Running head: ALIBI TAXONOMY

What Makes a Good Alibi? A Proposed Taxonomy

Elizabeth A. Olson and Gary L. Wells

Iowa State University

Portions of these data were presented at the 2001 Biennial Meeting of the Society for Applied Research in Memory and Cognition. This research is based on the first author's Master's thesis.
Correspondence concerning this article should be addressed to Elizabeth A. Olson, Department of Psychology, Iowa State University, W112 Lagomarcino Hall, Ames, Iowa 50011; email: vorndran@iastate.edu.

Abstract

A taxonomy of alibis is proposed involving two forms of supporting proof: physical evidence and person evidence. Levels of physical evidence and person evidence were combined to create twelve cells in the taxonomy. Participants (n = 252), who were asked to assume the role of detectives, evaluated alibis representing these 12 cells. The believability of the alibis generally followed the taxonomy's predicted pattern, but physical evidence, when present, tended to overwhelm the person evidence more than had been expected. In addition, alibi evaluators seemed to not consider the possibility that a stranger who corroborated an alibi might be mistaken about the identity of the person. Trait inferences regarding the alibi providers tended to follow the believability data, even when the traits themselves were not relevant to believability of the alibi. We call for the development of a literature on the psychology of alibis, recommend the taxonomy as a framework, and suggest several avenues of inquiry.

Alibis: A Proposed Taxonomy

> *Alibi*; n. A plea of having been at the time of the commission of an act elsewhere than at the place of commission.

Suppose you discovered, much to your surprise, that you were considered a suspect in a criminal investigation. You are innocent, so you simply have to account for your whereabouts for the time of the crime. If you remember where you were at the time of the crime, you have an alibi—an alibi is simply a claim that you were elsewhere. However, *having* an alibi is the easy part; *proving* an alibi is another matter altogether.

We first developed an interest in alibis in part through examination of cases in which innocent people were convicted by juries and their innocence was established later using forensic DNA testing that was unavailable at the time of their conviction (see Wells et al., 1998). Although the dominant evidence used against these innocent people was eyewitness identification testimony, we noticed that "weak alibis" were often exploited by prosecutors and used as incriminating evidence. Although the individuals were innocent of the offense (and, therefore, their alibis were likely true), their alibis were nevertheless ineffective.

These cases, in which innocent people's alibis did not prevent them from being erroneously convicted, led us to wonder, what is a weak alibi? Are there fundamental properties of alibis that can be identified as strong or weak? This seemed to us to be a natural interface between social psychology and law, and we expected to find an empirical literature addressing the issue. Instead, we have been surprised and a bit frustrated to discover that there is a dearth of empirical literature on alibis.

Although an empirical literature on alibis has not yet been developed, there are a few empirical studies that have used alibis to test hypotheses about other issues, principally

eyewitness identification issues. Leippe (1985) for example, used alibi testimony to examine mock juror judgments of various forms of eyewitness identification and non-identification evidence, but the alibi information was held constant, not manipulated. McAlister and Bregman (1989) manipulated whether an alibi witness positively identified the defendant (thereby corroborating the alibi) or not and, as expected, failure of the alibi witness to corroborate the defendant's alibi led to more guilty verdicts from mock jurors. Lindsay et al. (1986) manipulated whether or not an alibi witness was a relative of the defendant and found that only the non-relative alibi witness was able to reduce convictions when there was an eyewitness who had identified the defendant as the culprit. Although all three of these studies used alibi witnesses, the studies were focused primarily on issues of eyewitness identification, not alibis. Alibis were used merely as tools to find out how people think about eyewitness identification issues. The only empirical study that we have been able to find that was devoted to alibis per se was one conducted by Culhane and Hosch (2002). In their study, the alibi witness was either a neighbor or a girlfriend and the witness was either certain or not certain in making either an identification or non-identification of the defendant as being at his home during the time of the crime. Their results, like those of Lindsay et al., showed that mock jurors were persuaded by the alibi only if the alibi corroborator had no relationship with the defendant.

Clearly, the relationship between the defendant and the alibi corroborator affects the believability of the alibi. But this represents a small start to what seems to us to be a potentially rich literature. It is our intent to facilitate the development of an empirical and conceptual literature on the psychology of alibis, and accordingly, this article has two goals. First, we introduce some basic distinctions, concepts, and ideas that could serve as a framework for the development of an empirical literature on the psychology of alibis. In doing this, we hope to

prime the interest of psychological researchers in the fascinating applied and theoretical possibilities that can accrue from the systematic study of alibis. Second, we propose and test a taxonomy of alibis. We argue that these goals are closely related—a taxonomy of alibis is needed in order to develop a coherent framework for hypotheses about alibis and in order to study alibis empirically.

Some Basic Concepts and Distinctions Regarding Alibis

We make a distinction between *alibi* in the colloquial sense and *alibi* as used in this article. Generally, *alibi* is used to refer to the story, or claim, that is proffered by a suspect in a criminal investigation, and when people say "no alibi" they are referring to an alibi that has no accompanying proof (e.g., many people would consider a plea that the suspect was home alone to be 'no alibi' at all). However, we use the definition of *alibi* from the legal system: "a defense that places the defendant at the relevant time of the crime in a different place than the scene involved and so removed therefrom as to render it impossible for one to be the guilty party" (Nolan, 1990). Hence, being home alone is in fact an alibi because it is claim that places the person at a time and place that, if true, makes it impossible for that person to have committed the crime. From a legal defense perspective, however, the claim itself is only one component of the alibi because one must also have evidence, or proof, that is offered in support of the alibi claim.

Alibis can vary greatly in their surface attributes—the story line of the claim. By surface attributes, we mean the setting (e.g., home, work, public event), activities (e.g., sleeping, eating, playing, working), and other characteristics of the story. But, it is not the surface attributes themselves that make some alibis strong and others weak. Instead, it is the level of proof that can be produced in support of the story. An at-work alibi might be no more persuasive than an at-home alibi if there are no work records or co-worker corroborators to support the alibi. What

distinguishes one alibi from another is the level of proof that can be produced to support the alibi, not the surface attributes of the alibi itself. Although we acknowledge that there are key surface attributes to alibis (e.g., at-home setting versus public event setting) that affect the likelihood or difficulty of having proof, we assert that the important dimension is proof, not the surface attributes themselves.

Defining *alibi* in terms of the proof offered in support of the alibi allows us to discuss features of strong or weak alibis and compare various alibis with each other regardless of their surface attributes. The surface attributes of alibi stories can vary greatly, but the types of evidence offered to prove the alibi stories fall into one of two broad categories: physical evidence and person evidence. Physical evidence might include such items as receipts, videotape recordings, or other physical records that indicate the person was at a particular location at a given time. Person evidence might involve testimony from friends, relatives, or strangers who can corroborate the claim that the alibi provider was at a particular place at a given time[1]. An important quality of any corroborating evidence is that it must solve a time/space problem: The proof must speak to both the time of the crime and the space the alibi provider was occupying at that time. In short, a mere claim that one was at a particular location at a given time is likely to be insufficient; there must be some physical or person evidence to support the claim for an alibi to be considered at all believable.

In calling for a literature on the psychology of alibis, one distinction that we think should be made is between the alibi provider and the alibi itself. Undoubtedly, there are characteristics of alibi providers that make their alibi more or less believable than someone else who gives the same alibi with the same level of proof. The ex-convict who produces an ATM receipt from a distant location from the crime for the time the crime was committed is going to be less

believable than the business owner with no prior record who produces an ATM receipt. It is tempting to think that varying the characteristics of alibi providers is the same as studying alibis. But characteristics of alibi providers should not be confused with characteristics of alibis. The difference between the ex-convict and the businessman has nothing to do with alibis per se. Hence, it is important to make clear that an understanding of the believability of alibis requires us to study the alibis themselves, not the characteristics of the alibi provider.

The concept of alibi proof raises one of the most important questions that we can imagine regarding alibis. Even if innocent people can reconstruct where they were at a particular time, how often can they actually prove it? How often can we expect innocent people to have physical evidence to corroborate their alibis? How often are people alone, rendering them unable to provide person evidence to corroborate their alibi? How often are people around others in public who could not later recognize them so as to corroborate their alibi? How often are people with someone they know, but that person cannot absolutely attest to the time/space issue or remembers the time differently? These questions fall into a domain that we call "alibi generation."

A Proposed Taxonomy of Alibis

We envision a research program on alibi generation in which innocent people would have to account for their whereabouts for specific days and times in the recent past. We are currently collecting such data, but the initial research reported in the current article necessarily begins at a more rudimentary level. The current research is concerned with the development of a taxonomy of alibis.

We argue that a taxonomy of alibis is needed in order to move on to other research questions about alibis. For instance, suppose a researcher wanted to "score" alibis from actual

cases for descriptive purposes. How does one classify alibis for purposes of scoring them? Or, suppose a researcher hypothesized that people with certain demographic characteristics (e.g., unemployed, unmarried) are less likely to have good alibi proof than are people with other demographic characteristics (e.g., employed, married). How does one score such alibis for "goodness" of proof or for qualitative differences in types of alibis? There must be some way to organize the alibis that are generated under varying conditions. Our proposed taxonomy is intended to function as an organizing structure that can be put to such purposes.

We think that the most useful taxonomy is one that sorts alibis according to the extent to which alibi evaluators will accept the alibi, i.e., the extent to which evaluators *believe* that the alibi provider could not have committed the crime. Thus, the underlying psychological construct in our taxonomy is the "believability" of the alibi. Believability seems to be the basic psychological dimension along which all alibi evaluators, from detectives to prosecutors to judges to jurors, make their judgments about the alibi. Although there is no way to have a taxonomy of the "absolute level" of alibi believability (because alibis will interact with other details of a particular case to affect final judgments about the alibi provider), we can examine the *relative believability* of alibis within the taxonomy by holding non-alibi evidence constant. Believability is relatively easy to measure on a continuous scale; most people are accustomed to reporting such beliefs on a percentage scale. Accordingly, the believability construct not only drove our initial construction of the taxonomy but also served as the principal dependent measure for purposes of validating the taxonomy.

Whereas the underlying psychological dimension for our taxonomy is believability, the underlying structure of the taxonomy is based on the two forms of proof that can be used with alibis, namely physical evidence and person evidence. We can think of no other form of

supporting evidence for an alibi other than a person who is brought forward to corroborate the alibi or physical evidence to support the alibi. Because an alibi provider can have physical evidence, person evidence, both physical and person evidence, or neither physical nor person evidence, it seemed natural for us to create a factorial taxonomy with physical evidence and person evidence as the factors.

Another underlying construct in our proposed taxonomy is "perceived ease of fabrication." By perceived ease of fabrication, we mean the evaluator's perception of how easily the alibi provider can manufacture or orchestrate the evidence supporting the alibi. For example, an alibi provider who claims he was watching a television program at the time of the crime might support his alibi by describing events in the television program. However, it would be easy to have taped the program and watched it later or downloaded a program summary from the Internet at a later time. On the other hand, it would be difficult for an alibi provider whose image appears on a dated and timed security video from a store across town to have fabricated the video. Of course, it is possible to have tampered with the store's tapes, but the relative difficulty of fabricating the video should make the alibi proof very strong. We apply the same underlying construct to person corroborators. It would be easier to orchestrate alibi-corroborating testimony from a relative or friend than from a stranger, for instance, because their relationship to the alibi provider could motivate them to fabricate a story favoring the alibi provider.

Based on the ease-of-fabrication construct, we created multiple levels of person and physical evidence for the taxonomy. However, we felt that the ease-of-fabrication construct did not capture all of what needed to be captured for the person evidence dimension. A person who corroborates an alibi might be either lying or mistaken. A complete stranger runs the risk of misidentifying the alibi provider, but has no motive to lie. A person with a close relationship to

the suspect (e.g., spouse), on the other hand, is unlikely to misidentify the alibi provider, but has a potential to lie for the suspect in order to protect him or her. Attribution theories show strong support for the principle of discounting in which the presence of one explanation leads observers to discount another explanation (Gilbert & Malone, 1995). Hence, attribution theory might predict that the focal explanation for the alibi corroborator's claim (i.e., the corroborator actually saw the suspect at the critical time) would be discounted by the possibility of mistaken identification in the stranger case and by lying in the close relationship case. Perhaps the most believable corroborator is one who is familiar with the alibi provider but not motivated to protect or lie for the alibi provider—a non-motivated familiar other. Hence, we thought it would be important for the taxonomy to distinguish between three kinds of alibi corroborators: a motivated familiar other (who is not likely to be mistaken but might lie), a complete stranger (who is not motivated to lie but might be mistaken), and a non-motivated familiar other (who also is not motivated to lie for the person and also is not likely to have mistakenly identified the person).

Overview

We combined three levels of the factor physical evidence (none, easy to fabricate, and difficult to fabricate) and four levels of the factor person evidence (none, motivated familiar other, non-motivated familiar other, non-motivated stranger) into the taxonomy in Figure 1. The taxonomy is organized from the lowest level of predicted believability (upper left of the factorial taxonomy) to the highest level of believability (lower right of the factorial taxonomy).

We created several exemplar alibis for each cell of the taxonomy. Using multiple exemplars for each cell of the taxonomy (e.g., mother, friend, brother representing the motivated familiar other category) was important so as to not confuse the effect of a category-level cell (e.g., motivated familiar other) with a particular instance (e.g., mother) of that category cell (see

Wells & Windschitl, 1999). Participants read a brief description of a crime and then read the alibis of three suspects. Participants were asked to evaluate each alibi as though they were detectives investigating this criminal case. Each alibi was evaluated sequentially before reading the next alibi. Our primary measure was the believability of the alibi, but participants also made judgments of the likelihood of guilt. Participants also evaluated the alibi providers on a number of personality trait dimensions, some of which were relevant to alibi believability and some of which were not.

We expected each of the two factors in the taxonomy, i.e. physical evidence and person evidence, to have main effects on believability. We also expected an interaction. In particular, although we expected to find differences across levels of person evidence when physical evidence was non-existent or easy to fabricate, we expected levels of person evidence to not have an impact at the highest level of physical evidence. We expected the likelihood-of-guilt and relevant trait measures to follow the same general pattern as the believability data, but we did not expect these measures to be as sensitive to the taxonomy levels as the believability measure. We did not expect the traits that were irrelevant to believability to be affected by levels of the taxonomy.

Method

Participants

Participants were 252 students from a large Midwestern university recruited for an experiment titled “Police Detective Reasoning Skills.” Participants earned extra credit in psychology classes for their participation. All participants were treated in accordance with the “Ethical Principles of Psychologists and Code of Conduct” (American Psychological Association, 1992).

Design

This experiment was a 4 (Person Evidence) x 3 (Physical Evidence) mixed factorial design. Presentation of physical evidence was manipulated within subjects, and presentation of person evidence was manipulated between subjects. Because participants viewed three alibis within one level of person evidence, it was necessary to create three exemplars for each category of person evidence to reduce suspicion and to alleviate stimulus sampling concerns (Wells & Windschitl, 1999). For example, the exemplars from the motivated familiar other category were mother, brother, and best friend, while the non-motivated familiar other category included a grocery store cashier, a bookstore clerk, and a taxi driver. This multiple-exemplar approach ensured that a given participant would not encounter the same exemplar of a person corroborator in more than one alibi.

The order of the within-subjects physical evidence was counterbalanced so that each level of physical evidence occurred equally often in the first, second, and third alibis. For instance, one participant might see alibis with no physical evidence, easy-to-fabricate physical evidence, and difficult-to-fabricate physical evidence, in that order. Another participant might see alibis with difficult-to-fabricate physical evidence, no physical evidence, and easy-to-fabricate physical evidence, in that order. Likewise, the order of exemplars was counterbalanced so that each exemplar appeared equally often in the first, second, and third alibis, as well as equally often with each level of physical evidence.

Materials

Stimulus materials. The crime scenario was in the form of a police report consisting of two parts. The initial report contained an account of a police officer's investigation of an armed robbery at a convenience store. A follow-up report (shown on the same page) explained that the

investigation had reached a dead end with multiple suspects, and it had become necessary to evaluate the suspects on the basis of their alibis. There were 36 short alibi reports used in the experiment, three exemplars for each cell of the 4 X 3 taxonomy.

Table 1 gives 4 examples of the 36 alibis. Example A is from the no person evidence and no physical evidence cell, example B is from the motivated familiar-other and easy-to-fabricate physical evidence cell, example C is from the non-motivated familiar other and difficult-to-fabricate physical evidence cell, and example D is from the non-motivated stranger and difficult-to-fabricate physical evidence cell. Some readers might be concerned that the differences between one alibi and another alibi in Table 1 involve more than just the evidence offered because they also differ in the basic story line (e.g., one was eating pizza, one was taking a walk, and so on). However, it is crucial to note that a given story line (e.g., was in a grocery store at the time) was used as the setting for no physical evidence, easy-to-fabricate physical evidence (e.g., cash receipt), and difficult-to-fabricate physical evidence (security video) *equally often across participants*. Hence, although no individual participant evaluated more than one alibi with the same alibi story line, alibi story lines were not confounded with evidence levels. All 36 alibis can be viewed at the following internet address:
http://129.186.143.73/faculty/gwells/alibiSCENARIOsR.html

Dependent measures. For the primary dependent measure, participants rated each alibi on an 11-point Likert-type scale of believability (0 = *I do not believe him at all*, 10 = *I believe him completely*) immediately after reading each alibi (hereafter called the *belief* measure). Participants also rated on an 11-point Likert-type scale the probability that the suspect is the one who committed the crime (0 = *totally unlikely*, 10 = *he is certainly the gunman*) after reading each alibi (hereafter called the *likelihood* measure). This likelihood measure was not considered

a primary measure of believability because it does not focus on the alibi itself; rather, such a judgment could evoke a broad range of considerations, such as the number of other suspects or the absence of other evidence. Although not the primary measure, the likelihood-of-guilt judgment was included to see if it followed the same pattern as the primary measure (belief).

Participants also rated each alibi provider on 20 traits. Twelve of these traits we considered to be relevant to the believability of an alibi provider (*suspicious, cunning, scheming, deceitful, calculating, conniving, honest, intelligent, loyal, sincere, open, trustworthy)* and eight of these traits we considered to be irrelevant to the believability of an alibi provider (*shy, shrewd, caring, curious, funny, friendly, ambitious, content*). Traits were rated on a 7-point Likert-type scale (1 = *does not describe this suspect at all*, 7 = *describes this suspect perfectly*). After participants had read and rated each alibi individually, they were asked an open-ended, exploratory question regarding what made them believe and disbelieve each alibi. The open-ended question was not subjected to formal analyses, but was included merely to assist the current authors' development of future hypotheses about how people evaluate alibis.

Procedure

The crime scenario, consisting of an initial crime report and a follow-up report, was given to all participants upon arrival at the study. Participants were asked to assume the role of a detective and evaluate the alibis of six suspects. Although participants were told that they would evaluate six alibis, participants were given only three alibis to evaluate; this was done to help prevent order effects. If participants believed they were approaching the last of the possible suspects, they may have injected some strategy (e.g., “I didn’t think it was any of the others, so it must be this one”) into their decision, rather than evaluating the alibi itself.

After participants finished evaluating their three alibis, they were fully debriefed,

thanked, and dismissed.

Results

Overview

We expected that participants would differentiate among types of physical evidence and types of person evidence, rating some alibis as more believable than others in ways reflected in the taxonomy as shown in Figure 1. We also hypothesized that difficult-to-fabricate physical evidence would be powerful evidence and would overwhelm all person evidence with which it was combined, resulting in no effect across levels of person evidence for difficult-to-fabricate physical evidence.

Overview of analyses. Our primary measure, believability, was analyzed using a 4 X 3 mixed ANOVA. The results generally followed the pattern expected. As predicted, the interaction and both main effects were statistically significant, so we conducted a set of planned comparisons. First, we examined simple main effects for person evidence at each of the three levels of physical evidence. We found that person evidence had a significant influence only when there was no physical evidence. Then, we examined simple main effects for physical evidence at each level of the person evidence. We found that physical evidence had a significant effect at each level of person evidence except for the non-motivated stranger level of person evidence. Finally, we conducted single degree of freedom significance tests for person evidence when there was no physical evidence and single degree of freedom tests for physical evidence when there was no person evidence. The only single degree of freedom tests for person evidence that were significant were between the non-motivated familiar other condition and the no person evidence condition, between the non-motivated stranger condition and the no person evidence condition, and between the non-motivated stranger condition and the motivated other condition.

The only single degree of freedom tests for physical evidence that were significant were between the difficult-to-fabricate and the no physical evidence conditions and between the easy-to-fabricate and the no physical evidence conditions.

The likelihood-of-guilt and the trait measures were analyzed as 4 X 3 mixed ANOVAs. The likelihood measure and the relevant traits measures yielded both main effects but not the interaction. The irrelevant traits measure yielded a main effect for physical evidence, but no main effect for person evidence and no interaction.

The Belief Measure

Table 2 shows means and standard deviations for the *belief* measure according to condition, and Figure 2 shows the overall pattern of means. For ease of viewing, we switched the order of the non-motivated familiar other and non-motivated stranger conditions to reflect the order of the obtained means rather than the order in the taxonomy. A general 4 X 3 mixed ANOVA revealed a significant interaction between physical evidence and person evidence on how believable evaluators rated alibis, Wilk's $\Lambda = .95$, $F(6, 492) = 2.28$, $MSE = 4.46$, $p < .05$. There was also a significant main effect for person evidence, $F(3, 247) = 2.76$, MSE 5.38, $p < .05$, as well as a significant main effect for physical evidence, Wilk's $\Lambda = .81$, $F(2, 246) = 29.82$, $MSE = 5.45$, $p < .05$.

Planned comparisons were then performed to clarify the relationships among person and physical evidence. The first three simple main effects were tests of whether person evidence affected believability within each level of physical evidence. Results indicated that type of person evidence only affected believability when there was no physical evidence. There was no significant simple main effect of person evidence at the level of easy-to-fabricate evidence $F(3, 247) = .39$, $MSE = 5.38$, $p = .76$, or at the level of difficult-to-fabricate physical evidence $F(3,$

247) = .40, *MSE* = 5.45, $p = .76$. There was, however, a significant simple main effect for person evidence at the level of no physical evidence $F(3, 248) = 8.48$, *MSE* = 4.46, $p < .05$. Contrasts among the levels of person evidence at the level of no physical evidence showed that the non-motivated stranger was considered more believable than no person evidence, $t(251) = 4.90$, $p < .05$, $d = 0.62$, and was also considered more believable than the motivated other, $t(251) = 3.29$, $p < .05$, $d = 0.42$. Also, the non-motivated familiar other was considered more believable than no person evidence $t(251) = 2.77$, $p < .05$, $d = 0.35$.

The next four simple main effects were tests of whether physical evidence affected believability within each level of person evidence. Results indicated that physical evidence affected believability at all levels of person evidence except the non-motivated stranger level. When the alibis included no person evidence, physical evidence made a significant difference in the ratings of alibis, with no physical evidence the weakest and difficult-to-fabricate evidence the strongest, Wilk's $\Lambda = .62$, $F(2, 61) = 18.84$, $p < .05$. Likewise, there was a simple main effect of physical evidence at the level of motivated familiar-other person evidence, Wilk's $\Lambda = .76$, $F(2, 61) = 9.85$, $p < .05$, as well as at the level of non-motivated familiar-other person evidence, Wilk's $\Lambda = .80$, $F(2, 61) = 7.52$, $p < .05$. However, there was no significant simple main effect for physical evidence when a stranger was a corroborator, Wilk's $\Lambda = .96$, $F(2, 60) = 1.18$, $p = .31$.

Two additional single degree of freedom comparisons, these at the level of no person evidence, allowed a closer examination of the differences among types of physical evidence in the absence of person evidence. Easily fabricated physical evidence was rated significantly more believable than no physical evidence, $t(251) = 5.06$, $p < .05$, $d = 0.80$, 95% C.I.: 0.66, 2.64. Likewise, difficult-to-fabricate physical evidence was significantly more believable than no

physical evidence $t(251) = 5.60$, $p < .05$, $d = 0.71$, 95% C.I.: -0.58, 1.62.

The Likelihood Measure

Table 3 shows the means and standard deviations for the *likelihood* question according to condition, and Figure 3 shows the overall pattern of means for the *likelihood* question according to condition. A general 4 X 3 mixed ANOVA revealed no significant interaction between physical evidence and person evidence on judgments of likelihood that the alibi provider was the gunman, Wilk's $\Lambda = .98$, $F(6, 484) = 0.99$, $p = .43$. However, there was a significant main effect for physical evidence Wilk's $\Lambda = .74$, $F(2, 242) = 42.99$, $p < .05$, as well as a significant main effect for person evidence $F(3, 243) = 4.29$, $MSE = 25.57$, $p < .05$.

The Trait Measures

All trait ratings were scored so that higher scores bore a negative connotation. The trait measures were divided into traits that were relevant to believability and traits that were irrelevant to believability. Relevant traits (*honest, deceitful, suspicious, conniving, open, cunning, sincere, scheming, trustworthy, calculating, loyal, intelligent*) were significantly correlated with the *belief* measure ($.13 < r > .35$), while the irrelevant traits (*content, shrewd, friendly, caring, ambitious, curious, funny, shy*) were not significantly correlated with the *belief* measure ($.01 < r > .09$).

Relevant traits. The relevant trait ratings were collapsed for each participant into a single score for each participant. Table 4 shows means and standard deviations for the overall relevant traits rating according to condition. A general 4 X 3 mixed ANOVA revealed no significant interaction between physical evidence and person evidence on the relevant trait ratings of alibi providers, Wilk's $\Lambda = .97$, $F(6, 492) = 1.23$, $p = .29$. However, there was a significant main effect for physical evidence, Wilk's $\Lambda = .83$, $F(2, 246) = 25.47$, $p < .05$, as well as a significant main effect for person evidence, $F(3, 247) = 3.66$, $p < .05$.

Irrelevant traits. The irrelevant trait ratings were collapsed for each participant into a single score for each participant. Table 5 shows means and standard deviations for the overall irrelevant traits rating according to condition. A general 4 X 3 mixed ANOVA revealed no significant interaction between physical and person evidence Wilk's $\Lambda = .97$, $F(6, 490) = 1.40$, $p = .21$. Likewise, there was no main effect for person evidence, $F(3, 246) = 1.94$, $p = .12$. Interestingly, there was a significant main effect for physical evidence, Wilk's $\Lambda = .91$, $F(2, 245) = 12.19$, $p < .05$. Thus, participants rated those alibi providers with stronger physical evidence as funnier, more caring, more friendly, etc.

Discussion

Our proposed taxonomy of alibi believability received some support from participants' alibi evaluations. Main effects for person evidence and for physical evidence indicate that participants in the role of detectives tended to make distinctions among these alibis along the lines we had expected. In addition, the presence of an interaction followed the general idea that difficult-to-fabricate physical evidence overwhelms person evidence. Nevertheless, there were some aspects to the results that were not expected. Our interaction prediction was that the difficult-to-fabricate physical evidence would trump the person evidence, but we had not expected that even the easy-to-fabricate physical evidence was sufficient to render the person evidence irrelevant. Some caution is required in generalizing this conclusion because the impact of a manipulated variable is sensitive to the particular ways in which it is operationalized in a given study. Nevertheless, we were surprised that even easy-to-fabricate physical evidence (e.g., a cash receipt) wiped out the effects of having a person to corroborate the alibi even when that person had no apparent motivation to lie for the suspect.

Another mild surprise was that participants did not find the alibi corroborators to be more

credible when they were non-motivated familiar others than when they were non-motivated strangers. In both cases there was no motive for the alibi corroborator to lie, but the stranger should be more likely than the familiar other to be honestly mistaken. If anything, alibi corroboration from the non-motivated stranger was seen as more credible than corroboration from the non-motivated familiar other (although we remind the reader that this difference was not statistically significant). This suggests that people do not naturally consider mistaken identification to be a potential reason to discount an alibi corroborator, or perhaps they do not consider the chances of mistaken identification to be any higher in the stranger case than in the familiar-other case. The eyewitness identification literature tends to support the view that people are not sophisticated in their understanding of factors that affect mistaken identification (e.g., Wells, 1984). If this is generally true, then perhaps the taxonomy should be simplified by dropping the distinction between non-motivated familiar others and non-motivated strangers. This is one domain, however, where we think that alternative methods of evaluating alibi believability could show support for this distinction in the taxonomy. Specifically, we suspect that cross-examination at the trial level would sharpen the distinction between non-motivated strangers and non-motivated familiar others. A prosecutor, for example, would likely cross-examine the stranger alibi witness along lines suggesting that the stranger was mistaken in the identification, a line of questioning not available when cross-examining the non-motivated familiar other. For this reason, we advise maintaining the taxonomy distinction between non-motivated familiar others and strangers even though it did not exhibit the expected differences in the current study.

As expected, the pattern of responses for likelihood of guilt followed the (inverse) pattern of responses for believability of the alibi. More interesting is the pattern of responses for the

personality trait ratings of the alibi provider. Although we did not have a strong prediction for the trait ratings, we assumed that some traits, such as deceitful and conniving, would be related to alibi believability. We expected that other traits, however, such as caring and funny, should not be affected by alibi proof. A correlational analysis indicated that the twelve relevant traits were each significantly correlated with alibi believability whereas none of the eight irrelevant traits were significantly correlated with alibi believability. The results showed that both relevant and irrelevant traits were affected by the taxonomy manipulations. The effects on irrelevant traits were not as strong as the effects on relevant traits, but we were surprised to find irrelevant traits to be affected at all by alibi proof. In hindsight, this seems not so surprising. A suspect whose alibi is evaluated negatively is perceived as someone who is likely to have committed the crime in question and is, therefore, likely to suffer from a number of negative inferences about his or her character. Serendipitously, perhaps we have discovered an implicit, or at least indirect, method for assessing alibi believability. As a proxy for directly asking people to evaluate the believability of an alibi, they could be asked to make trait inferences about the alibi provider.

We find it interesting that the highest mean believability in any condition in this study was only 7.4 out of a possible 10.0. Generally, it is not advisable to interpret a scale mean relative to the endpoint possibilities. Nevertheless, we think that this tells us something about the incredible skepticism that greets alibi providers. Consider, for instance, that this highest mean (a mere 7.4 out of 10.0) was obtained in a condition in which there was a non-motivated familiar other who corroborated the alibi *and* an extremely strong level of physical evidence supporting the alibi (e.g., security videos from bookstores, taxis, or ATM machines during times that covered the crime). If this level of proof results in a mean believability rating of only 7.4, what kind of supporting evidence would it take to obtain a score an 8 or 9, let alone a 10, in alibi

believability? Is it even possible to score a 10 in alibi believability?

What might these results suggest about the social cognitive processes involved in alibi evaluation? There is an impressive amount of research in social, cognitive, and developmental psychology indicating that comprehension and acceptance occur simultaneously, leading to a bias to accept statements as true and only later in the process can the statement be "unaccepted" (Gilbert, 1991). This comprehension/acceptance process, along with other processes such as anchoring and insufficient adjustment (Koehler, 1991), leads to a general bias of people to be more accepting (believing) of statements than they should be. In other words, an observer's initial representation of a statement is that the statement is true and this serves as an anchor from which insufficient adjustments are made. Given the profound incredulity of alibi evaluators in the face of even the strongest alibis in the current work, we propose that the process of comprehension and acceptance probably begins before the alibi statement is made. Perhaps the mere status of an alibi provider as a suspect leads to an implicit initial comprehension along the lines "This person committed crime X" and the alibi statement ("I was at location Y and, therefore, could not have committed crime X") must overcome this initial representation.

Our primary motivation for developing this taxonomy was to help launch a systematic literature on the psychology of alibis. We think that this taxonomy represents a reasonable starting point for thinking about alibis. In developing this taxonomy, we tried to maintain a degree of simplicity that would permit easy comprehension while serving a heuristic role for generating new hypotheses. We invite others to refine this taxonomy or to propose alternative taxonomies, but we think any useful taxonomy will need to be based on the underlying concept of proof to support the alibi rather than on the surface attributes of the alibi. This is not to say that the surface attributes of the alibi story cannot affect alibi evaluations at all. Consider two

men who claim to have been alone at home at the time of the crime but one says he was smoking crack at the time and the other says he was reading his Bible. Clearly, inferences about the character of the two men based on the surface attributes of their alibi stories are likely to lead to different evaluations. Still, we remind readers that these differences are more at the level of differences in alibi providers, not differences in the alibis themselves.

It would be misleading to suggest that our alibi taxonomy, or any taxonomy, can capture all the dynamics that determine the believability of an alibi. Clearly, the taxonomy assumes that other variables are held constant. These other variables include such things as background characteristics of the alibi provider (e.g., prior record, no prior record) and behavioral cues of the person giving the alibi story (e.g., nervousness, pauses). We also think that alibi evaluators might be especially suspicious of someone who is able to readily provide an alibi for a distant past event when there seems to have been no reason for an innocent person to be able to recall what she or he was doing that day. And, what if the person initially errs in recall ("I was playing golf with friends that afternoon") only to discover later that she recalled the wrong day and has to change the alibi story? These dynamic features are not captured in the taxonomy and we do not think that the taxonomy itself can be easily modified to capture them. Nevertheless, we think that the taxonomy represents a useful structure from which to study these other variables. For example, we suspect that these other variables (e.g., prior record, nervousness, change of alibi story) will affect alibi credibility primarily when the level of proof is weak (upper left cells of the taxonomy) and will matter less when levels of proof are strong (lower right cells of the taxonomy). Hence, the taxonomy should be useful for researchers even when their hypotheses concern alibi variables that are not part of the taxonomy itself.

A big door that we hope we have opened on the study of alibis concerns the ability of

innocent people to generate and prove alibis. Clearly, some innocent people simply will not be able to reconstruct where they were at a particular time in the past. Others might mistakenly recall where they were and then later have to change their alibi story, a behavior that is likely to be met with considerable incredulity. But, even when innocent people can accurately recall where they were, how often could they provide proof at one of the higher levels of our taxonomy? How often would people be stuck in the upper-left cell of the taxonomy (i.e., no person or physical evidence to support the alibi)? We suspect that physical evidence is rather rare and that person evidence, when it exists, will usually come from a motivated familiar other (friend or relative), which is the second-lowest level of believability in the taxonomy. At the same time, we suspect that people overestimate their ability to prove their alibi. For example, they might think that their girlfriend or cousin will somehow be a powerful form of corroboration or that their corroborator will always recall the time frame the same way that they do. In theory, it seems that the alibi defense ought to work with some regularity for innocent people; in practice we worry that it will rarely work because of the difficulty of proving the alibi and the incredible skepticism that seems to accompany alibi evaluations. We challenge psychological researchers to take up some of these important questions.

References

American Psychological Association. (1992). Ethical principles of psychologists and code of conduct. *American Psychologist, 47*, 1597-1611.

Culhane, S. E, & Hosch, H. M. (2002). *An Alibi Witness's Influence on Jurors' Verdicts.* University of Texas-El Paso manuscript under editorial review.

Gilbert, D. T., & Malone, D. S. (1995). The correspondence bias. *Psychological Bulletin, 117*, 21-38.

Gilbert, D. T. (1991). How mental systems believe. *American Psychologist, 46,* 107-119.

Koehler, D. J. (1991). Explanation, imagination, and confidence in judgment. *Psychological Bulletin, 110*, 499-519.

Leippe, M. R. (1985). The influence of eyewitness nonidentifications on mock jurors' judgments of a court case. *Journal of Applied Social Psychology, 15*, 656-672.

Lindsay, R. C. L. Lim, R., Marando, L., & Cully, D. (1986). Mock-juror evaluations of eyewitness testimony: A test of metamemory hypotheses. *Journal of Applied Social Psychology, 15,* 447-459.

Nolan, J. R. (1990). *Black's law dictionary* (6th ed.). St. Paul, MN: West Publishing Co.

McAllister, H. A., & Bregman, N. J. (1989). Juror underutilization of eyewitness nonidentifications: A test of the disconfirmed expectancy explanation. *Journal of Applied Social Psychology, 19,* 20-29.

Wells, G. L. (1984). How adequate is human intuition for judging eyewitness testimony? In G. L. Wells & E. F. Loftus (Eds.), *Eyewitness testimony: Psychological perspectives.* New York: Cambridge University Press.

Wells, G. L., Small, M., Penrod, S., Malpass, R. S., Fulero, S. M. & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads, *Law and Human Behavior, 22*, 603-647.

Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling in social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115-1125.

Footnotes

[1]The term *alibi provider* refers to the suspect or defendant who is being questioned regarding his or her whereabouts at the time of the crime. We call a person whose statements are put forward to support the alibi an *alibi corroborator*. Although an alibi corroborator is "providing an alibi" for the suspect in the colloquial sense, we reserve the term *alibi provider* for the suspect him or herself.

Table 1.

Alibi Examples

Type of Alibi	Example
A. No person evidence No physical evidence	Suspect A. M. initially could not remember where he was between 7:30 and 8:30 on the evening in question. Later in the interview he claimed that he had been out for a walk in his neighborhood on the east side of the city. A. M. has no record of gun ownership.
B. Motivated familiar other Easy to fabricate	Suspect B. L. said he was entertaining his brother in his home on the west side of the city on the evening in question. He said they had ordered pizza and provided a pizza delivery receipt, paid in cash, timed 8:07 pm. A statement from the brother was taken; the brother claimed they had been in the home for the entire time between 7:30 and 8:30 pm. Suspect does not own a gun.
C. Non-motivated familiar other Difficult to fabricate	Suspect C. Z. claimed he had been at a Check-Into-Cash store in the central city between 8:00 and 8:20 pm on the evening in question. He said that he regularly goes to that store to get cash. The teller at the store recognized a picture of the suspect and agreed that he is a regular customer. The teller also indicated that he was there that night. Security camera video from the store showed the suspect in the store between 8:03 and 8:18. Suspect has no history of gun ownership.

Type of Alibi	Example
D. Non-motivated stranger Difficult to fabricate	Suspect C. Z. claimed that he had been at a grocery store in the city between 7:45 and 8:30 pm on the evening in question. He said that he had been grocery shopping, and then stopped at the ATM in the grocery store to withdraw cash. After seeing a picture of the suspect, a cashier at the customer service desk said that she remembered seeing the suspect at the store that night at approximately 8 pm. Camera video from the ATM showed the suspect withdrawing money at 8:26 pm. Suspect has no history of gun ownership.

Table 2.

Mean Alibi Belief as a Function of Evidence

	Physical Evidence			
Person Evidence	None	Easy-to-Fabricate	Difficult-to-Fabricate	Overall Mean
None	4.79 (1.88)	6.44 (2.18)	6.97 (2.22)	6.07 (2.28)
Motivated				
Familiar Other	5.40 (2.20)	6.83 (2.28)	7.19 (2.21)	6.47 (2.35)
Non-Motivated				
Familiar Other	5.83 (2.29)	6.46 (2.39)	7.41 (2.23)	6.57 (2.38)
Non-Motivated				
Stranger	6.63 (2.06)	6.68 (2.43)	7.11 (2.66)	6.81 (2.39)
Overall Mean	5.66 (2.20)	6.60 (2.31)	7.17 (2.33)	

Note: Standard deviations are given in parentheses. Scale range is from 0 to 10. Higher numbers indicate greater belief in the alibi.

Table 3.

Mean Likelihood of Guilt Rating as a Function of Evidence

	Physical Evidence			
Person Evidence	None	Easy-to-Fabricate	Difficult-to-Fabricate	Overall Mean
None	5.41 (1.77)	3.98 (2.37)	3.35 (2.20)	4.25 (2.29)
Motivated				
Familiar Other	4.59 (2.04)	3.32 (2.35)	3.22 (1.98)	3.71 (2.21)
Non-Motivated				
Familiar Other	4.49 (2.19)	3.60 (2.41)	2.66 (2.10)	3.59 (2.35)
Non-Motivated				
Stranger	3.98 (2.16)	3.37 (2.23)	2.81 (2.05)	3.38 (2.19)
Overall Mean	4.62 (2.10)	3.56 (2.34)	3.01 (2.09)	

Note: Standard deviations are given in parentheses. Scale range is from 0 to 10. Higher numbers indicate greater likelihood that the alibi provider is the gunman.

Table 4.

Mean Relevant Traits Rating as a Function of Evidence

	Physical Evidence			
Person Evidence	None	Easy-to-Fabricate	Difficult-to-Fabricate	Overall Mean
None	4.15 (0.71)	3.68 (0.90)	3.58 (0.90)	3.80 (1.11)
Motivated				
Familiar Other	3.90 (0.82)	3.37 (0.95)	3.30 (0.90)	3.52 (1.10)
Non-Motivated				
Familiar Other	3.89 (0.74)	3.61 (0.90)	3.41 (0.74)	3.64 (1.10)
Non-Motivated				
Stranger	3.60 (0.75)	3.47 (0.79)	3.40 (0.86)	3.49 (1.10)
Overall Mean	3.89 (0.78)	3.53 (0.89)	3.42 (0.85)	

Note: Standard deviations are given in parentheses. Scale range is from 0 to 10. Higher numbers indicate greater ratings of the alibi provider as dishonest, suspicious, deceitful, untrustworthy, etc.

Table 5.

Mean Irrelevant Traits Rating as a Function of Evidence

	Physical Evidence			
Person Evidence	None	Easy-to-Fabricate	Difficult-to-Fabricate	Overall Mean
None	4.39 (0.53)	4.21 (0.57)	4.24 (0.61)	4.28 (0.96)
Motivated				
Familiar Other	4.29 (0.62)	3.94 (0.60)	4.00 (0.65)	4.08 (0.96)
Non-Motivated				
Familiar Other	4.26 (0.62)	4.23 (0.64)	4.15 (0.54)	4.21 (0.96)
Non-Motivated				
Stranger	4.29 (0.60)	4.10 (0.58)	4.13 (0.59)	4.18 (0.96)
Overall Mean	4.31 (0.60)	4.12 (0.60)	4.13 (0.60)	

Note: Standard deviations are given in parentheses. Scale range is from 0 to 10. Higher numbers indicate greater ratings of the alibi provider as shy, uncaring, unfriendly, unfunny, etc.

Figure Captions

Figure 1. A taxonomy of alibi believability.

Figure 2. Mean believability rating as a function of corroborating evidence. Bars represent one standard error.

Figure 3. Mean likelihood rating as a function of corroborating evidence. Note: Bars represent one standard error.

Figure 1

	Physical Evidence		
Person Evidence	None	Easy to Fabricate	Difficult to Fabricate
None	Least believable		
Motivated Familiar Other (easy to fabricate)			
Non-Motivated Stranger (difficult to fabricate, but possibly mistaken)			
Non-Motivated Familiar Other (difficult to fabricate, not likely mistaken)			Most believable

Figure 2.

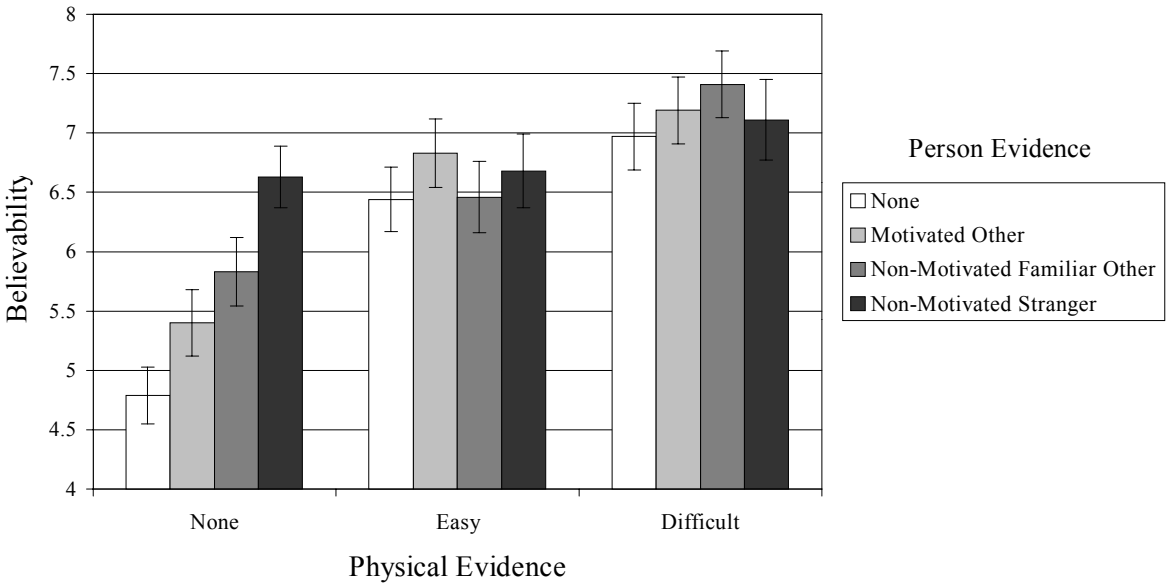

Figure 2. Mean believability rating as a function of corroborating evidence. Bars represent one standard error.

Figure 3.

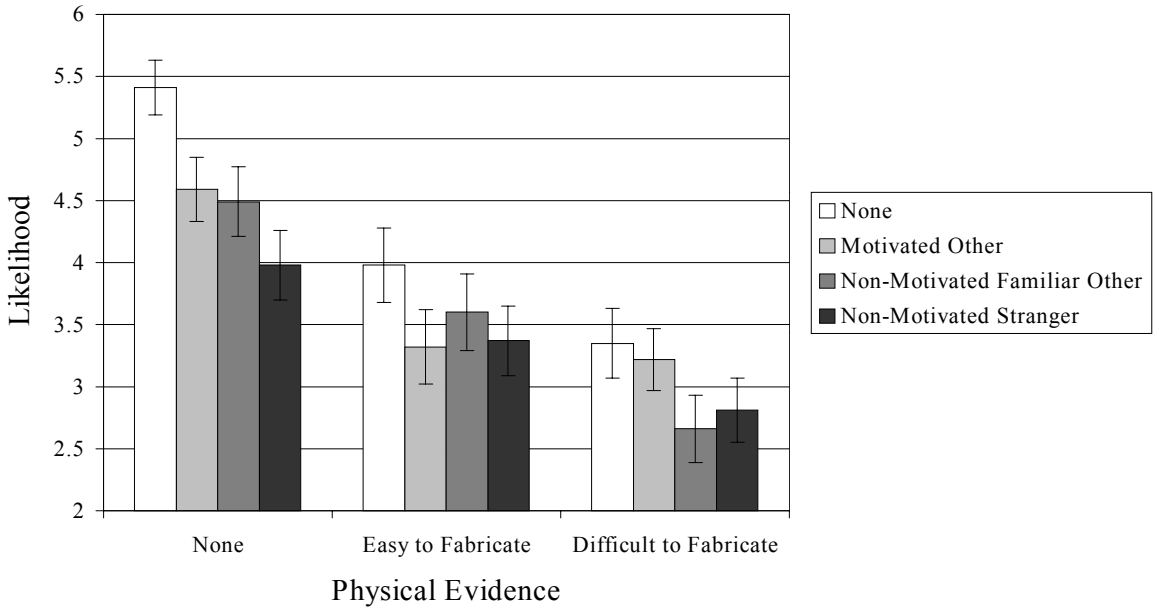

Figure 3. Mean likelihood rating as a function of corroborating evidence. Note: Bars represent one standard error.