Eyewitness Identification: The Importance of Lineup Models

Gary L. Wells and John W. Turtle
University of Alberta, Edmonton, Alberta, Canada

A detailed analysis is made of lineup models for eyewitness identification. Previous treatments of eyewitness identification have not distinguished between the all-suspect model and the single-suspect model. The single-suspect model allows for the occurrence of foils identifications, a known-error category, whereas the all-suspect model does not. A Bayesian analysis of posterior probabilities of the guilt of a given suspect under various prior probabilities shows that the all-suspect model may be more or less diagnostic than the single-suspect model depending on the extent to which the use of suspects rather than foils increases the prior likelihood that the actual target is in the lineup. On the other hand, the lineup-wise error rate (which is the likelihood that any suspect will be falsely identified) is considerably higher with the all-suspect lineup. Field data show that the all-suspect lineup is sometimes used by police departments, and some data suggest that police do not appreciate the distinction between the two models with regard to lineup-wise error rates. It is recommended that either a mixed model or a preceding blank lineup be used to replace all-suspect models in actual cases.

Although eyewitness testimony received some attention from experimental psychologists over 75 years ago (Munsterberg, 1908), it has only been within the last decade that significant efforts have been made to build a useful literature. Six recent books have been devoted entirely to the psychology of eyewitness testimony (Clifford & Bull, 1978; Lloyd-Bostock & Clifford, 1982; Loftus, 1979; Shepherd, Ellis, & Davies, 1982; Wells, & Loftus, 1984; Yarmey, 1979), and articles have appeared in almost every major psychological and legal journal. There are now hundreds of eyewitness experiments, many dealing with the issue of eyewitness identification.

Despite experimental evidence showing relatively high rates of false identification, controversy exists over the extent to which false identification rates are also high in real-world cases, and, if they are high, what can be done about them (see Egeth & McCloskey, 1984; Loftus, 1983; Wells, 1984a). The experimental research shows false identification rates of 20% or even 60% under some conditions. Clearly, the rates obtained in these experiments are well above the false identification rates that are obtained in real-world cases. Yet this apparent discrepancy between laboratory findings and real-world cases may be better explained at another level.

We argue that the rates of false identification obtained in experiments might in fact be a fair representation of the frailties of eyewitness identification. This is not to say that false identification rates in real cases are as high as 60% or even 20%. As discussed in the next section, not all misidentifications qualify as false identifications, and there are base rates (or prior probabilities) that must be considered as well. The analyses that we present show how the likelihood of false identification can be quite low or quite high depending on the "lineup model" used in the case. (For the precise meaning of lineup model see the section on Single-Suspect Versus All-Suspect Models.) Lineup models have not been considered previously in the psychological or legal literature on eyewitnessing, yet they might account not only for the apparent discrepancies between laboratory eyewitness research and real-world cases, but also for differences between police jurisdictions in their rates of false identification.

Before discussing models of lineups, two important distinctions must be introduced. These distinctions are (a) known versus unknown errors and (b) target-present versus target-absent lineups.

Necessary Distinctions

Known and Unknown Errors

Most eyewitness researchers recognize immediately the fact that not all errors committed by eyewitnesses are equal in consequence. A false identification of a suspect from a lineup, for...
example, can be construed as more harmful than a no-identification response from an eyewitness. In the former case, a person is accused falsely, and the true target remains at large; in the case of a no-identification response, the true target is at large but no one is accused falsely (Wells, 1978). However, not all misidentifications are necessarily false identifications, a fact that is understood only when one distinguishes between suspects and foils (Wells & Lindsay, 1985).

A foil is a member of a lineup or photo-spread who is not suspected of the crime in question. The main purpose of foils is to have distractors so as to make the eyewitness’s task one of discrimination. Courts of law generally require multiple lineup members or photo-spread members rather than a “show-up” (i.e., a one-to-one confrontation between a suspect and a witness) because of the suggestiveness of show-ups (Sobel, 1972). Eyewitness researchers endorse the need for a true lineup rather than a show-up and have gone further to argue for distractors who at a minimum match the suspect on general physical characteristics (e.g., Doob & Kirschenbaum, 1973; Malpass, 1981; Wells, Leippe, & Ostrom, 1979). Empirical research supports the idea that physical similarity between foils and suspect is important (Lindsay & Wells, 1980).

Note, however, that the identification of a foil, although a true identification error, is a “known error.” That is, in an actual lineup or photo-spread situation, the identification of a foil will be detected as an error. The identification of a foil does not result in charges being brought against the identified person. In other words, the identification of a foil is not a false identification in the forensic sense. Throughout the remainder of this article, we use the term false identification only to refer to the identification of an innocent suspect; we call inaccurate identifications of foils foil identifications.

If the model being used by police involves foils, then there exists a category of eyewitness behavior that can be called a known error. Without foils, however, there can be no known-errors in real cases. A no-identification response, for example, may or may not be an error depending on whether or not the true target is in the lineup or photo-spread. In experimental research, of course, we can always know whether an identification or no-identification response is accurate or inaccurate. In actual cases, however, there must be at least one foil among the distractors in order to have any chance at classifying a witness’ response as an error. In this limited sense it is not true that police can never know if an eyewitness has made an identification error.

**Target-Present Versus Target-Absent Lineups**

One of the most fundamental problems in the eyewitness identification literature is the failure of many studies to include target-absent lineups or photo-spreads in their experimental designs. Detailed arguments as to why it is necessary to include target-absent lineup conditions have been discussed elsewhere, and, therefore, only an overview of the reasons will be given here (see Lindsay & Wells, 1980, 1985; Malpass & Devine, 1981, 1984; Wells, 1984b; Wells & Lindsay, 1980, 1985; Wells & Murray, 1983, 1984).

A major argument for the use of target-absent lineups is similar to that involved in signal detection theory regarding the importance of “noise alone” trials, namely the need to control for response bias. The issue is less obvious with lineups, however, than it is in the case of signal detection tasks because lineups use distractors. Therefore, there is a tendency to think of the distractors as some form of control that will rule out response bias explanations of a given set of results. This would be true, however, only if the lineup or photo-spread were a forced-choice task, whereas ecological validity requires eyewitness identification tasks that allow for a “none-of-the-above” response. Therefore, it is only through the use of target-absent lineups that the noise-alone aspect of identification can be examined.

In actual cases it sometimes occurs that the true target is not in the lineup or photo-spread. This fact alone makes it important that we know how the data come out in target-absent lineups and photo-spreads. What is not known is the frequency or probability of eyewitnesses being exposed to target-absent lineups. Wells and Lindsay (1980) have pointed out that the prior probability of the eyewitness being tested with a target-present versus target-absent lineup has strong effects on the information that is gained from a given response by an eyewitness (i.e., an identification of a suspect, an identification of a foil, or no identification). Because their interest was in comparing information gain for identifications versus nonidentifications, this prior probability was of considerable import. Wells and Lindsay dealt with this problem by using Bayesian statistics to calculate information gain across all possible prior probabilities (from 0 to 1.0). Because the current article is concerned with comparisons between different models, we use a similar Bayesian solution to the problem.

There are numerous other reasons why it is important to use target-absent lineups in research designs. First, data from target-absent lineups have interpretational value so as to discriminate between response bias and memory strength (cf. signal detection theory). Second, real-world cases sometimes involve lineups in which the suspect(s) are innocent; thus generalization to real-world situations requires some data on what happens under these conditions. Finally, several techniques for improving lineup identification by eyewitnesses have failed to show effects for target-present lineups while showing marked improvements for target-absent lineups (via reduced misidentification rates). These techniques include warnings that the target may not be present in the lineup (Malpass & Devine, 1981), the use of higher functional-size lineups (Lindsay & Wells, 1980), the use of dual lineups (Wells, 1984b), and the use of sequential rather than simultaneous presentations of lineup members (Lindsay & Wells, 1985).

Thus, there are reasons on almost all fronts for including target-absent lineups in research designs. These reasons include ecological considerations (it happens in real cases), mathematical considerations (the way prior probabilities affect distributions), methodological considerations (to control for response biases), and empirical considerations (research shows that it makes a difference to results and conclusions).

---

1 “Functional size” refers to the number of physically similar foils in a lineup as distinct from “nominal size,” which refers to the number of members in the lineup regardless of their physical appearance. Although there are some technical difficulties with operationally defining the functional size of a lineup (see Malpass, 1981; Wells, Leippe, & Ostrom, 1979), the conceptual import of functional size and related measures (e.g., effective size) is apparent.
Table 1
Distribution of Eyewitnesses' Photo-lineup
Choices by Lineup Condition

<table>
<thead>
<tr>
<th>Lineup condition</th>
<th>Lineup member, %</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Target present</td>
<td>2.1</td>
</tr>
<tr>
<td>Target absent</td>
<td>4.2</td>
</tr>
</tbody>
</table>

Note. n = 48 for both conditions. N/A = "none-of-the-above" response. Adapted from Wells (1984b).

* This lineup member is the target (in the target-present condition) or her or his substitute (in the absent condition). The remaining lineup members are the same in both lineup conditions.

Single-Suspect Versus All-Suspect Models

In view of the factors and distinctions put forward thus far, it is now possible to discuss the two generic models of lineups and photo-spreads. The single-suspect model is one in which there is only one suspect, the distractors being known-innocent foils. The foils may be police officers, persons from holding cells, or private citizens, but the important point is that they are not suspects and that an identification of any of them is a known error. Many previous attempts at detailed analysis of the psychology of lineups have been based on the single-suspect model on the assumption that it was the prevailing model in use by all police departments (e.g., Lindsay & Wells, 1980, 1985; Wells, 1984b; Wells & Lindsay, 1985). However, a survey of 22 police departments in the U.S. midwest calls that assumption into question. We discuss this survey in more detail in the section Current Lineup Practices.

The all-suspect model contains no foils; each member of the lineup or photo-array is a suspect. This means that there is no response that the eyewitness can make that can be classified as a known error in actual cases that use the all-suspect model. Any identification of a lineup member is likely to result in charges being laid against that lineup member in the all-suspect lineup.

In order to make clearer some of the implications of the distinction between the all-suspect versus single-suspect lineup, consider the data presented in Table 1. These data (from a staged-theft study reported in Wells, 1984b) illustrate a distribution of responses that is typical of eyewitness identification studies in several respects. The following aspects to the pattern of data are worthy of note.

1. The individual lineup member receiving the greater number or proportion of responses is the actual target. This is what should be expected given that eyewitness performance is above chance levels.

2. None-of-the-above responses are more frequent given that the target is not in the lineup than when the target is in the lineup. Again this is consistent with what should be expected if eyewitness performance is above chance levels.

3. The chances that an innocent lineup member will be selected by an eyewitness are greater if the lineup does not contain the target than if the lineup contains the target.

4. Choices of innocent lineup members are not distributed randomly.

Most researchers do not report their data in a way that allows a breakdown like that given in Table 1. However, this type of pattern (as described in Observations 1-4 above) appears to hold for data reported fully enough to be subjected to such a breakdown (e.g., Lindsay & Wells, 1980; Loftus, 1976; Wells, Ferguson, & Lindsay, 1981; Wells & Leippe, 1981) and fits well with a logical analysis of above-chance performance.

Table 2 is the choice data from Table 1 converted into outcomes under the all-suspect and single-suspect models. Notice that there are several outcomes that cannot occur, and these differ as a function of the particular combination of lineup condition and model. Notice also that the models are identical in their expected rate of accurate identifications as well as their expected rates of correct and false rejections. Differences between the models occur for the outcomes of foil identification and false identification.

Distinguishable Responses

Although Table 2 breaks the data down into five outcome categories (accurate, false, and foil identifications; correct and false rejections), only three responses are distinguishable in cases where the guilt or innocence of the suspect is in doubt. This is, of course, the case in all real-world eyewitness problems. Specifically, a witness can identify a suspect, identify a foil, or make a none-of-the-above response. We can move from these three response categories to the five categories only if it is known whether or not the suspect is guilty. The question of concern is in fact whether or not the suspect is guilty (or the witness is accurate). Therefore, the goal is to estimate the probability of witness accuracy (or suspect guilt) given one of the three distinguishable responses under various conditions (e.g., the prior probability of suspect guilt and the lineup model being used). Note also that in the case of the all-suspect model there are only two distinguishable responses, namely the identification of a suspect and a none-of-the-above response. As will become apparent later, the diagnostic properties of the all-suspect model are complicated by the fact that it can yield only two distinguishable responses rather than the three that can be ascertained from the single-suspect model.

Even a casual glance at Table 2 shows that false identifications, in the forensic sense as defined earlier, are much more likely with the all-suspect model than with the single-suspect model. This is true regardless of whether or not the lineup contains the target. However, it does not necessarily follow that the single-suspect model will always prove superior to the all-suspect model in the likelihoods of accurate and false identifications. What Table 2 indicates is simply that the all-suspect model is much more likely to produce false identifications than the single-suspect model.
Table 2
Distribution of Outcomes (in Percentages) by Lineup Condition and Model

<table>
<thead>
<tr>
<th>Lineup condition</th>
<th>Identification of suspect</th>
<th>Known error</th>
<th>“None-of-the-above” response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Accurate</td>
<td>False</td>
<td>Foil identification</td>
</tr>
<tr>
<td>Target present</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All-suspect model</td>
<td>60.4</td>
<td>20.9</td>
<td>*</td>
</tr>
<tr>
<td>Single-suspect model</td>
<td>60.4</td>
<td>*</td>
<td>20.9</td>
</tr>
<tr>
<td>Target absent</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>All-suspect model</td>
<td>*</td>
<td>71.0</td>
<td>*</td>
</tr>
<tr>
<td>Single-suspect model</td>
<td>*</td>
<td>31.3</td>
<td>*</td>
</tr>
</tbody>
</table>

Note. The rows should add up to 100% after accounting for rounding. * = the outcome cannot occur.

does not reflect is the fact that as one increases the number of suspects, the likelihood of having a target-present lineup should likewise increase. Thus, in the next section we use Bayesian statistics to examine the relative trade-off between the fact that an all-suspect model increases the likelihood that the target is present while increasing the likelihood that an identification falls into the false identification (unknown error) category rather than the foil identification (known error) category.

Information Gain

Wells and Lindsay (1980) devised a measure, information gain, that reflects the informativeness of a given eyewitness response. Information gain is reflected in the difference between a prior probability and a posterior probability. For example, if the prior probability that a suspect is guilty is .75 and the posterior probability (e.g., given a positive identification) is .83, then the information gain value is .08. Of course, the prior probability can be greater than the posterior probability if the new information is of the form that is exonerating rather than incriminating. Thus, information gain is expressed as an absolute difference between the prior and posterior probabilities.

In general, information gained from an identification of a suspect is defined as

\[ |p(S = T) - p(S = T|IDS)|, \]

where \( p(S = T) \) is the prior probability that the suspect is the true target and \( p(S = T|IDS) \) is the posterior probability that the suspect is the target given an identification of the suspect. The value of \( p(S = T) \) is unknown, but its role can be understood by calculating information gain across all possible prior probabilities, that is, for \( p(S = T) \) from 0 to 1.0. The value of \( p(S = T|IDS) \) can be determined for any set of data such as that provided in Tables 1 and 2. Specifically:

\[
p(S = T|IDS) = \frac{p(IDS|S = T)p(S = T)}{p(IDS|S = T)p(S = T) + p(IDS|S \neq T)p(S \neq T)},
\]

where \( S \neq T \) is a condition where the suspect is not the target.

Similarly, information gained from a rejection or none-of-the-above response is defined as

\[ |p(S = T) - p(S = T|NA)|, \]

where \( NA \) is a none-of-the-above response. The value of \( p(S = T|NA) \) is derived from the equation

\[
p(S = T|NA) = \frac{p(NA|S = T)p(S = T)}{p(NA|S = T)p(S = T) + p(NA|S \neq T)p(S \neq T)}
\]

Finally, the informativeness of a foil identification is defined as

\[ |p(S = T) - p(S = T|FI)|, \]

where \( FI \) is a foil identification and

\[
p(S = T|FI) = \frac{p(FI|S = T)p(S = T)}{p(FI|S = T)p(S = T) + p(FI|S \neq T)p(S \neq T)}
\]

Using this definition of information gain one can compare information gained from an identification of a suspect with that gained from a foil identification or a none-of-the-above response (as in Wells & Lindsay, 1980). However, the purpose here is to compare information gained from the single-suspect model with that gained from the all-suspect model.

Likelihood Ratios

One approach to comparing the single-suspect model with the all-suspect model is to hold constant the prior probability value of \( p(TP) \), which is the prior probability that the target is present in the lineup. Under these conditions, the likelihood ratio \( p(IDS|TP) \div p(IDS|TA) \) determines the diagnosticity of an identification of the suspect for estimating the probability that the lineup includes the target. Determining the probability that the lineup contains the target is important because that probability is identical to the probability that the suspect is the true target when dealing with the single-suspect model. It also is important when dealing with the all-suspect model because the probability that the lineup contains the target is a parameter that must be calculated as an intermediate step toward calculating the probability that a given suspect is the true target. In the case of the data in Table 2, the likelihood ratio for the all-suspect lineup is
approximately 1.15 (i.e., \([60.4 + 20.0] / 71.0\)), and for the single-suspect lineup is 1.93 (i.e., \([60.4 + 31.3]\)). These likelihood ratios indicate that the all-suspect lineup is less informative than the single-suspect lineup for diagnosing the likelihood that the target is present in the lineup. For the single-suspect model, this ratio of 1.93 means that witnesses were almost two times more likely to identify the guilty suspect than they were to identify the innocent suspect. The case of the all-suspect model is more complex, as discussed in the next section.

The identification of a foil also has diagnostic value for deciding whether or not a suspect is the target. As shown in previous work (Wells & Lindsay, 1980), foil identifications are more likely to occur if the target is absent from the lineup than if the target is present in the lineup. Because foil identifications have exonerating value for the accused (i.e., foil identifications lower the likelihood that the suspect is guilty), the likelihood ratio for foil identifications takes the form \(p(\text{FIS} \neq T) / p(\text{FIS} = T)\). In the case of the single-suspect lineup data in Table 2, the obtained likelihood ratio is 1.90 (i.e., \([39.7 + 20.9]\)). This means that the witnesses were almost nine times more likely to identify a foil if the target was not in the lineup than if the target was in the lineup. Note, however, that the likelihood ratio for the all-suspect lineup regarding foil identifications necessarily is zero because, by definition, foil identifications cannot occur with an all-suspect lineup.

The likelihood ratio for none-of-the-above responses takes a form similar to that for foil identifications, namely \(p(\text{NAIS} \neq T) / p(\text{NAIS} = T)\). This value is identical for the two lineup models and is 1.55 (i.e., \([29.2 + 18.8]\)) for the data in Table 2. The obtained likelihood ratio in this case can be interpreted to mean that a witness is 1.55 times more likely to make a none-of-the-above decision if the lineup does not contain the guilty person than if the lineup does contain the guilty person. Because the likelihood ratio for none-of-the-above responses is unaffected by the lineup model, we pay this response category little subsequent attention in this article.

**Lineup Models, Likelihood Ratios, and Prior Probabilities**

Despite the foregoing discussion, it is not clear that the single-suspect model is superior to the all-suspect model. At least two levels of the issue must be explored in order to make precise conclusions regarding the two models. First, there is the issue of how much an identification of a suspect changes the posterior probability of guilt relative to a given prior probability of guilt under both models. Such an analysis is concerned with an individual suspect. After exploring the parameters of such an analysis with respect to the data in Table 1, a second analysis will focus on what we call "lineup-wise" error rate, which is the probability that a false identification will occur in the lineup as a whole (rather than at the level of an individual suspect).

The traditional forensic concern regarding eyewitness identification is focused at the level of an individual suspect after she or he has been identified by an eyewitness. Expressed in the form of conditional probabilities, the question can be phrased as \(p(S = T | \text{IDS})\) as defined earlier. As noted earlier, in order to estimate this posterior probability, the prior probability, \(p(S = T)\), must be specified. This prior probability can be expressed as follows for the all-suspect model:

\[
p(S = T) = p(S = T | TP)p(TP) + p(S = T | TA)p(TA).
\]

This expression can be abbreviated, however, by the fact that \(p(S = T | TA)\) takes on a value of zero by definition. Thus,

\[
p(S = T) = p(S = T | TP)p(TP)
\]

for the all-suspect model. For the single-suspect model, however, the expression can be abbreviated even further by the fact that \(p(S = T | TP) = 1.0\) in all cases. Thus,

\[
p(S = T) = p(TP)
\]

for the single-suspect model.

A meaningful comparison of the two models requires that the prior probabilities of the suspect's guilt be equivalent under both models. This can be achieved either by holding constant \(p(S = T | TP)\) and varying \(p(TP)\) with the all-suspect model or by holding constant \(p(TP)\) and varying \(p(S = T | TP)\) with the all-suspect model or some combination of the two. For the first analysis, we chose to hold constant \(p(TP)\) for the all-suspect model and we set its value at 1.0. Analyses based on this decision are quite favorable to the all-suspect model in that it presumes that the target is in the lineup in all cases. This analysis, therefore, accepts the most extreme argument that those advocating all-suspect lineups could possibly hold, namely that by using so many suspects one is certain to have the true target in the lineup. In this case, the posterior probability is calculated as

\[
p(S = T | \text{IDS}) = \frac{p(\text{IDS}|S = T)p(S = T)}{p(\text{IDS}|S = T)p(S = T) + p(\text{IDS}|S \neq T)p(S \neq T)}
\]

Using a numerical example from Table 2, suppose that the prior probability \(p(S = T) = .60\). In this case the posterior probability \(p(S = T | \text{IDS})\) equals

\[
\frac{.604(.60)}{.604(.60) + .209(.40)} = .801.
\]

A second analysis of the posterior probabilities for the all-suspect lineup set \(p(S = T | TP)\) equal to \(p(TP)\). Thus, for example, the prior probability \(p(S = T) = .64\) results from \(p(S = T | TP) = .80\) and \(p(TP) = .80\). In this second analysis the posterior probabilities are considerably lower than that obtained when \(p(TP) = 1.0\), even when the prior probabilities of \(p(S = T)\) are identical. The reason for this is that the second analysis introduces some possibility that the target is absent and false identification rates are higher in target-absent conditions than they are in target-
present conditions. In this case, the posterior probability expression requires an expansion to account for the fact that \( p(TP) \neq 1.0 \) and therefore, \( p(S \neq T) \) has two components, \( p(S \neq T|TP) \) and \( p(S \neq T|TA) \). The general expression is

\[
p(S = T|IDS) = \frac{p(IDS|S = T)p(S = T)}{p(IDS|S = T)p(S = T) + p(IDS|S \neq T)p(S \neq T|TP)p(TP) + p(IDS|S \neq T)p(S \neq T|TA)p(TA)}
\]

Using a numerical example, suppose that the prior probability \( p(S = T) = .64 \). In this case, the posterior probability \( p(S = T|IDS) \) equals

\[
.604(.64) + [(.209)(.20) .80] + [(.708)(1.0)(.20)] = .688.
\]

Unlike the all-suspect lineup, there is only one source of variance for the prior probability \( p(S = T) \) with the single-suspect lineup, namely \( p(TP) \). Figure 1 presents posterior probabilities across all prior probabilities for the two all-suspect analyses and the single-suspect analysis. The diagonal is an identity line; data points would fall on this line if identifications of the suspect had no diagnostic value. The distance of a given point on a curve from the identity line is a measure of information gain, as discussed earlier. Note that the all-suspect model where \( p(TP) = 1.0 \) is associated with the greatest information gain, whereas the all-suspect model where \( p(S = T|TP) = p(TP) \) has little information gain. Which of these two curves best represents the all-suspect model cannot be definitively argued. On the one hand, the all-suspect curve where \( p(TP) = 1.0 \) is unrealistic because one can never be certain that the true target is in the lineup and this curve probably represents the upper limit on the all-suspect model. On the other hand, the all-suspect curve where \( p(S = T|TP) = p(TP) \) probably represents a reasonable lower limit on the all-suspect model.

It is apparent from Figure 1 that there is no firm basis for arguing that the all-suspect model is either more or less informative than the single-suspect model regarding the posterior

![Figure 1](image-url)
probability that a given suspect is in fact the target. If we can assume that the all-suspect model greatly increases the likelihood that the true target is in the lineup, then the all-suspect model might be slightly more informative; if the all-suspect model improves the likelihood of the target’s presence by only a modest amount, the single-suspect model might be more informative.

At this point we return to an issue raised earlier (see Footnote 3). Suppose that the innocent suspect in the target-absent condition was not Lineup Member 4 (see Table 1) and, therefore, was not the most frequently chosen lineup member. Suppose instead that Lineup Member 2 was the suspect. How would this change our conclusions? This would have no implications for the all-suspect model, but it does have implications for the single-suspect model. The effect is to make the single-suspect model much more informative than the previous analyses suggest. In the previous analyses, the magnitude of the difference between the prior and posterior probabilities for the single-suspect model was a function of the likelihood ratio 60.4 / 31.3, or 1.93. If Lineup Member 4 was the suspect, however, the likelihood ratio would be 60.4 / 10.4, or 5.72. Were these data to be plotted in Figure 1, the height of the obtained curve would vastly exceed the all-suspect model (even when the all-suspect model is given the unrealistic advantage of perfect certainty that the target is in the lineup). In other words, by choosing Lineup Member 4 as the innocent suspect for the single-suspect model, we have stacked the deck somewhat against our general argument that the single-suspect model is superior to the all-suspect model. Therefore, the analyses in this section might understate the true differences favoring the single-suspect model over the all-suspect model.

**Lineup-Wise Error Rates**

The posterior probability that a given suspect is the target given an identification of that suspect is importantly distinct from the overall probability of a false identification. In many ways this distinction is analogous to the distinction between error rate for a given statistical contrast and experiment-wise error rate. As concluded in the previous section, there is no definitive difference between the all-suspect model and the single-suspect model as they relate to the posterior probabilities of a given suspect’s guilt. However, the fact that the single-suspect model partitions eyewitness identification errors into foil identifications and false identifications (whereas every identification error is a false identification with the all-suspect model) has important implications for the overall expected rate of false identifications.

We call the overall probability of a false identification the “lineup-wise error rate” to distinguish it from the error rate associated with a given suspect. If one is concerned only with evaluating the posterior probability that a given suspect is guilty (the traditional concern for courts), then the previous section is most pertinent and it appears that the lineup model makes little difference. However, if one is concerned with police practices and the likelihood that a given practice will produce more false identifications in the long run, then it is the lineup-wise error rate that is of most concern.

The two models can be compared and contrasted easily with regard to lineup-wise error rates. The conditional probability expression is

\[
p(FID) = p(FID|TP)p(TP) + p(FID|TA)p(TA),
\]

where \(p(FID)\) is the probability of a false identification and the other terms are as defined previously. Again, \(p(TP)\) is a free parameter and Figure 2 presents \(p(FID)\) for all possible values of \(p(TP)\) for the data presented in Table 1 for both lineup models.

Note that one major difference between the two models is the fact that the single-suspect model is capable of achieving a lineup-wise error rate of zero when \(p(TP) = 1.0\), whereas the all-suspect model can never achieve a zero error rate. This is because the only identification error that can occur in a target-present/single-suspect lineup is the identification of a foil (i.e., a known error) and these are not false identifications in the forensic sense. The all-suspect model, however, can produce false identifications (unknown errors) even when the target’s presence is an absolute certainty. The lower rate of false identifications with the single-suspect lineup than with the all-suspect lineup maintains across all levels of \(p(TP)\); the lineup-wise error rate for the single suspect lineup is .313 at maximum, where \(p(TP) = 0\), and for the all-suspect lineup is .712 at maximum.

Despite the profound differences in lineup-wise error rates, it is possible for the all-suspect lineup to have an equal or lower error rate than the single-suspect lineup, depending on the ability of the all-suspect model to guarantee high levels of \(p(TP)\). Specifically, the lineup-wise error rates are equal if \(p(TP) = 0\) for the single-suspect model and \(p(TP) \geq .792\) for the all-suspect model, or if \(p(TP) \geq .332\) for the single-suspect model and \(p(TP) \geq 1.0\) for the all-suspect model. Theoretically therefore, the lineup-wise error rates for the all-suspect model could be equal to or lower than that for the single-suspect model; realistically, however, it cannot be argued that the lineup-wise error rate would ever favor the all-suspect model. The situation where \(p(TP) = 0\) in the single-suspect case is in fact a contradiction in logic because it is equivalent to a priori certainty that the suspect is innocent (i.e., there is no suspect). As long as the a priori probability that the suspect is the target is greater than .332 in a single-suspect lineup, then the all-suspect model produces higher lineup-wise error rates (see Figure 2).

As a final note about lineup-wise error rates, we return again to the issue raised in Footnote 3. The issue concerns how our conclusions might change if the innocent suspect in the single-suspect lineup were not Lineup Member 4 (the most frequently chosen member of the target-absent lineup, see Table 1). Suppose instead that Lineup Member 2 was the innocent suspect. This change makes no difference for the all-suspect model, but it does have implications for the single-suspect model. Specifically, the line depicting lineup-wise error rates for the single-suspect model in Figure 2 would remain unchanged, but the line depicting these error rates for the single-suspect model would drop to one-third of its current height (beginning at zero and climbing to a maximum height of .104 instead of .313). Thus, our conclusion about the superiority of the single-suspect lineup over the all-suspect lineup is robust with regard to our choice of Lineup Member 4; in fact, the choice of any other lineup member would have made the differences between the two lineup models even larger than that depicted in Figure 2.

**Mixed Model**

A mixed lineup model is one in which there is more than one suspect but, unlike the all-suspect model, there are also foils.
Not surprisingly, the statistical parameters of the mixed model for both the posterior probabilities and for the lineup-wise error rates fall between the single-suspect model and the all-suspect model. In order to estimate diagnosticity and lineup-wise error rates for the mixed model for the data in Table I, one could assign the single-suspect lineup's foils randomly to the categories of foil and suspect using various ratios of foil to suspect.

An important aspect of the mixed model is the fact that it can always stand as an alternative to the all-suspect model in a given case, whereas the single-suspect model might not be a viable alternative to the all-suspect model in a given case. Consider, for example, a case in which there are 6 suspects. In this case it may be unrealistic to conduct six separate single-suspect lineups. However, a mixed model can be created by adding 6 foils for which we can expect lower lineup-wise error rates. Although not as efficient as a single-suspect model for holding down lineup-wise error rates, the mixed model can significantly improve the lineup-wise error rate over that expected from the all-suspect model. There is a continuum of possible mixed models that can be defined by reference to either the proportion of lineup members who are suspects or the ratio of suspects to foils.

A 50% mixed model, for example, is one in which half of the lineup members are foils and half are suspects. In the case described above where there were 6 suspects, a 50% mixed model would require a 12-person lineup (i.e., the 6 suspects embedded among 6 foils). It is important to note that the difference between a 6-person all-suspect lineup and a 12-person mixed lineup is a qualitative difference rather than merely a quantitative difference.

Unlike the all-suspect lineup, the mixed lineup provides for the possibility of a known error occurring (i.e., a foil identification). Note also that a mixed model can be described easily in quantitative terms with a straightforward interpretation. In general, the lower the proportion of suspects the lower the expected likelihood of a false identification (i.e., the lower the lineup-wise error rate).

An important point is that experimental data and lineup theory suggest that the addition of foils should have little effect on the likelihood that the true target will be identified in target-present
lineups (Lindsay & Wells, 1980; Wells, 1984a). Instead, the primary effect of the addition of foils should be to distribute errors across lineup members. Hit rates should drop only slightly according to a formula relating guess rates to number of lineup members. Specifically, we propose that

$$\Delta H = p(G) \frac{1}{L_1} - p(G) \frac{1}{L_2}$$

where $\Delta H$ is the percentage change in hit rate, $p(G)$ is the percentage of witnesses who are guessing, $L_1$ is the number of members in one lineup, and $L_2$ is the number of members in the second lineup. Thus, suppose there were 6 persons in one lineup and 6 new foils were added for the second lineup. Suppose further that 20% of all witnesses were guessing. In this case there would be a trivial 1.6% drop in hit rates for the second lineup. Given the trivial cost to hit rates associated with the addition of foils along with the predictable and robust reduction in lineup-wise error rates, we argue that a mixed lineup model should always be preferred to an all-suspect model.

Logically there is some point at which our equation for changes in hit rates as a function of the number of lineup members breaks down. For example, a 30- or 40-member lineup may produce a form of interference that will more profoundly affect hit rates than our equation predicts. Under such conditions we agree with the Law Reform Commission of Canada (Brooks, 1983) that the task must be treated as a mugshot task and that no live lineup can be justified.

**Current Lineup Practices**

Do police use the all-suspect model? The first author obtained data relevant to this question while visiting 22 police departments in the midwestern United States in 1983–1984. (Further details can be obtained by writing to the first author.) The purpose of the visits was to obtain a sample of actual lineups and photo-spreads for which photographic records were kept so that functional-size analyses could be conducted. For each lineup or photo-spread, police were asked to indicate which lineup member was the suspect, a necessary bit of information for calculating functional-size analyses could be conducted. After visiting the second police department it became clear that the first question should be “How many members of this lineup were suspects and how many were distractors who were not actually suspects?” rather than “Which lineup (photo-spread) member was the suspect?” There were 56 lineups and photo-spreads for which photographic records were kept, but a determination of the number of suspects was not possible for all cases. Of the 56 lineups and photo-spreads, 8 used an all-suspect model, 15 used a single-suspect model, 9 used a mixed model, and the remainder were indeterminate. Part of the problem with these frequencies is that the police departments visited did not code their lineups in a manner that would allow an assessment of whether the nonidentified members were suspects or foils. Thus, the lineups for which photographic records were available had to be presented to the investigating officers for each case in question. Some investigating officers were no longer available and others could not remember, thus the high rate of indeterminate cases. Two cases initially classified as single-suspect lineups by the public relations officer were found to be all-suspect lineups when the investigating officers were interviewed. This kind of problem may have existed for other lineups classified as single-suspect lineups because the officers were prone to define suspects according to a post hoc criterion of who was identified.

In general, there was no evidence to indicate that police departments were particularly sensitive to the issue of lineup models. Indeed, in each of the 22 departments, one experienced investigator was presented with a hypothetical scenario in which an eyewitness to a robbery viewed the robber for 25 s from 15 ft in twilight conditions and was presented with a lineup 2 weeks later. A 6-person lineup was described as being composed of 1 suspect and 5 known-innocent stand-ins or as composed of 6 suspects. Each officer was asked to indicate the likelihood of a false identification of a suspect under these conditions. Their mean estimates were 12.3% in the single-suspect case and 10.8% in the all-suspect case. These results may not be surprising given that the problem is akin to a base-rate problem and there is ample evidence that people ignore or underutilize such statistical dimensions (e.g., Kahneman & Tversky, 1973).

**Conclusions**

Previous treatments of eyewitness identification have failed to consider the role of lineup models. A single-suspect model has important distinct outcome possibilities from that of an all-suspect model in that the former allows for a category of known errors. As well, the likelihood of falsely identifying a given suspect can be quite different for the two models depending on the extent to which the use of an all-suspect model increases the probability that the actual target is present in the lineup. Lineup-wise error rates (i.e., the likelihood that any suspect will be falsely identified) can be quite different for the two models depending on the extent to which the use of an all-suspect model increases the probability that the actual target is present in the lineup. To the extent that one police jurisdiction runs a reliably higher rate of false identifications than another, the lineup models used in those jurisdictions would seem to be the likely culprit; after all, it seems unlikely that eyewitnesses in one jurisdiction are inherently worse than they are in some other jurisdiction. There are no definitive data on this question. However, data from the 22 police departments showed that those reporting having experienced the most difficulty with false identifications (where $I = \text{no difficulty at all}$ and $7 = \text{has been a difficult problem}$) were also more likely to use an all-suspect lineup rather than either a mixed model or single-suspect model (point-biserial correlation $= .41, p \approx .06$).

The single-suspect model is not necessarily a viable alternative to the all-suspect model in a given case unless police switch to a multiple-lineup system. What should police do when there are
several suspects? Minimally they should consider a mixed model. That is, every lineup should contain at least some foils so as to yield the possible result of a known error. Another possibility is to use a “preceding blank lineup,” a technique recently tested in which the actual lineup (i.e., the one with the suspect[s]) is preceded by a lineup composed entirely of foils. The blank lineup has been shown to serve effectively as a lure for eyewitnesses who are most prone to making false identifications and, because all lineup members are foils in this blank lineup, these errors are relatively harmless (Wells, 1984b).

It is interesting to note the implications of the fact that the lineup models differ profoundly at the level of lineup-wise error rate but not at the level of posterior probabilities of false identification for a given suspect. One implication of this is that a defendant has no argument for his defense on the basis of police having used an all-suspect model rather than a single-suspect model. It cannot be said that the individual suspect’s degree of protection or due process is jeopardized by the lineup model being used. Perhaps this is why courts have never addressed the issue. Indeed, the most comprehensive documentation of legal cases and speculative commentary on eyewitness identification (Sobel & Pridgen, 1984) never discusses the issue. In recent years, experimental psychologists have tried to make use of expert testimony to highlight some of the conditions most closely associated with eyewitness errors (for a variety of perspectives on this, see Egeth & McCloskey, 1984; Loftus, 1983; Wells, 1978). The issue of lineup models, however, does not fit comfortably into the expert courtroom testimony mold because the posterior probability of the defendant having been falsely identified cannot be said to be higher with the all-suspect model than with the single-suspect model. Nevertheless, the rate of false identifications in the justice system as a whole is expected to be profoundly greater when the all-suspect model is used, and thus the all-suspect model should be jettisoned.

References


