then we introduce prior-by-posterior curves, which are fairly common displays in Bayesian probability, to show that the impact of the base rate is nonlinear. We then display information–gain curves, which are derived from prior-by-posterior curves. We use these information–gain curves to compare information gain for five different system variables, and note how the lineup is consistently more proficient at discriminating the guilty than it is at exonerating the innocent. Finally, we introduce a new type of Bayesian curve, which we call base-rate effect–equivalency (BREE) curves. These BREE curves tell us how much of an increase in the base rate is needed to match the impact of each of these five system variables (e.g., to make a show-up perform as well as does a lineup).

Simple Point Estimates From Bayes’s Theorem

We use the expression \( p(\text{IDS}\mid\text{SC}) \) to represent the probability of identification of the suspect (\( \text{IDS} \)) given that the suspect is the culprit (\( \text{SC} \)). In Table 1, \( p(\text{IDS}\mid\text{SC}) \) would be 0.461. Likewise, \( p(\text{IDS}\mid\text{SNC}) \) is the probability of identification of the suspect (\( \text{IDS} \)) given that the suspect is not the culprit (which would be 0.134 in Table 1). We use the expression \( p(\text{Filler}\mid\text{SC}) \) to represent the probability of identification of a filler given that the suspect is the culprit (.212 in Table 1) and \( p(\text{Filler}\mid\text{SNC}) \) to represent the probability of identification of a filler given that the suspect is not the culprit (.345 in Table 1). Finally, let \( p(\text{NIDS}\mid\text{SC}) \) represent the probability of no identification given that the suspect is the culprit (.327 in Table 1) and \( p(\text{NIDS}\mid\text{SNC}) \) represent the probability of no identification given that the suspect is not the culprit (.520 in Table 1).

How does one get from the experimental data, which are of the form \( p(\text{IDS}\mid\text{SC}) \), to the trier-of-fact’s probability of interest, namely \( p(\text{SC}\mid\text{IDS}) \)? A version of Bayes’s theorem tells us the answer:

\[
 p(\text{SC}\mid\text{IDS}) = \frac{p(\text{IDS}\mid\text{SC}) \times p(\text{SC})}{p(\text{IDS})} \]
Consider two base rates, one a 25% culprit-present base rate and the other a 75% culprit-present base rate. Using the data in Table 1 and applying Equation 1, the 25% base rate would yield a probability that the suspect is the culprit given an identification of the suspect:

\[
p(\text{SC} \mid \text{IDS}) = \frac{p(\text{IDS} \mid \text{SC})p(\text{SC})}{p(\text{IDS} \mid \text{SC})p(\text{SC}) + p(\text{IDS} \mid \text{SNC})p(\text{SNC})} \tag{1}
\]

1 Base rate and prior probability are interchangeable terms in the context of the lineup problem. Both refer to the prior or base rate probability that the suspect in the lineup is the actual culprit.

Notice that eyewitness identification experiments have some of the information needed to answer this question, but are missing \( p(\text{SC}) \) and \( p(\text{SNC}) \). But, because \( p(\text{SNC}) = 1 - p(\text{SC}) \), only one parameter is missing. That parameter, of course, is the base rate or prior probability that the suspect is the culprit.\(^1\)

In contrast, if the base rate for the suspect being the culprit is 75%, then the probability that the suspect is the culprit given a nonidentification is:

\[
p(\text{SC} \mid \text{NID}) = \frac{p(\text{NID} \mid \text{SC})p(\text{SC})}{p(\text{NID} \mid \text{SC})p(\text{SC}) + p(\text{NID} \mid \text{SNC})p(\text{SNC})} \tag{2}
\]

Using the data in Table 1 and applying Equation 2, the 25% base rate would yield a probability that the suspect is the culprit given a nonidentification:

\[
p(\text{SC} \mid \text{NID}) = \frac{.327(.25)}{.327(.25) + .520(.75)} = .173
\]

In contrast, if the base rate for the suspect being the culprit is 75%, then the probability that the suspect is the culprit given a nonidentification is:

\[
p(\text{SC} \mid \text{NID}) = \frac{.327(.25)}{.327(.25) + .520(.75)} = .654
\]

Notice that nonidentifications always reduce the probability that the suspect is the culprit. Interestingly, however, the powerful influence of the base rate is evident here in more ways than one. Notice, for example, that the probability that the suspect is the culprit is lower when the witness identifies the suspect under the 25% base rate (.534) than when the witness rejects the lineup (no identification) under the 75% base rate (.654). Indeed, the base rate is such a powerful variable that it can overwhelm the informational value of an eyewitness’s identification behavior. The relative impact of base rates and witnesses’ identification decisions will be explored more fully in the section in which we present information—gain curves.

Finally, we can also calculate the probability that the suspect is the culprit given a filler identification:

\[
p(\text{SC} \mid \text{FillerID}) = \frac{p(\text{FillerID} \mid \text{SC})p(\text{SC})}{p(\text{FillerID} \mid \text{SC})p(\text{SC}) + p(\text{FillerID} \mid \text{SNC})p(\text{SNC})} \tag{3}
\]

Using the data in Table 1 and applying Equation 3, the 25% base rate would yield a probability that the suspect is the culprit given a filler identification:

\[
p(\text{SC} \mid \text{FillerID}) = \frac{.212(.25)}{.212(.25) + .345(.75)} = .170
\]

In contrast, if the base rate for the suspect being the culprit is 75%, then the probability that the suspect is the culprit given a filler identification is:

\[
p(\text{SC} \mid \text{FillerID}) = \frac{.212(.75)}{.212(.75) + .345(.25)} = .648
\]

Notice that, like nonidentifications, filler identifications lower the probability that the suspect is the culprit.

**Base Rates Can Vary Widely and Are Under Control of the Justice System**

In the previous section, we used the base rates of .25 and .75 to make it clear that the base rate can have a huge impact on forensically relevant outcomes. When presented with this lineup base-rate problem, there is a tendency for people to ask “but . . . what is the base rate for culprit-present and culprit-absent lineups in the real world?” As noted more than 20 years ago, there is no single base-rate number for culprit-present lineups (Wells, 1993). In fact, speculating that there is some single, unknown base-rate number “out there” seriously misconstrues the problem.

There is no single culprit-present versus absent base rate for lineups conducted in the real world because the base rate will vary widely from one jurisdiction to another and even vary from one detective to another within a jurisdiction (Wells, 1993). The culprit-present versus absent base rate is influenced by the judgments and behaviors of individual crime investigators and these behaviors are not currently constrained by policies. More specifically, there are no necessary conditions that have to be met in order to place someone in an identification procedure (Wells, 2006). Hence, “grounds” for placing an individual in a lineup can be as low as a wild hunch in the practice of one detective or police
department or as high as near certainty for another detective or police department. The decision to conduct an identification procedure is not constrained by any laws in the U.S. Moreover, we have been unable to find any set of policies and procedures in a U.S. police department that provides detectives with any cautions whatsoever against conducting lineups when there is a low probability that the person of interest is the culprit. A national survey of U.S. law enforcement officers indicated that more than one third of crime investigators indicated that they needed either no evidence at all or a mere hunch that a person might be the culprit before they would conduct a lineup (Wise, Safer, & Maro, 2011).

Behrman and Richards (2005) examined records from 306 Northern California cases in which lineups were conducted and resulted in the witnesses making an identification of someone. Each of these cases was coded for how much evidence existed against the suspects prior to conducting the lineup. Behrman and Richards found that 30% of the lineups were conducted under conditions in which evidence against the suspect was considered “minimal.” Even more shocking is that more than 40% of the lineups were conducted under conditions in which there was no extrinsic evidence whatsoever against the suspect in the lineup. Hence, 70% of the lineups were conducted with either minimal or no evidence against the suspect. This is consistent with Wise et al.’s (2011) finding that more than one third of crime investigators reported needing no evidence at all before placing someone in a lineup.

Clearly, a police department or individual detective that routinely conducts lineups with little or no actual evidence against the suspect is going to run a lower base rate than will a department or detective who requires more evidence before conducting a lineup. As a result, that department or detective is going to run a much lower ratio of accurate to mistaken identifications than will a department or detective who requires significant evidence before putting a person of interest into the jeopardy of a lineup test. Although police departments appear to not be exercising systemic control over this base-rate variable, they could adopt policies and procedures to do so. Wells (2006) described one type of policy in which police departments could require their investigators to get approval from a chief of detectives before conducting a lineup. Of course, it is up to the legal system, not psychological researchers, to decide what might constitute reasonable suspicion to put someone in a lineup.

In many ways, the problem being addressed here is one that is well known in medical diagnostics. When a medical diagnostic test is performed on individuals for whom there is little reason to suspect have a particular disease, the rate of false positives can be quite high. This is why, for example, prostate screening is not recommended for men under the age of 30. Although the prostate test itself is just as accurate for men under 30 as it is for men over 50, the base rate for prostate cancer for men under the age of 30 is so low that almost all positive tests are actually false positives (see Vollmer, 2005 for a Bayesian analysis of PSA tests by age). Indeed, the seemingly constant shifts in medical policies and recommendations regarding when and to whom certain diagnostic tests and screenings should be given (e.g., breast cancer, pap smears) is based primarily on concerns that medical testing on low risk and symptom-free individuals yields a high false positive rate relative to the true positive rate owing to the fact that the base rate for the illness in question is very low.

Again, our point is that it makes no more sense to ask “what is the base rate for the culprit being present in lineup tests?” than it does to ask “what is the base rate for cancer being present in medical tests?” In both cases, it depends on the practices of the individuals who make the decisions on when and whether to conduct the test. In the medical world, there are medical boards, governing bodies, and insurance companies that discourage testing in cases where the base rate is low (low risk, asymptomatic) and these bodies establish guidelines that help minimize the risk of false positives that result from using such tests in low base-rate situations. But, this is not the case in the justice system. We will return to this issue of the legal system taking control of the base rate in the final Discussion section.

Prior-by-Posterior Displays

We turn now to a common type of Bayesian display, namely the prior-by-posterior curve. Above the diagonal line in Figure 1 is a prior-by-posterior curve (the solid-line curve) for the identification-of-suspect data from Table 1. The prior (across the X axis) is the base rate and the posterior (on the Y axis) is the posterior (postlineup) probability that the suspect is the culprit given that the witness identified the suspect or $p(\text{SC|IDS})$. The straight diagonal line is an “identity” line that represents how the prior would relate to the posterior if the identification had no diagnosticity. In other words, if the eyewitnesses were just as likely to identify the innocent suspect in the culprit-absent lineup as they were to identify the guilty suspect in the culprit-present lineup, then the prior and the posterior probabilities would be identical and their relation would be defined by the straight diagonal line. As the behavior of the witnesses becomes more diagnostic, the prior-by-posterior curve bends away from the identity line.

![Figure 1. Prior (base rate) by posterior probability of guilt curves for identifications of suspects, fillers, and nonidentifications. Based on data from Clark et al. (2008) meta-analysis of 94 eyewitness identification studies.](image-url)
Earlier in this article we calculated two of the points on the Figure 1 curve using Equation 1. We calculated that when the base rate is 25% the value of $p(\text{SC_1|IDS}) = .534$ and when the base rate is 75% the value of $p(\text{SC_2|IDS}) = .912$. For purposes of reference, we have marked those two points on the solid-line curve in Figure 1 as Point A and Point B. Readers will note that Point A on the curve corresponds to the 25% base rate and a posterior probability of .534 and Point B corresponds to the 75% base rate and a posterior probability of .912 (which we calculated earlier). Figure 1 is simply a curve of all $p(\text{SC|IDS})$ points across all possible base rates from zero to 100%.

Below the identity line in Figure 1 is the prior-by-posterior curve for nonidentifications (dashed line) and the prior-by-posterior curve for filler identifications (triangle-marked line). These curves for nonidentifications and for filler identifications bend below the identity line because they are diagnostic of innocence rather than guilt and therefore reduce the posterior probability to a value that is lower than the prior probability.

Non-Identifications Versus Filler Identifications: A Striking Finding

Examination of the curves for the filler identifications and for the nonidentifications in Figure 1 reveals a striking finding that merits a closer look. Specifically, notice that the prior-by-posterior curves for filler identifications and for nonidentifications are nearly identical. Indeed, the diagnosticity ratios that are driving the two curves are 1.59 for nonidentifications and 1.63 for filler identifications. An examination of the 94 studies in Clark’s meta-analysis shows that 45 studies produced diagnosticity ratios that were higher for the nonidentifications and 49 studies produced diagnosticity ratios that were higher for identifications of fillers.

Many researchers are likely to be surprised by this finding. Although there have been analyses of some individual eyewitness identification experiments that have shown diagnosticity for filler identifications that meets or exceeds the diagnosticity of nonidentifications (e.g., Clark & Wells, 2008; Wells & Lindsay, 1980; Wells & Olson, 2002), the presumption seemed to be that this would not usually be the case. Instead, the presumption has been that nonidentifications would typically be more diagnostic of the innocence of the suspect than are filler identifications. After all, nonidentifications tend to be construed in the eyewitness identification literature as “rejections” in which the eyewitness is indicating that the culprit is not present in the lineup. In contrast to the nonidentifiers, an eyewitness who selects a filler is clearly mistaken (by definition) and there is a tendency to think that such a witness is responding in a way that is not meaningful in terms of what it tells us about whether the suspect is the perpetrator.

The fact that we found the impact of filler identifications to be effectively equal to that of nonidentifications on posterior probabilities of guilt across a large number of studies leads us to take more seriously the possibility that one of our standard presumptions in the eyewitness literature (that nonidentifications are better evidence of innocence than are filler identifications) might be incorrect. Unfortunately, many meta-analyses (and the original articles on which they were based) fail to properly report filler identification rates for both culprit-present and for culprit-absent lineups. Hence, as we attempted to expand the database beyond the 94 comparisons that were recoverable from Clark et al.’s (2008) meta-analysis, we were frustrated. We were, however, able to use a recent meta-analysis of 72 studies that examined simultaneous versus sequential lineups (Steblay, Dysta, & Wells, 2011) because tests of sequential versus simultaneous lineups tend to routinely report filler identification rates. The Steblay et al. meta-analysis shares some studies with the Clark et al. meta-analysis, but they are not totally redundant. Like the 94 studies in Clark et al., the Steblay et al. studies show the diagnostic value of filler identifications to be nearly equivalent and even slightly greater than nonidentifications (diagnosticity ratios of 1.94 and 1.78, respectively).

Despite the near equivalence on average of diagnosticity for filler identifications and nonidentifications across these studies, we noticed a considerable amount of heterogeneity. In some studies, for example, filler identifications were up to seven times more diagnostic than nonidentifications and in other studies nonidentifications were up to nine times more diagnostic than filler identifications. Because the 94 Clark et al. (2008) and the 72 Steblay et al. (2011) studies vary in numerous ways, we were not able to discern a clear pattern for when filler identifications versus nonidentifications are more diagnostic.

Information–Gain Curves

Prior-by-posterior curves are a common way of examining Bayesian probabilities. However, there is another type of curve, called an information–gain curve, that is even more interesting because it allows for more direct comparisons between witnesses’ behaviors that are incriminating versus those that are exonerating. Information–gain curves, first introduced by Wells and Lindsay (1980), were developed specifically for eyewitness identifications from lineups, but now have been adopted for other purposes as well, such as polygraphs (e.g., Honts & Schweinle, 2009). Information–gain curves display the absolute difference between the prior probability that the suspect is the culprit and the posterior probability that the suspect is the culprit as a function of the behavior of the witness across all possible prior probabilities. For example, information gain for an identification of the suspect (behavior of the witness across all possible prior probabilities. For example, information gain for an identification of the suspect is incriminating, which is consistent with the higher base rate but goes against being less diagnostic than filler identifications and in other studies nonidentifications were up to nine times more diagnostic than filler identifications. Because the 94 studies in Clark et al. (2008) and the 72 Steblay et al. (2011) studies vary in numerous ways, we were not able to discern a clear pattern for when filler identifications versus nonidentifications are more diagnostic.

Information–Gain Curves

Prior-by-posterior curves are a common way of examining Bayesian probabilities. However, there is another type of curve, called an information–gain curve, that is even more interesting because it allows for more direct comparisons between witnesses’ behaviors that are incriminating versus those that are exonerating. Information–gain curves, first introduced by Wells and Lindsay (1980), were developed specifically for eyewitness identifications from lineups, but now have been adopted for other purposes as well, such as polygraphs (e.g., Honts & Schweinle, 2009). Information–gain curves display the absolute difference between the prior probability that the suspect is the culprit and the posterior probability that the suspect is the culprit as a function of the behavior of the witness across all possible prior probabilities. For example, information gain for an identification of the suspect is incriminating, which is consistent with the higher base rate but goes against being less diagnostic than filler identifications and in other studies nonidentifications were up to nine times more diagnostic than filler identifications. Because the 94 studies in Clark et al. (2008) and the 72 Steblay et al. (2011) studies vary in numerous ways, we were not able to discern a clear pattern for when filler identifications versus nonidentifications are more diagnostic.

Information–Gain Curves

Prior-by-posterior curves are a common way of examining Bayesian probabilities. However, there is another type of curve, called an information–gain curve, that is even more interesting because it allows for more direct comparisons between witnesses’ behaviors that are incriminating versus those that are exonerating. Information–gain curves, first introduced by Wells and Lindsay (1980), were developed specifically for eyewitness identifications from lineups, but now have been adopted for other purposes as well, such as polygraphs (e.g., Honts & Schweinle, 2009). Information–gain curves display the absolute difference between the prior probability that the suspect is the culprit and the posterior probability that the suspect is the culprit as a function of the behavior of the witness across all possible prior probabilities. For example, information gain for an identification of the suspect is incriminating, which is consistent with the higher base rate but goes against being less diagnostic than filler identifications and in other studies nonidentifications were up to nine times more diagnostic than filler identifications. Because the 94 studies in Clark et al. (2008) and the 72 Steblay et al. (2011) studies vary in numerous ways, we were not able to discern a clear pattern for when filler identifications versus nonidentifications are more diagnostic.
is shown by the dashed-line curve in Figure 2. As with the filler identifications, presenting the interesting possibility that a nonidentification or a nonidentification of the suspect is diagnostic of the suspect's innocence (exonerating direction). In this case, the new information (that the witness made no identification) is consistent with a lower base rate and goes against the higher base rate. Information gain for nonidentifications peaks when the base rate is 60%.

Readers will not be surprised at this point to note in Figure 2 that the information–gain curve for filler identifications is nearly identical to the information–gain curve for nonidentifications. This is, of course, the same phenomenon that we observed using the prior-by-posterior displays.

Why Are Filler Identifications Diagnostic of the Innocence of the Suspect?

One of the simple ways to think about the informational value of witness behaviors that work in the exonerating direction is to note that such behaviors are more likely to occur when the suspect is innocent than when the suspect is the culprit. Hence, it is readily intuitive that eyewitnesses should be less likely to make a nonidentification decision (identify no one) when the suspect is in the lineup than when the suspect is not in the lineup. This is simply another way of saying that correct rejections are more frequent than false rejections. Indeed, it is expected that correct rejection rates will be greater than false rejection rates when eyewitnesses are performing above chance.

But, why should witnesses be more likely to identify fillers when the suspect is innocent than when the suspect is guilty? There is general agreement that when witnesses make an identification they tend to identify the person who looks the most like their memory of the culprit compared to the other lineup members (Wells, 1984). In fact, the idea that witnesses choose the person who best matches their memory (when they choose someone) is a fundamental assumption that it is one of the elementary features of computational models of eyewitness identification (Clark, 2003). It would appear nonsensical (in the absence of outside influence) to suggest that an eyewitness would select someone other than the person who best matches their memory if they are going to select anyone at all. Hence, witnesses who identify a filler are saying that the filler looks more like their memory of the culprit than does the suspect, which, of course, is more likely to happen when the suspect is innocent (culprit-absent lineup) than when the suspect is guilty (culprit-present lineup).

Perhaps one of the reasons why it is difficult to grasp the idea that filler identifications are as diagnostic of innocence as nonidentifications is the fact that eyewitnesses who make filler identifications are clearly mistaken regardless of whether or not the suspect is the culprit. As a result, the tendency seems to be to dismiss filler identifications as being uninformative regarding the guilt or innocence of the suspect. In fact, there is evidence indicating that lineup administrators are so dismissive of filler identifications that they do not even bother to make a record of the fact that the eyewitness identified a filler. In analyses of police files attempting to score the results of photo-lineups administered in actual cases, researchers have found that lineup administrators failed to make a record of filler identifications whereas they consistently made records of suspect identifications (e.g., Behrman & Davey, 2001; Tollesstrup, Turtle, & Yuille, 1994). Consistent with this, controlled experiments show that research participants serving as lineup administrators reported all identifications (both suspect and filler) when they did not know whether the witness picked the suspect or a filler (double-blind administrators) but...
blood types between a suspect’s blood and trace evidence blood. For example, finding that a match of gloves, wiping of surfaces). Asymmetries can operate in the absence of the culprit’s prints at the crime scene (e.g., use their incriminating versus exonerating values. For example, finding a match between a suspect’s fingerprints and latent fingerprints (recovered from a crime scene) is more incriminating than finding a mismatch of these blood types is exonerating.

Why does the incrimination/exoneration asymmetry occur with eyewitness identification? The answer is probably different for filler identifications than it is for nonidentifications. In the case of filler identifications, the exonerating power is probably limited by the fact that it relies exclusively on witnesses who are mistaken and, hence, probably have fairly weak memories. After all, they identified a filler when they could have selected no one.

But why should the exonerating value of nonidentifications be considerably less than the incriminating value of identifications of the suspect? As far as we can tell, that question has never been asked in the eyewitness identification literature. We can only speculate, but we see two related factors. First, nonidentifications are not necessarily rejections. A rejection is when the witness claims that the culprit is not in the lineup. But, what about those witnesses who cannot make a positive identification of anyone but, perhaps knowing they have a weak memory, also cannot clearly reject the idea that the culprit is among those in the lineup? Presumably, many of those witnesses will make no identification at all. Hence, they are nonidentifiers, but they are not really claiming that the culprit is not in the lineup, but instead are treating the nonidentification decision as a type of “don’t know” response. One would not expect such responses to be diagnostic of whether the culprit is in the lineup or not.

Among the nonidentifiers, it might be only a small proportion that are actually making a rejection decision. In fact, it is a long-standing contention in the lineup literature that people have a great deal of trouble recognizing the absence of the culprit in a lineup (Wells, 1984). Many researchers have speculated that for recognition tasks, rejections require an editing system called recall to reject (e.g., Rotello & Heit, 2000) or what others have called rejection recollection (e.g., Brainerd, Reyna, Wright, & Mojardin, 2003; Lampinen, Odegard, & Neuschatz, 2004). The idea is that the person has to be able to recall the original (correct) item in order to reject a similar-but-new item. A positive response, in contrast, does not require recall, it only requires recognition. Hence, we speculate that the relatively lower exonerating value of nonidentifications relative to identifications of the suspect is due at least in part to the nonidentifiers being a heterogeneous population; some are rejecters, who could be expected to be diagnostic, whereas many others are making no identification because they simply don’t have a good enough memory to make an identification.

Information Gain as a Function of System Variables

We examined information–gain curves as a function of five different system variables. System variables have been a central focus of eyewitness identification research because of their potential to inform the justice system about methods of conducting lineups that produce the best results. There has been considerable debate about how to best compare the different levels of system variables. An increasingly dominant view in recent years has been to use Receiver Operating Characteristic (ROC) curves (e.g., Carlson & Carlson, 2014; Gronlund, Wixted, & Mickes, 2014; Mickes, Flowe, & Wixted, 2012; Wixted & Mickes, 2014). It is not the purpose of this article to critique the ROC curve approach to comparing eyewitness identification procedures. But the idea be-
hind the ROC approach is to examine differences in psychological discriminability independently of response bias. Response bias, however, is an important variable in practice, and many of the system variables that influence the reliability of eyewitness identification evidence might do so by impacting response bias (i.e., eyewitnesses’ propensities to choose). Whereas ROC analysis tries to separate discriminability from response bias, the Bayesian approach takes the data as they are and estimates the outcomes of interest to prosecutors, judges, and juries. For these justice system actors, the question is “what is the probability that the suspect is guilty given this particular behavior by the eyewitness (e.g., identified the suspect, identified a filler, or made no identification)?”

Our Bayesian treatment of these system variables is critical to the overall thrust of this article, which focuses on comparing the impact of these system variables to the impact of the base-rate variable using BREE curves. There are various researchers who have aggregated laboratory data on system variables. We used Clark’s (2012) meta-analytic data, which included accurate and false identification rates for biased versus unbiased lineup instructions, simultaneous versus sequential lineups, higher versus lower similarity between fillers and the suspect, more or less influence from a lineup administrator, and show-ups versus lineups across numerous published laboratory experiments. We created information gain curves for each of these system variables, which are a critical step in our development of BREE curves.

Clark (2012) did not break down the system-variable data into filler identifications versus nonidentifications. That is unfortunate because, as we noted earlier, sometimes filler identifications produce greater information gain than do nonidentifications and sometimes the reverse is true. If we had been able to separate filler identification from nonidentifications, we could have examined whether the relative information gain from filler identifications versus nonidentifications varied across different levels of the system variables. Future studies might examine that question. For current purposes, however, we calculated information–gain curves for a combined category (fillers and nonidentifications) that we call “no identification of suspect” (abbreviated as No IDS). Readers should not confuse nonidentifications (eyewitness makes no identification from the lineup, which we abbreviated earlier as NID) with “no identification of suspect” (witness makes either no identification or identifies a filler, which we abbreviated No IDS).

### Higher Versus Lower Similarity Fillers

There is broad agreement that physical similarity between the suspect and the fillers in a lineup is important so as to not permit the suspect to stand out. Experiments have long shown that an innocent suspect is more likely to be mistakenly identified if the suspect fits the general description of the perpetrator whereas the fillers do not (Lindsay & Wells, 1980). Although there is debate about the best strategies for selecting fillers to use in a lineup, including a concern about whether some strategies might produce too much similarity (Clark & Tunnicliff, 2001; Wells, Rydell, & Seelau, 1993), the evidence across a large number of studies clearly shows that low-similarity fillers result in a poorer ratio of accurate to mistaken identifications than do moderate- or higher-similarity fillers (Fitzgerald, Price, Oriet, & Charman, 2013; Fitzgerald, Oriet, & Price, 2014). Clark’s (2012) meta-analysis reports 67% correct identifications and 31% false identifications for lower-similarity fillers and 59% correct identifications and 16% false identifications for higher-similarity fillers. Figure 3 shows information gain curves for lower-similarity fillers (left panel) versus higher-similarity fillers (right panel) based on Clark’s meta-analysis.

The most obvious observation from the information–gain curves in Figure 3 is that the informational value of an identification of the suspect is greater for higher-similarity fillers than for lower-similarity fillers. This makes good common sense because the identification of the suspect from a lineup in which the fillers are dissimilar to the culprit could be attributable merely to the suggestiveness of the lineup rather than to recognition memory. But there are other interesting observations from Figure 3. In particular, notice how the information–gain curves for identifications of the suspect intersect with the information gain curves for no identification of the suspect. In the case of the low-similarity lineups, the point at which the two curves intersect is when the
base rate is 52%. When the base rate is below 52% for low-similarity lineups, identifications of the suspect have more information value (in the incriminating direction) than do no identifications of the suspect (in the exonerating condition).

Notice as well in Figure 3 that the information–gain curves also intersect for the high similarity lineups. However, the intersection of the two curves occurs at a higher base rate (when the base rate is 79%) for the higher-similarity lineups than for the lower-similarity lineups. That higher point of intersection is to be expected because the difference in information gain for identifications of the suspect and nonidentifications is greater for the high than for the low similarity lineups. As we will show with some other system variables, it is not always the case that information–gain curves for identifications of the suspect intersect and cross over with information–gain curves for nonidentifications. When the difference between the information–gain curves for identifications of the suspect and nonidentifications is large enough (i.e., when identifications of the suspect are much more informative of guilt than nonidentifications of the suspect are of innocence), this difference will maintain itself across the entire base rate and the curves will not intersect. Nevertheless, even when suspect identifications are always more informative than are nonidentifications, the difference in informational value between identifications of the suspect and nonidentifications diminishes dramatically as one approaches higher base rates (e.g., above 90–95%). In effect, these curves illustrate that the tendency to perceive that identifications of the suspect are more informative (of guilt) than nonidentifications are (of innocence) should diminish as the base rate for guilt approaches high values.

**Biased Versus Unbiased Instructions**

Biased lineup instructions are those that either fail to warn the witness that the culprit might not be in the lineup or imply that the culprit is in the lineup. Unbiased instructions, in contrast, warn the witness that the culprit might not be in the lineup. Biased instructions increase the rate of choosing, both when the culprit is present and when the culprit is absent (Malpass & Devine, 1981; Steblay, 1997). However, the ratio of accurate to mistaken identifications is better for unbiased lineup instructions than for biased lineup instructions. The estimate from Clark’s (2012) review of the literature is 59% correct identifications and 15% false identifications for biased instructions and 50% correct identifications and 9% false identifications for unbiased lineups. Figure 4 shows information gain curves for these data. Two observations are important to note here. First, information gain from an identification of the suspect will intersect with a nonidentification curve when the following condition is met: $0 < \frac{LR_{ids} + LR_{nid} - 2}{2(LR_{ids} - 1)(1 - LR_{nid})} < 1$

where $LR_{ids}$ is the likelihood ratio,

$$LR_{ids} = \frac{p(IDS|SC)}{p(IDS|SNC)}$$

and $LR_{nid}$ is the likelihood ratio

$$LR_{nid} = \frac{p(NID|SC)}{p(NID|SNC)}$$

$2 An information–gain curve for identifications of the suspect will intersect with a nonidentification curve when the following condition is met:

$0 < \frac{LR_{ids} + LR_{nid} - 2}{2(LR_{ids} - 1)(1 - LR_{nid})} < 1$

where $LR_{ids}$ is the likelihood ratio,

$$LR_{ids} = \frac{p(IDS|SC)}{p(IDS|SNC)}$$

and $LR_{nid}$ is the likelihood ratio

$$LR_{nid} = \frac{p(NID|SC)}{p(NID|SNC)}$$
overall (Lindsay & Wells, 1985; Steblay et al., 2011). The estimate from Clark’s (2012) review of the literature is 54% correct identifications and 15% false identifications for simultaneous presentations and 43% correct identifications and 9% false identifications for sequential presentations. Figure 6 shows information gain curves for simultaneous versus sequential lineups. As expected, information gain from an identification of the suspect is greater for sequential lineups than for simultaneous lineup. As with lineup instructions and administrator influence, however, the opposite trend occurs for witnesses who make no identification of the suspect. Among those who make no identification of the suspect, the simultaneous lineups are somewhat more informative (in the innocence direction) than are sequential lineups.

Lineups Versus Show-Ups

A show-up is an identification procedure that, unlike a lineup, does not use fillers at all. Instead, the eyewitness is presented with one individual and asked if that person is the culprit. Show-up identification procedures are relatively common under conditions in which a crime occurred very recently (e.g., within the last hour) and an individual who fits the witness’s description of the culprit is detained in near vicinity to the crime. The legal rationale is that the police do not have probable cause to arrest the person (and, hence, the person cannot be detained long enough to conduct a full lineup procedure) and yet setting the person free could be a public safety issue. The general finding in the literature is that witnesses are less likely to make an identification with a show-up than they are with a lineup (Steblay, Dysart, Fulero, & Lindsay, 2003). However, a lineup yields a lower rate of mistaken identification for an innocent suspect than does a show-up because identification errors with lineups tend to get distributed to fillers, a type of control that does not exist with show-ups (Dysart, Lindsay, & Dupuis, 2006). Clark’s (2012) review estimated 41% correct identifications and 18% false identifications for show-ups and 43% correct identifications and 11% false identifications for lineups. As shown in Figure 7, information gain from an identification of the

![Figure 4](image-url)  
Figure 4. Information–gain curves for identifications of suspects and no identifications of suspects for biased lineup instructions (left panel) and unbiased lineup instructions (right panel). Based on Clark’s (2012) meta-analysis of 23 studies.

![Figure 5](image-url)  
Figure 5. Information–gain curves for identifications of suspects and no identifications of suspects for higher administrator influence (left panel) and lower administrator influence (right panel). Based on Clark’s (2012) meta-analysis of 11 studies.
suspect is greater for lineups than for show-ups. And, unlike the other four system variables, information gain from not identifying the suspect is also greater for lineups than for show-ups. In other words, the lineup versus show-up information–gain curves are the only ones among the five system variables for which a manipulated system variable improved information gain for both identifications of the suspect and also for nonidentifications of the suspect. Moreover, the show-up versus lineup variable produces the largest differences in information gain among the traditional system variables. And information gain from the lineup versus the show-up might be even more pronounced in the real world than in lab studies because show-ups typically involve clothing bias (see Dysart & Lindsay, 2007; Dysart, Lindsay, & Dupuis, 2006) and other suggestive practices (e.g., suspect in handcuffs) that would further reduce the informational value associated with witness identification behavior from show-ups.

Information Gain as a Function of Witness Confidence

Eyewitness identification confidence has long been an interest of eyewitness identification researchers. The first formal meta-analysis of the relation between eyewitness identification confidence and accuracy was conducted by Sporer, Penrod, Read, and Cutler (1995). Their analysis of 30 studies showed that the confidence-accuracy relation is strong among choosers with a mean standardized difference in confidence between correct and incorrect choosers approaching a full standard deviation ($d = .89$). For nonchoosers, in contrast, the mean standardized difference was small ($d = .24$). In 1996, Juslin, Olsson, and Winman (1996) demonstrated that calibration methods showed great promise for using confidence to estimate accuracy in eyewitness identification. In 1998, a white paper of the American Psychology–Law Society (and Division 41 of the American Psychological Association) on recommendations for lineup procedures concluded that it is imperative to explicitly collect a confidence statement from identifying witnesses (Wells et al., 1998). Even more recently, a review of the eyewitness identification literature by the National Research Council (2014) of the National Academies of Science made a very strong recommendation that law enforcement should document the witness’s level of confidence verbatim at the time when she or he first identifies a suspect. It is critical that this confidence statement

![Figure 6](image6.png)

**Figure 6.** Information–gain curves for identifications of suspects and no identifications of suspects for simultaneous lineups (left panel) and sequential lineups (right panel). Based on Clark’s (2012) meta-analysis of 51 studies.

![Figure 7](image7.png)

**Figure 7.** Information–gain curves for identifications of suspects and no identifications of suspects for show-ups (left panel) and lineups (right panel). Based on Clark’s (2012) meta-analysis of 15 studies.
be documented at the time of the identification (not at a later time) because we know that social influence variables, especially suggestions to mistaken eyewitnesses that they are correct, can dissociate confidence from accuracy (Wells & Bradfield, 1998, 1999). In addition, researchers have recently been using Receiver Operating Characteristic (ROC) curves to analyze differences between eyewitness identification procedures (e.g., simultaneous vs. sequential lineups, show-ups vs. lineups) and conditions (e.g., good vs. poor witnessing conditions) in an attempt to separate discriminability from response bias (e.g., Carlson & Carlson, 1999). In addition, researchers have recently been using ROC curves to analyze differences between eyewitness identification procedures (e.g., simultaneous vs. sequential lineups, show-ups vs. lineups) and conditions (e.g., good vs. poor witnessing conditions) in an attempt to separate discriminability from response bias (e.g., Carlson & Carlson, 1999; Gronlund, Wixted, & Mickes, 2014; Mickes, Flowe, & Wixted, 2012; Wixted & Mickes, 2014). These ROC analyses depend heavily on the relation between eyewitness identification confidence and accuracy.

In the current section, we offer the first Bayesian analysis of the confidence-accuracy relation in eyewitness identification. In our analysis, we calculate information–gain curves for all three outcomes (suspect identifications, filler identifications, and nonidentifications) at each of five levels of confidence. As far as we can tell, there is only one published study that includes data on confidence in filler identifications and nonidentifications and is large enough to generate stable estimates at several levels of confidence. The Juslin et al. (1996) study, for example, had only 64 witnesses per condition to examine five levels of confidence (averaging fewer than 13 dichotomous data points per level of confidence). This is not atypical of eyewitness identification experiments; they are often underpowered. Moreover, the Juslin et al. study did not collect confidence for nonidentifiers. This is also typical of eyewitness identification experiments as the interest tends to focus (rather myopically we think) on the confidence only of those witnesses who make identifications of the suspect.

Fortunately, there is one large-scale published experiment (Brewer & Wells, 2006) that included 1,200 participant-witnesses and collected confidence for every witness (those who identified suspects, identified fillers, and nonidentifiers). We extracted the relevant data from witnesses’ attempts to identify a waiter they had seen in a video-recorded restaurant scene in which there was a theft of a credit card. The participant-witnesses viewed a lineup that either contained the culprit plus seven fillers (for 600 witnesses) or contained eight innocent people (for the other 600 witnesses). All witnesses were warned that the culprit might not be in the lineup and were given an explicit option to indicate that none was the culprit. After making an identification or nonidentification, witnesses were asked to indicate their confidence in their decision by selecting one of 11 responses from 0% confident to 100% confident in 10% intervals.

The confidence responses were grouped into five confidence categories, namely 0% to 20%, 30% to 40%, 50% to 60%, 70% to 80%, and 90% to 100%, to stabilize the sample sizes within levels of confidence. The sample sizes for the five confidence levels were 91, 213, 375, 330, and 191 for the lowest to highest levels of confidence, respectively. We calculated information–gain curves for each of the three possible responses (identification of suspect, identification of filler, and nonidentification) at each level of confidence. We then calculated the probability of identifying the culprit separately for each level of confidence by dividing the frequency of culprit identifications in the culprit-present condition (at that level of confidence) by the total number of witnesses in the culprit-present condition who were at that same level of confidence. Likewise, we calculated the probability of identifying the innocent suspect separately for each level of confidence by dividing the frequency of innocent suspect identifications in the culprit-absent condition (at that level of confidence) by the total number of witnesses in the culprit-absent condition who were at that same level of confidence. This same method was used for nonidentifications and for filler identifications. Finally, we calculated an information–gain curve for each confidence level at any given base rate. Hence, for example, we can use Figure 8 to ask how confident an eyewitness needs to be in his or her identification for us to claim a 90% or greater probability that the suspect is the culprit. The answer, of course, depends on the base rate. At any given base rate, one can simply add the information–gain value from the information–gain curve for a given level of confidence to the base rate to yield the posterior probability that the suspect is the culprit. For example, if the base rate is 40%, the witness needs to be 90% confident to conclude that the probability is greater than 90% that...
the suspect is the culprit. If the base rate is 60%, then the witness needs to be only 30% to 40% confident in his or her identification to reach 90% probability that the suspect is the culprit. If the base rate is a mere 20%, however, even an eyewitness who is 90% to 100% confident cannot raise the probability that the suspect is the culprit to 90%.

There is a general consensus that, unlike witnesses who make identifications, the confidence-accuracy relation tends to be weak among nonidentifiers (e.g., see Leippe & Eisenstadt, 2006; Sporer et al., 1995). Figure 9 shows information–gain curves for nonidentifications as a function of the five levels of confidence and under the condition of confidence ignorance for the Brewer and Wells (2006) experiment. Consistent with our earlier observation that nonidentifications tend to yield less information gain (in the direction of innocence) than do identifications of the suspect (in the direction of guilt), it is clear from comparing Figure 8 with Figure 9 that the heights of the information–gain curves are lower for nonidentifications of the suspect than for identifications of the suspect. Moreover, the ordering of the information–gain curves by confidence is not quite as well-behaved for nonidentifications as it was for identifications of the suspect. Notice, for example, that the information–gain curve for 50% to 60% confidence is lower than the information-again curve for 30% to 40% confidence. Nevertheless, the highest levels of confidence in nonidentification produce substantial amounts of information gain and the lowest levels of confidence produce almost no information gain. Consider, for example, a base rate of 75%. If the base rate is 75%, then a nonidentification made at the lowest level of confidence (0% to 20%) reduces the probability that the suspect is the culprit by approximately 4% (to a 71% probability that the suspect is the culprit). A nonidentification made at the highest level of confidence (90% to 100%), in contrast, reduces the probability that the suspect is the culprit by approximately 37% (from 75% to a mere 38% probability that the suspect is the culprit). Clearly, confidence in a nonidentification matters.

The considerable level of information gain for high-confidence nonidentifications raises some important questions about why both the eyewitness literature and police practices tend to give little attention to nonidentifications and show almost no interest in the confidence that witnesses express in nonidentifications. Indeed, it is not a common practice in eyewitness identification experiments to collect confidence data from nonidentifiers. And, as mentioned before, 37% of U.S. law enforcement agencies do not even write a report making a record of the fact that a nonidentification occurred, let alone collect a statement of witness confidence in a nonidentification (Police Executive Research Forum, 2013).

But if the purpose of a lineup is to test the hypothesis that the suspect is the culprit, then there is every reason to fully document any results suggesting that the suspect is NOT the culprit rather than only documenting results suggesting that the suspect is the culprit. And because the confidence expressed by nonidentifiers tells us something important about how much to trust the nonidentification (which is clearly evidenced in Figure 9), failure to assess the confidence of nonidentifiers appears to be an indefensible practice.

Finally, we examined information gain for filler identifications as a function of confidence. Figure 10 shows these information–gain curves. Like the identification of suspect curves, the ordering of the curves by confidence is well-behaved with each increase in a confidence category producing a higher curve than the previous. Although information gain is lower for filler identifications than for either suspect identifications or for nonidentifications, it is still quite consequential, especially when the witness is highly confident. For example, when the base rate is 70%, a filler identification by a witness who is at the highest confidence level reduces the probability that the suspect is the culprit by 24%, making it more likely than not that the suspect is in fact innocent (70% prior minus 24% = 46% posterior probability that the suspect is the culprit). Figure 10 depicts what we think is a provocative phenomenon that has never previously been noted. Specifically, Figure 10 shows that the more confident an eyewitness is when making a known error (picking a filler), the more likely it is that the suspect is innocent. Another way of saying this is that the greater the amount of false confidence that an eyewitness has in his or her filler identification, the more likely it is that the suspect in the lineup is innocent. This makes sense to the extent that one realizes that the identification of a filler is a type of rejection of the proposition that the suspect is the culprit because the witness is saying that the filler is a better likeness of the culprit than is the suspect. And, the more confident the witness is in picking the filler, the more the witness is rejecting the proposition that the suspect is the culprit. Remember that in actual cases fillers are known a-priori to be innocent; they are in the lineup to “fill it out” and make it fair to the suspect. The lead author of the current article has been advocating to law enforcement for many years that they should ask witnesses about their confidence even if they pick a filler. The response is typically “Why would I do that? The witness picked a filler so I already know that the witness is mistaken.” But in Figure 10 we see vividly that the confidence with which an eyewitness identifies a filler has informational value about the innocence of the suspect.

Figure 9. Information–gain curves for nonidentifications as a function of witnesses’ confidence in their nonidentification decisions. Data from Brewer and Wells (2006), n = 1,200.
The final innovation in our Bayesian treatment of eyewitness identification evidence revolves around the idea that we alluded to earlier in this article, namely that the base-rate variable is a system variable. The base-rate variable is a system variable because judgments, decisions, and policies that are under the control of the justice system serve to drive the base rate up or down. We will have more to say about this type of control over the base rate in the final section of this article. In this section, we introduce a new type of curve that displays how much change in the base rate is equivalent to the effect of a given system variable in terms of achieving the same level of the posterior probability from an identification of the suspect.

Consider, for example, the lineup instructions system variable. As we saw in Figure 4, information gain from an identification of the suspect is greater for unbiased instructions (right-hand panel of Figure 4) than it is for biased instructions (left-hand panel in Figure 4). That, in turn, means that the posterior probability of guilt from an identification of the suspect is greater for unbiased instructions than it is for biased instructions. Figure 11 displays the prior-by-posterior curves for biased and unbiased instructions for an identification of the suspect (based on Clark’s (2012) meta-analysis data). We have marked on Figure 11 two different points on the base rate, one at 40% and one at 80%. The width between the two dotted lines represents the amount of increase in the base rate that would be required to make the biased instructions equivalent to the unbiased instructions in terms of the posterior probability of guilt following an identification of the suspect. At the 40% base rate, for example, the biased instructions yield a posterior probability of .72 whereas the unbiased instructions yield a posterior probability of .79. For the biased instructions to reach the .79 posterior probability reached by the unbiased instructions, the base rate for the biased instructions would need to be increased from 40% to 49%, an increase of 9%. Notice that, as the base rate approaches the upper values or the lower values, it takes less change in the base rate to make the biased instructions perform as well as the unbiased instructions. Hence, for example, when the base rate is 80% the base rate would need to increase only 5% in the biased instructions to perform as well as the unbiased instructions. This 5% in the latter example is the BREE score for lineup instruction bias when the base rate is 80%.

A BREE score can be calculated for any given base rate using the following equation for the purpose of determining how much increase in the base rate is needed to make a lower diagnostic procedure produce a posterior probability equal to that of the higher diagnostic procedure:

\[ \Delta BR = \frac{BR \left( 1 - \frac{d_l}{d_h} \right)}{1 - BR + \frac{d_l}{d_h}} \]

where \( BR \) is base rate, \( d_l \) is the ratio of accurate to mistaken identifications of suspects for the lower diagnostic procedure and \( d_h \) is the ratio of accurate to mistaken identifications of suspects for the higher diagnostic procedure. Figure 12 displays BREE curves, which are simply all BREE scores for every point in the base rate from base rates of zero to 100%. We have created these curves for five system variables, namely filler similarity, instructions, administrator influence, simultaneous versus sequential, and show-ups versus lineups. The relative heights of each curve are directly related to the effect sizes of the individual system variables. Hence, show-ups versus lineups has the highest curve and simultaneous versus sequential has the lowest curve because the effect size (on posterior probabilities) is the largest for show-ups versus lineups and the smallest for simultaneous versus sequential lineups.
Notice that all the BREE curves tend to peak when the base rate is in the 45% to 50% range. Consider the show-up versus lineup BREE curve. At its maximum (which occurs when the base rate is 46%), the show-up needs a base-rate increase of 13.4% to produce a posterior probability of guilt equivalent to that of a lineup. At their maximums, the other system variables need 13.3% (filler similarity), 9.8% (administrator influence), 8.7% (instructions), and 7.0% (simultaneous vs. sequential). But these are the maximum points. If the base rate is 75%, the amount of increase needed in the base rate is 8.7% (show-up vs. lineup), 8.6% (filler similarity), 6.6% (administrator influence), 5.9% (instructions), and 4.9% (simultaneous vs. sequential).

Our BREE curve analysis in Figure 12 permits us to quantitatively display one of the most important conclusions of our entire analysis, namely that even relatively modest increases in the base rate can have more impact on the reliability of eyewitness evidence than do any of the traditional system variables. These traditional system variables have dominated the eyewitness identification literature for more than two decades. Recommendations from scientific psychology to the legal system regarding eyewitness identification have revolved exclusively around the traditional system variables (e.g., see the “white paper” of the American Psychology-Law Society and Division 41 of the American Psychological Association, Wells et al., 1998). At one level, this makes sense. After all, eyewitness researchers are primarily cognitive and social psychologists whose interests are in memory and social influence. The base-rate variable, however, is neither a cognitive nor a social influence variable. Nevertheless, it was psychology that prominently recognized and demonstrated the propensity for people to fail to understand the impact that base rates have on the actual probabilities of events (Kahneman, 2011). Moreover, it was psychology that defined the idea of system variables in eyewitness identification as any variable over which the justice system has control that affects the reliability of eyewitness identification evidence (Wells, 1978).

For the most part, the problem has not been that eyewitness researchers have failed to consider the role of base rates in eyewitness identification. Almost 35 years ago, the first treatment of the role of base rates in eyewitness identification appeared in the psychological literature (Wells & Lindsay, 1980). The problem has been that the base-rate variable in eyewitness identification has been inadequately developed and the base rate has been treated as though there was some unknown single number “out there” in the real world that characterizes lineups. But there is no mysterious single number “out there” that represents the base rate. Instead, there is an entire spectrum of base rates that can run from very low numbers to very high numbers depending on the policies and practices of the individual detectives who decide when or whether to conduct a lineup. There are no laws (nor have we been able to find any police department guidelines) regarding lineups that have anything to say about how much evidence they ought to have against a suspect before placing that suspect into a lineup. A national survey of U.S. law enforcement officers found that more than one third of crime investigators indicated that they needed either no evidence at all or only a mere hunch that a person might be the culprit before they would conduct a lineup (Wise, Safer, & Maro, 2011).

Cases of mistaken identification in which innocence was proven using DNA evidence are rife with examples in which there was little or no evidence to suggest that the person placed in an identification procedure was the culprit (Garrett, 2011). Brenton Butler was a 15-year-old living in Jacksonville, FL when he was walking through a neighborhood close to an area in which a woman had recently been murdered. As police stood around remarking that they knew little about the shooter other than that he was a Black male with a fisherman’s hat, someone spotted Butler and said “There is a Black male.” Soon, Butler found himself mistakenly identified by the case’s sole witness. Despite the fact that there was no fisherman’s hat, no gun, no stolen purse, and no bloody clothes, the prosecution pressed forward against Butler. The Butler case, vividly retold in the documentary film Murder on a Sunday Morning, does not prove anything about the base rate “out there.” But it helps illustrate our point that the decision to subject someone to the jeopardy of an identification procedure lies solely within the discretion of individual law enforcement officers for whom there are no guidelines or training on the type or amount of evidence they might want to have before making someone the subject of an identification procedure (Wells, 2006).

Consider two hypothetical police departments. In one of these departments, which we will call the Lax Police Department, police investigators are quite willing to conduct an eyewitness identification procedure based on a mere hunch—perhaps based on someone who comes to mind or on a convenient possible suspect who broadly fits the description. Over a period of time, the Lax Police Department is likely to be running a fairly low base rate. Consider another department, which we will call the Strict Police Department. In the Strict Police Department, investigators make sure that they have actual evidence pointing to a specific person before making that person the subject of an identification procedure. Over time, we can expect the Strict Police Department to run a much higher base rate than the Lax Police Department. The result is fewer mistaken identifications and a greater ability to trust the identifications obtained by the Strict Police Department than by the Lax Police Department.

Our novel BREE curves in Figure 12 are one way to think about how much a given increase in the base rate increases trust in an
identification of the suspect relative to the impact of other system variables. Suppose, for example, the Lax Police Department was running a base rate of 60% with their lineups. Increasing that base rate by 15% (to 75%) would produce far more trust in an eyewitness’s identification of a suspect than any other system-variable improvement. In fact, even the bane of all eyewitness identification procedures, the much maligned show-up, yields more trustworthy results from an identification than does the lineup if the show-ups use the 75% base rate and the lineup uses the 60% base rate.

None of this work on BREE curves should be taken as an indication that the traditional system variables should be forsaken in a pursuit of higher base rates. After all, an increase in base rates plus good use of the traditional system variables will produce better outcomes than either alone. But, BREE curves show that construing of the base rate as a system variable is at least as important as the traditional system variables in terms of its impact on outcomes.

The Reasonable-Suspicion Proposition

The observation that even modest increases in the base rate substantially reduce the chances of mistaken identification and produce substantially higher posterior probabilities of guilt has led to a proposition that there be reasonable suspicion before placing an individual into the jeopardy of an identification procedure (Wells, 2006). Figure 1, which was based on a meta-analysis of 94 published studies, helps make this point quite clearly. Consider points A and B in Figure 1, representing a 25% base rate and a 75% base rate. Keep in mind that these two data points represent exactly the same eyewitnesses with exactly the same ability to discriminate between a guilty and an innocent suspect. But, when the base rate is 25%, the posterior probability of guilt is a mere 53.4%. In other words, 46.6% of all identified suspects under this base rate are innocent. If the base rate is 75%, however, the posterior probability of guilt is 91.2% and only 8.8% of all identified suspects under this base rate are innocent. We can readily translate this to frequencies. For a police department running a 25% base rate, 53 of every 100 witnesses who identify the suspect will be accurate and 47 of every 100 identifications of the suspect will be cases of mistaken identification. For a police department running a 75% base rate, 91 of every 100 witnesses who identify the suspect will be accurate and 9 of every 100 identifications of the suspect will be cases of mistaken identification. These are huge differences that make very salient the idea that it is rather dangerous to be conducting lineups when the base rate is low.

The fact that the mistaken identification rate is vastly greater with the lower base rate than with the higher base rate is fundamentally no different than the observation we made earlier in this article in which prostate tests for cancer produce mostly false alarms for men under 30 years of age but produce vastly fewer false alarms for men over the age of 50. This difference in false alarms from the PSA test is not attributable to the test itself being less reliable for younger males than for older males. Instead, it is attributable to the very low base rate for prostate cancer in younger males. Hence, the recommendation is to not do PSA tests on men under 30 even though the PSA test is just as diagnostic for men under 30 as for men over 50.

The reasonable-suspicion proposition for lineups (Wells, 2006) states that there ought to be some actual evidence indicating that there are reasonable grounds for believing that the suspect is the culprit before placing the suspect in the jeopardy of a lineup. The idea of reasonable suspicion, and the dangers of acting against a person without reasonable suspicion, is not at all foreign to the legal system. In the United States, the term “reasonable suspicion” is not of constitutional derivation (unlike probable cause, which is explicitly recognized in the fourth amendment). Instead, the idea of reasonable suspicion was fashioned by the U.S. Supreme Court in 1968 (Terry v. Ohio) to describe a level of suspicion lower than probable cause. For example, whereas probable cause is required for an arrest, only reasonable suspicion is required for temporary investigative detention.

The legal system has not given any probability number for reasonable suspicion, just as it has not given a probability number for probable cause or for the concept of reasonable doubt. And it is not expected that the legal system will ever put a numerical probability on these concepts. Nevertheless, it is not necessary to use a number in order to make the idea of reasonable suspicion a meaningful working concept in the legal system. In the eyewitness identification area, for example, the system could start with the general idea that putting someone in the jeopardy of an identification procedure ought to be based on something more than a mere hunch. How much more than a mere hunch is something that policymakers in the legal system would have to determine, but unless some kind of minimal criterion is instituted for justifying placing a person in an identification procedure, it will continue to be the case that a higher standard is required to detain someone on the street for questioning (or pulling their car over) than is required for putting that person into an identification procedure.

A reasonable-suspicion requirement would not be difficult to implement. For example, a police department could simply institute a requirement that their detectives receive permission from a supervisor of detectives before putting together a lineup. The supervisor would simply ask some questions such as “Who is the suspect?” “What evidence do you have that leads you to think he is the culprit in this case?” Departments would have to work out some kind of criteria that prevent individuals from being placed in lineups based on mere hunches or SPECIOUS reasons so as to minimize the chances of conducting lineups at the low end of the base rate, where mistaken identifications are going to occur with fairly high probabilities.

Science cannot determine what the criteria should be for reasonable suspicion any more than it can specify what constitutes probable cause or reasonable doubt. But what we can do, and have done here, is show the profound implications of what happens when investigators conduct lineups at the low end of the base-rate continuum. Our development of a new type of analysis, BREE curves, also allows us to compare the relative value of traditional system variables, such as unbiased prelineup instructions, in units of base-rate percentages. The BREE curves show that when the base rate is low it takes very little increase in the base rate to have far more power than any of the traditional system variables. For example, raising the base rate from a level of 30% to a level of 40% has a greater impact on raising the posterior probability of guilt following an identification of the suspect than does changing from the use of biased lineup instructions to the use of unbiased lineup instructions. We can also use BREE curves to examine
 equivalents when the base rate is lowered. For example, a mere 7% reduction in the base rate (from 80% to 73%) makes a lineup perform no better than a show-up in terms of the posterior probability that a person identified is guilty.

Caveats Regarding the Reasonable-Suspicion Proposition?

A Freshness of Memory Cost?

A criticism of the argument for a reasonable-suspicion criterion might be found in a potential loss in the freshness of an eyewitness’s memory that might result from attempts to raise the base rate. At the extreme, we could imagine having little or no evidence on a possible suspect named Barry S. (e.g., for whom the expected base rate might be 20%) within 12 hours of the crime. With more investigation (e.g., 24 hours later) it might be possible to find more evidence against this possible suspect or to discover that it is far more likely (e.g., 75%) that the actual culprit is Peter N. and it is Peter N. who should be in the lineup. Here, we have a potential dilemma. If Barry S. was the perpetrator and investigators took another 24 hours to find enough evidence to surpass the reasonable suspicion criterion on Barry S., then the passage of time could have produced some forgetting and lowered the chances that the witness would be able to identify Barry S. On the other hand, if the search for more evidence led to Peter N. and he was the culprit, then it was wise to not have conducted the lineup with Barry S. After all, if the witness had identified anyone from a lineup built around Barry S., the witness has either made a mistaken identification or has been spoiled for any later identification of the actual culprit, Peter N.

The most famous of the DNA exoneration cases is the case of Ronald Cotton, a man who was misidentified by Jennifer Thompson and ultimately freed after 11 1/2 years in prison when DNA testing proved that her attacker was actually Bobby Poole. Shortly after Cotton was misidentified and arrested, Poole arose as a possible suspect (Simon, 2012) but the police already had their guy – Ronald Cotton. Counterfactuals can never be known with certainty, but it is reasonable to speculate that had investigators been a bit more patient and not been so quick to place Cotton in a lineup (based on a highly questionable “tip”), the involvement of Poole (a convicted rapist) would have surfaced and he would have been the subject of the lineup instead of Cotton.

Although the Cotton case is an example of what we might call a “premature” decision to conduct a lineup, the question remains as to what the trade-off might be between taking the extra time needed to develop reasonable suspicion on the one hand and the freshness of the witness’s memory on the other. Unfortunately, the literature on the passage of time and eyewitness identification is relatively small, tends to be an older literature, and has not taken a parametrically systematic approach. A meta-analysis by Shapiro and Penrod (1986) confirmed that accurate identifications diminish with delay between the witnessed event and the time of the identification procedure. But, the rate of decline is highly variable across the small number of studies that have been conducted. Perhaps all we can conclude at this point is that it seems likely that most of the forgetting probably occurs in the first moments after the culprit is no longer in view, more occurs in the first hour than the second hour, more in the first day than the second day, more in the first week than the second week, and so on. In other words, it seems most likely that the ability to recognize the culprit follows the classic negatively accelerating forgetting curve (Ebbinghaus, 1885). Hence, the decision to delay a lineup by another 24 hours so as to gather more evidence related to reasonable suspicion is likely a different one if the crime occurred only 24 hours ago than if it occurred one week ago or one month ago. In a study of 1,039 actual lineups conducted in the United Kingdom, Horry et al. (2012) found that the median delay to an identification procedure was 31 days. Other than show-ups (which are almost exclusively restricted to no more than 1–2 hours after the crime), it is rare for a lineup to be conducted in less than 24 hours and, based on the Horry et al. data, most were conducted after at least three weeks.

It is not our intent to be dismissive of the argument that attempts to ensure a higher base rate could harm the freshness of an eyewitness’s memory, but if the median lineup is conducted 30 days after the witnessed event, the addition of another 24 to 72 hours seems unlikely to have any significant effect on this “freshness” factor.

The Exceptional Eyewitness Exception to Reasonable Suspicion?

Our entire Bayesian analysis assumes that there is a nontrivial error rate in eyewitness identifications. And there are good reasons to believe that this is true based on laboratory data (e.g., see Table 1), DNA exoneration cases, and the increasingly compelling data showing that witnesses pick known-innocent people from actual lineups approximately 33% of the time that they make an identification (e.g., Behrman & Davey, 2001; Horry et al., 2014; Memon, Havard, Clifford, Gabbert, & Watt, 2011; Wells, Stelbay, & Dysart, 2015). But, it is also the case that the impact of the base rate depends critically on there being a nontrivial error rate. If, for example, eyewitnesses were so reliable that they always identified the culprit in culprit-present lineups and always rejected lineups in which the suspect was not the culprit, then the base-rate variable would have no impact at all.

This idea of an errorless eyewitness suggests a potential exception to the reasonable-suspicion proposition. Suppose, for example, a victim-eyewitness had been abducted by a culprit and held for two weeks. During these two weeks the victim observed the unmasked culprit for hours at a time, day after day. In effect, the victim came to know the person quite well, forming a deep and detailed memory. After the two weeks, the victim escapes and the police have a hunch about who might be the culprit. Under these conditions, we can reasonably speculate that there is little or no risk of a mistaken identification resulting from showing the victim-witness a lineup that does not include the actual culprit because the witness’s memory of the culprit is so good that the witness would readily reject the culprit-absent lineup. Hence, if these circumstances existed in an actual case, the concern about needing reasonable suspicion before placing a possible suspect in a lineup is moot because there is no inherent risk of mistaken identification.

A Need-to-Avoid-Flight Exception to Reasonable Suspicion?

Another possible exception to requiring reasonable-suspicion might exist when there is a possible suspect who might be a “flight
risk.” Consider an example. A victim-witness describes her assailant as a White male, mid 20s in age, short brown hair, medium build, and no facial hair. The case detective thinks that it might be Timothy C. but has no evidence beyond the general description. Fitting this general description is not itself reasonable suspicion for putting Timothy C. in a photographic lineup. Perhaps if the detective investigated the case for another 48 hours or so, there might be actual evidence to make a case for reasonable suspicion of Timothy C. or, alternatively, to point to an alternative suspect for whom there are sufficient grounds for reasonable suspicion. However, suppose that the detective has some fact-based reasons to believe that Timothy C. is making plans to leave the area and is a significant flight risk. Because there is not sufficient evidence to pass a reasonable-suspicion requirement, there clearly is not probable cause (a higher standard), which would be needed to arrest Timothy C. In this type of situation, there is a direct conflict between the need for reasonable suspicion to conduct a lineup and the risk of flight emanating from the additional time needed to make a case for reasonable suspicion. Because courts consider the identification of a suspect to be probable cause for arrest, however, the detective might be able to solve the case and prevent flight of a culprit by conducting the identification procedure now rather than waiting until there is reasonable suspicion.

In this hypothetical case, the flight risk consideration in no way changes the inherent jeopardy of the lineup, nor does it change the base-rate probability that the suspect is the culprit or any of the math we have presented here. However, from a legal policy perspective, the flight risk example might well constitute an exception to the reasonable suspicion requirement that, on balance, might allow a lineup without reasonable suspicion. The legal system makes these kinds of exceptions in a variety of domains. For example, police are required to make a case for probable cause and obtain a search warrant from a magistrate or judge before they can enter a person’s private residence against the wishes of the resident. But, if police are in a “hot pursuit” or have reason to believe that there would be imminent destruction of evidence before a warrant can be obtained, no warrant is needed to enter, even forcibly. The point of this is that the legal system is quite capable of structuring some rule (such as reasonable suspicion being required before placing someone in the jeopardy of a lineup) while carving out well-reasoned exceptions for conditions in which there might be reasons to override the rule.

**Bayesian, ROC, and Calibration Approaches: Similarities and Differences**

Readers should note that we deliberately did not title this section Bayesian versus ROC and Calibration approaches. We do not necessarily see these three approaches as rivals for what is the “proper” way to approach eyewitness identification data. Each approach is attempting to answer a somewhat different question. By this point, it is clear what questions the Bayesian is trying to answer. The Bayesian approach is concerned with the posterior probability that a suspect is guilty given a behavioral response of the eyewitness (identification of suspect, nonidentification, filler identification). Necessarily, such probabilities require consideration of the base rate. Derivations from these calculations (such as information gain) tell us how much we need to revise the hypothesis that the suspect is the culprit once we know the behavioral response of the witness. And BREE curves tell us how much change in the base rate would be needed to equal the impact of a traditional system variable (such as biased vs. unbiased instructions) in terms of achieving the same posterior probability of guilt. The posterior probability is the primary interest of the triers of fact (e.g., judges and juries) who are trying to assess the likely guilt or innocence of the identified person.

**The ROC Approach**

Unlike the Bayesian approach, Receiver Operating Characteristic (ROC) curves purport to control for response bias, thereby yielding a clean measure of discriminability (Mickes, Flowe, & Wixted, 2012). In fact, however, it is not clear that the ROC approach is properly controlling for response bias or that it measures discriminability. The problem is that the ROC approach treats all filler identifications as if they were rejections. But filler identifications are not rejections, they are false positive identifications. And because the false positive identification rate is critical for calculating response bias (which is a propensity to make a positive identification), it is unclear how ROC curves on lineups are controlling for response bias when it treats filler identifications (positive responses) as if they were rejections (negative responses).

Our point, however, is that the ROC approach attempts to control for response bias. The Bayesian approach, in contrast, does not control for differences in response bias. Instead, the Bayesian approach calculates posterior probabilities and ancillary measures (such as information gain and BREE scores) with both discriminability and response bias driving the results. In this sense, Bayesian analyses depict the data as they are, regardless of whether observed differences are due to discriminability differences or response bias.

Another difference between ROC curves and the current Bayesian approach is that ROC curves examine only identifications of the suspect. What they do not tell us is how good witnesses are at rejecting the lineup (indicating “none of the above”) because ROC analyses treat filler identifications and rejections as the same thing, collapsing them into the denominator of the calculations.

**The Calibration Approach**

The calibration approach to eyewitness identification data attempts to assess how well the confidence of an eyewitness relates to the probability that an identification is accurate. Ideally, an eyewitness who is 90% confident should have a 90% chance of being accurate, an eyewitness who is 70% confident should have a 70% chance of being accurate, and so on. Juslin et al. (1996) proposed the calibration-curve approach as an alternative to the common point-biserial correlation, pointing out that the point-biserial correlation can be low even when calibration is quite good. In their original article, Juslin et al. reported very good calibration from their experiment. Other researchers have found that calibration tends to be good for those making an identification but much poorer for nonidentifiers (e.g., Brewer & Wells, 2006). We agree that calibration can be a useful metric for examining confidence-accuracy relations in eyewitness identification and that it sometimes makes more sense than using the point-biserial correlation.

An important metric in calibration is overconfidence/underconfidence. Overconfidence occurs when the percent of witnesses
whose identifications are correct is lower than their percent confidence and underconfidence occurs when the percent of witnesses whose identifications are correct is higher than their percent confidence.

Despite some broad claims that eyewitnesses tend to be quite well calibrated, it turns out that estimates of overconfidence and underconfidence depend on the base rate. This is because changes to the base rate influence the proportions of accurate and mistaken identifications but do not change the confidence of the witnesses.

To illustrate this, we used the Brewer and Wells (2006) data to examine calibration curves under the assumption of base rates of 30%, 50%, and 70%. Figure 13 shows these three calibration curves. The dashed line represents perfect calibration. A calibration curve below the diagonal line represents overconfidence and a calibration curve above the line represents underconfidence. The middle curve is based on the 50% base rate and shows what appears to be relatively good calibration with only slight underconfidence for those whose confidence was below 50% and overconfidence for those whose confidence was above 50%. But if the base rate is 70%, the curve indicates underconfidence for low confidence witnesses and nearly perfect calibration for very high confidence witnesses. If the base rate is 30%, however, the curve shows appreciable overconfidence throughout the entire curve except for the very lowest confidence witnesses.

We believe that calibration analyses have a lot to offer in terms of studying the confidence-accuracy relation in eyewitness identification. However, we see that forensically relevant analyses of the eyewitness identification problem simply cannot escape the fact that base rates play a heavy role in conclusions that we might want to reach about overconfidence and underconfidence.

Summary, Discussion, and Conclusions

We began our Bayesian analysis of eyewitness identification evidence with the observation that the purpose of lineups in the real world is to test the hypothesis that the suspect is the culprit rather than testing the reliability of the witness. This requires consideration of the base rate, for which a Bayesian approach is a natural fit. We then showed how different behavioral responses of eyewitnesses (identifications of the suspect, nonidentifications, and filler identifications) have a curvilinear impact on the posterior probability that the suspect is the culprit as a function of the base rate. We followed this by using information–gain curves to capture the extent to which the probability that the suspect is the culprit changes as a function of the behavioral response of the witness across all possible levels of the base rate. Information–gain curves were then created for five different system variables to show how different levels of these system variables impact information gain across all levels of the base rate. Finally, we created an entirely new type of curve, the base-rate effect–equivalency curve. BREE curves show how much change in the base rate is required for one level of a system variable to perform as well as another level of a system variable in terms of being able to match the posterior probability that the suspect is the culprit following an identification of the suspect. The BREE curves show that even relatively small increases in the base rate can match the impact of these traditional system variables and that the amount of base-rate increase required to do this depends on the base-rate starting point.

In unfolding our Bayesian analysis, we interwove the often-overlooked observation that the base rate in lineups is not any single number in the real world but is instead a system variable. The base rate is a system variable because it can be increased or decreased by the justice system through the practices and policies that the system puts in place (just as with any other system variable). Moreover, as the BREE curves dramatically illustrate, changes to the base rate have far greater potential to reduce mistaken identifications and increase accurate identifications (hence, higher posterior probabilities of guilt if the witness identifies the suspect) than do the traditional system variables. We described one possible mechanism for avoiding very low base-rate practices that is already quite familiar to the legal system, namely the mechanism of requiring reasonable suspicion prior to placing a suspect in a lineup.

Our analysis raises a provocative question: Why has the legal system ignored the question of whether there ought to be some kind of reasonable-suspicion criterion before placing an individual into the inherent risk of an eyewitness identification procedure? We see several possible reasons.

First, the failure of the legal system to concern itself with the base rate in lineups might simply be an example of the legal system being a reflection of normal human intuition. Human intuition commonly neglects (or at least greatly underutilizes) base rates when estimating the probabilities of various events. And, in fact, a compelling case has been made that the legal system mirrors the same biases and cognitive illusions that characterize the individual perceiver (Simon, 2012). But why is the medical system in tune with base rates while the legal system is not? After all, medical diagnostic tests are very similar to eyewitness identification tests. In both cases there is a probabilistic test of some sort (e.g., a PSA prostate test or a lineup) that runs the risk of a false alarm or a false rejection. Given a positive result, there is some increase in the posterior probability that a proposition is true (that the person has cancer or that the suspect is guilty). But that posterior probability is not simply a function of the diagnosticity of the test; it depends

Figure 13. Calibration curves for choosers as a function of three possible base rates based on Brewer and Wells (2006) data.
very much on the base rate for the truth of the proposition (the rate of cancer in the population tested or the rate of guilt in the lineups conducted). The medical world studies this problem very seriously and develops clear guidelines for when the test should and should not be used, for example, telling doctors to stop giving PSA prostate tests to males under 40. But in the legal system there are no guidelines, nor is there even discussion, about the dangerous increase in false alarms that can arise from the practice of using low base-rate lineups. The difference between the medical system and the legal system might be explained in part by the fact that medicine has long embraced modern science and mathematics in its development of tests. In contrast, the legal system’s tests, with only one exception, have not been subjected to the rules of science and mathematics.3

A second possible reason that the legal system has never addressed the question of whether to require reasonable suspicion before placing an individual in an identification procedure is that the system assumes that eyewitness identification tests are extremely accurate. In fact, it is true that if one is using a perfectly accurate test, base rates play no role at all in the chances of error. This is true for any test, including medical tests. Of course no test is perfect but, at the extreme (e.g., a test that has less than 1/10000th chance of an incorrect classification), the role of base rates would be very minimal. But, except perhaps for the exceptional witness (recall our hypothetical abduction witness), there is no reason to believe that eyewitness identification has a level of accuracy that makes the influence of base rates minimal. Across the 94 lab studies analyzed by Clark et al. (2008), for example, witnesses mistakenly identified an innocent suspect more than 13% of the time they were shown a culprit-absent lineup. Outside of the lab, in lineups conducted by law enforcement in actual cases, witnesses identify a known-innocent filler 33% of the time that they make an identification. And mistaken eyewitness identification is implicated in approximately 75% of DNA exoneration cases.

A third possible reason why the legal system has never considered a reasonable-suspicion criterion before doing an identification procedure is because the system assumes that there already is reasonable suspicion; otherwise why would this person be placed into an identification procedure? Indeed, eyewitnesses might make this assumption as well. But, as we noted already, a large share of U.S. police report that they are quite willing to conduct an identification procedure even when they have no evidence at all or are relying on a mere hunch that the person is the culprit (Wise, Safer, & Maro, 2011). And, in a study that examined actual lineups, it was found that 40% of the lineups were conducted under conditions in which there was no extrinsic evidence against the suspect (Behrman & Richards, 2005). Moreover, the mere assumption that an individual becomes the subject of an identification procedure only when there is reasonable suspicion belies the common practice of using random mug book search procedures with witnesses. And, as we have already noted, we have found no police department guidelines, procedures, or training manuals in the U.S. that ever touch on the issue of considering whether there is actual evidence against a person before deciding to conduct an identification procedure. Hence, we see nothing at all to indicate that there is any reason to believe that a reasonable-suspicion type of consideration is playing any part in the legal system’s understanding of how and when to use eyewitness identification procedures.

3 Readers curious about this claim should read the National Research Council’s (2009) assessment of the state of forensic “science” testing. With the exception of forensic DNA testing, which was developed by scientists, none of the common forensic science tests (e.g., tool marks, bite marks, handwriting analysis, hair matching, bloodstain pattern analysis, ballistics) was developed by scientists, their error rates are mostly unknown, and all remain largely untested using scientific methods.

References


Dysart, J. E., Lindsay, R. C. L., & Dupuis, P. R. (2006). Show-ups: The
critical issue of clothing bias. Applied Cognitive Psychology, 20, 1009–
1023. http://dx.doi.org/10.1002/acp.1241

Ebbinghaus, H. (1885). Memory: A contribution to experimental psy-

in eyewitness lineups: A literature review and a novel methodology. Law
and Human Behavior, online first. http://dx.doi.org/10.1007/hb0000095

Garrett, B. (2011). Convicting the innocent: Where criminal prosecu-
tions go wrong. Cambridge, MA: Harvard University Press.

Greathouse, S. M., & Kovera, M. B. (2009). Instruction bias and lineup
presentation moderate the effects of administrator knowledge on eye-
doi.org/10.1007/s10979-008-9136-x

Hons, C. R., & Schweinle, W. (2009). Information gain of psychophys-
iological detection of deception in forensic and screening settings. Ap-
org/10.1007/s10484-009-9096-z

analyses of eyewitness identification test outcomes: What can they tell
us about eyewitness memory? Law and Human Behavior, 38, 94–108.
http://dx.doi.org/10.1007/s10979-008-9136-x

Eyewitness identification decisions from video lineups in England: A
10.1007/s10979-012-9359-8

innoenceproject.org/

of confidence in eyewitness identification: Comments on what can be
inferred from the low confidence–accuracy correlation. Journal of Ex-
perimental Psychology: Learning, Memory, and Cognition, 22, 1304–
1316. http://dx.doi.org/10.1037/0278-7393.22.5.1304

Straus & Giroux.


Lampinen, J. M., Erickson, W. B., Moore, K. N., & Hittson, A. (2014). The
effects of distance on face recognition: Implications for eyewitness
dx.doi.org/10.3758/s13423-014-0641-2

recollection rejection in the memory conjunction paradigm. Journal of
Experimental Psychology: Learning, Memory, and Cognition, 30, 332–
342. http://dx.doi.org/10.1037/0278-7393.30.2.332

confidence-accuracy relationship in memory for people. In R. C. L.
Lindsay, D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), Handbook of
eyewitness psychology: Memory for people (pp. 377–425). Mahwah, NJ:
Erlbaum.

Handbook of eyewitness psychology: Memory for people (Vol. 2). Mah-
wah, NJ: Erlbaum.

Lindsay, R. C. L., & Wells, G. L. (1980). What price justice? Exploring the
relationship of lineup fairness and identification accuracy. Law and
Human Behavior, 4, 303–313. http://dx.doi.org/10.1007/BF01040622

Lindsay, R. C., & Wells, G. L. (1985). Improving eyewitness identifica-
tions from lineups: Simultaneous versus sequential lineup presentation.
0021-9010.70.3.556

University Press.

instructions and the absence of the offender. Journal of Applied Psy-
chology, 66, 482–489. http://dx.doi.org/10.1037/0021-9010.66.4.482

construction and lineup fairness. In R. C. L. Lindsay, D. F. Ross, J. D.
Read, & M. P. Toglia (Eds.), Handbook of eyewitness psychology: Memory

field evaluation of the VIPER system: A new technique for eliciting eyewitness identification evidence. Psychology, Crime & Law, 17, 711–
729. http://dx.doi.org/10.1080/1068316090352433

characteristic analysis of eyewitness memory: Comparing the diagnostic
accuracy of simultaneous versus sequential lineups. Journal of Experi-
mental Psychology: Applied, 18, 361–376. http://dx.doi.org/10.1037/a003069

National Research Council. (2009). Strengthening forensic science in the
Press.

National Research Council. (2014). Identifying the culprit: Assessing eye-

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in
org/10.1037/1089-2680.2.2.175

Pickel, K. L. (2007). Remembering and identifying menacing perpetrators:
Exposure to violence and the weapon focus effect. In R. C. L. Lindsay,
D. F. Ross, J. D. Read, & M. P. Toglia (Eds.), The handbook of eyewitness
psychology: Vol. 2. Memory for people (pp. 339–360). Mahwah, NJ:
Erlbaum.

Police Executive Research Forum. (2013). A National Survey of Eyewit-
ness Identification Procedures in Law Enforcement Agencies. Retrieved

Rodriguez, D. N., & Berry, M. A. (2014). The effect of lineup adminis-
strator blindness on recording of eyewitness identification decisions.
Legal and Criminological Psychology, 19, 69–79. http://dx.doi.org/
10.1111/lcpr.2014.20.0058.x

Rozenthal, R., & Rubin, D. B. (1978). Interpersonal expectancy effects:
The first 345 studies. Behavioral and Brain Sciences, 1, 377–386.
http://dx.doi.org/10.1017/S0140525X00007506

dx.doi.org/10.3758/BF03209339

0033-2909.100.2.193

Simon, D. (2012). In doubt: The psychology of the criminal justice process.
harvard.9780674065116

confidence, and accuracy: A meta-analysis of the confidence-accuracy
relation in eyewitness identification studies. Psychological Bulletin, 118,
315–327. http://dx.doi.org/10.1037/0033-2909.118.3.315

Stelbay, N. M. (1997). Social influence in eyewitness recall: A meta-
analytic review of lineup instruction effects. Law and Human Behavior,