‘But can you prove it?’ – examining the quality of innocent suspects’ alibis

Elizabeth A. Olson\textsuperscript{a*} and Steve D. Charman\textsuperscript{b}

\textsuperscript{a}Department of Psychology, University of Wisconsin–Whitewater, Whitewater, USA; \textsuperscript{b}Department of Psychology, Florida International University, Miami, USA

(Received 15 June 2009; final version received 22 June 2010)

Despite an assumption in the legal system that innocent people can generate accurate alibis, little research has examined the process of alibi generation. The current study examined this process with respect to three theoretical reasons why innocent suspects may fail to generate convincing alibis: they may lack the necessary memory, generate mistaken alibis, and generate weak or uncorroborable alibis. Undergraduates ($N=255$) were asked to report four initial alibis – each for a different time – along with corroborating physical and person evidence. Participants attempted to corroborate that evidence before returning 48 hours later. Upon return, participants reported their investigated alibis for the same four time periods. Results indicated that, despite participants’ willingness to generate initial alibis, a substantial proportion of these alibis (36\%) were mistaken, requiring either a change in narrative or a change in corroborating evidence. The majority of investigated alibis relied on evidence that evaluators would consider weak. Distant-past alibis were more likely to be mistaken than near-past alibis. Results indicate that innocent alibi providers may find convincing alibis difficult to generate, and we explain these results within a quantity–accuracy trade-off framework.

Keywords: alibi; alibi generation; autobiographical memory; alibi witness; corroborator

Introduction

Where were you three weeks ago between 8 pm and 9 pm? What about two months ago from 10 am to 11 am? Now, most importantly – can you prove it? What if these questions were asked of you by a police officer who was trying to ascertain whether you are a murderer? Would your alibi be strong enough to convince the officer that you are not a suspect in the case, or would you, an innocent person, be in jeopardy because your alibi was perceived as weak?

Despite the importance these types of questions have with respect to understanding innocent suspects’ alibis, very little psychological research has been devoted to answering them. What empirical research does exist has mostly occurred within what Olson and Wells (2004) have termed the evaluation domain: Investigating how evaluators, such as police investigators and jurors, ascertain the strength of alibis (Allison & Hawkins, 2005; Chong & Dysart, 2009; Culhane, 2005; Culhane & Hosch, 2004; Jolly, Simonds, & Hosch, 2009; Olson & Wells, 2004; Sommers & Douglass,
This domain is important; within the context of a real case, evaluators must classify alibis as being either truthful or not, and the process by which they make these evaluations is largely unknown. Nonetheless, this focus has come at the cost of neglecting what Olson and Wells termed the generation domain of alibis: investigating how alibi providers generate alibis and, perhaps more importantly, the proof to corroborate them (see also Burke, Turtle, & Olson, 2007; Culhane, Hosch, & Kehn, 2008).

Burke et al. (2007) proposed that alibi generation occurs in two phases: the story phase and the validation phase. In the story phase, the alibi provider offers what may colloquially be referred to as an alibi – the narrative account of the alibi provider’s whereabouts at the time in question. The alibi story is very similar to other narrative recollections, such as a story one might tell to a friend or a report of memory in autobiographical memory research. Police investigators, however, do not take an alibi story at face value; they actively investigate and require proof of the story’s veracity. This is what Burke et al. referred to as the validation phase: the period of time during which evidence is gathered to support the alibi story. Olson and Wells (2004) demonstrated that this corroborating evidence determines the perceived strength of an alibi, and can come either in the form of a person corroborator or physical evidence that places the suspect at a location other than the crime scene during the time of the crime.

Although the strategies guilty suspects employ when generating fictional stories are certainly open for investigation, we argue the more compelling area of inquiry within the generation domain is the creation of alibis by innocent people, for it is here where susceptibility to memory error is especially important. The purpose of the current study is to examine whether innocent alibi providers can generate strong, convincing alibis, or whether their alibis will tend to be weak, putting them at risk of prosecution.

How might alibi generation put innocent alibi providers at risk?

There are at least three reasons why innocent alibi providers may fail to provide a convincing alibi: First, in the story phase, alibi providers may lack memory for their whereabouts. Second, it may be discovered in the validation phase that alibi providers accidentally generated a mistaken memory report of their whereabouts. Third, also in the validation phase, the alibi may be unable to be corroborated, or may be only weakly corroborated. Note these three reasons are not mutually exclusive: it is possible, for example, to generate a mistaken alibi which is also weakly corroborated, perhaps by a relative who is uncertain of a specific time but agrees to the alibi. For the sake of the current research, we are more concerned with the unique contributions each of these processes make to unconvincing alibis as opposed to interactions among them, and thus for conceptual clarity, we discuss them as separate processes.

Lack of memory

When initially asked for an alibi story, innocent alibi providers simply may not have a strong memory of their whereabouts during the time period in question. Alibi-providers-to-be are unlikely to recognize the significance of the time for which they
will later have to provide an alibi as it is occurring because only in retrospect will they recognize its significance. It is especially likely that alibi providers will lack a memory of their whereabouts if they were engaged in routine tasks (Brewer, 1988). Recent work suggests that this places alibi-providers in a particularly awkward situation, because (1) memory for an event is likely to be poor if one lacks the motivation to remember it at the time of encoding, (2) motivation to remember an event at the time of retrieval does not improve memory, and (3) evaluators of someone else’s memory erroneously believe motivation to remember at time of retrieval improves memory (Kassam, Gilbert, Swencionis, & Wilson, 2009).

Furthermore, a substantial body of research indicates that latency between encoding an event and the recall of that event degrades one’s memory (see Schacter, 1999). Since the innocent alibi provider might be asked to generate an alibi days, weeks, months, or years after a crime, any episodic memory for the critical time period might have already degraded, even if originally encoded. Although this memory degradation will occur most quickly for weak memories (which likely comprise most innocent alibi providers’ alibis, due to their relative abundance), significant degradation can occur even for events that were highly salient (and thus strongly encoded). For example, flashbulb memories, highly accurate and detailed memories of one’s whereabouts when one initially learned of major public or personal events (Brown & Kulik, 1977), have been shown to degrade over time much like everyday memories (Talarico & Rubin, 2007). A lack of encoding coupled with memory degradation is thus problematic for an innocent alibi provider because this may lead to weak or non-existent alibis, which are not well-believed by alibi evaluators (Olson & Wells, 2004).

**Mistaken memory reports**

A second way in which alibi generation may put innocent alibi providers at risk occurs when alibi providers accidentally provide information that turns out to be mistaken. This mistaken information can come in one of two forms. First, the narrative account of one’s whereabouts may be mistaken. For example, an innocent person may confuse two dates and report having been with a friend at the movies when in fact he was with his parents at home. Second, the basic narrative account of one’s whereabouts may be correct, but the evidence one reports to have may be mistaken. For example, an alibi provider may correctly remember having been shopping during a particular time period, but may mistakenly report having numerous receipts from various stores that he in fact does not have. Because these two types of inaccuracies may be perceived differently (for example, detectives may be more suspicious of a mistaken narrative account than a mistaken report of a piece of evidence that turns out not to exist), we separate them for the purposes of this manuscript. We refer to the former as narrative errors, and to the latter as evidence errors.

Innocent alibi providers may make narrative errors for a number of reasons. Because many innocent alibi providers are likely already starting out with weak and incomplete memories, they may rely on various strategies to reconstruct events (e.g. Brewer, 1996; see Kriati, Goldsmith, & Pansky, 2000). One natural tendency people have when asked to recall past events is to fill in these gaps in their memories with script-consistent details (Fiske & Taylor, 1991; Lampinen, Faries, Neuschatz, &
Toglia, 2000). But because those ‘recalled’ behaviors are based on a behavioral script and not on memory, any deviations from the script that might have actually occurred will not be reported by the alibi provider. For example, a college student who is asked to generate an alibi for a Wednesday two weeks prior at 2:00 pm may not have an explicit memory for his whereabouts but may reason that he normally has class at that time. Consequently, he reports having been in class (a script-consistent behavior). However, if he happened to miss class that day for some reason, and if this error is uncovered through a criminal investigation, the student’s erroneous script-consistent extrapolation is likely to lead evaluators to question his veracity. People are especially likely to rely on scripts (as opposed to episodic memory) when recalling past events after long latencies, making alibi providers especially susceptible to providing erroneous narratives (Eldridge, Barnard, & Bekerian, 1994).

Erroneous narratives may occur for other reasons as well. Autobiographical memory can be error-prone, with participants in memory research often mistakenly ‘recognizing’ plausible but false descriptions of events (Barclay & Wellman, 1986) or combining elements from multiple actual events into an erroneous memory for a single event (known as a memory conjunction error; see Lampinen et al., 2000; Odegard & Lampinen, 2004). As alibi providers attempt to reconstruct the memory of their whereabouts, this combination of elements from previously experienced events may result in a confident reporting of an unintentionally fabricated memory, such as when a store clerk mistakenly reports helping several particular customers in succession, when in fact those customers had come to the store on different days. Consequently, even innocent alibi providers highly motivated to come up with an accurate recollection of their whereabouts during a specific time period may provide erroneous narratives.

Even though narrative errors may be more suspicious to evaluators, even evidence errors may taint the evaluation of any subsequently-generated corrected alibi, as the very fact that the suspect changed some aspect of the alibi may be perceived as highly suspicious. Culhane (2005) presented alibi vignettes to participant-evaluators and asked them to judge the guilt or innocence of the alibi provider. Some of the alibis in the vignettes were reported to change over time with respect to the evidence available to support them: some alibis grew stronger by gaining corroborative evidence and some alibis grew weaker by losing corroborative evidence. The alibis that did not change were rated the most believable and garnered the lowest conviction rates. The alibis that changed in some way, even those that improved, were rated as less believable than the consistent alibis and garnered higher conviction rates. The evidence suggests once a mistaken alibi has been generated, it may be near-impossible for an innocent alibi provider to convince evaluators of his innocence.

Uncorroborated alibis

A third way in which alibi generation may put innocent alibi providers at risk occurs when an alibi cannot be corroborated. An innocent alibi provider who has a strong autobiographical memory for the time period in question, and who reports it accurately, may nonetheless fail to convince evaluators if the accuracy of that alibi cannot be verified independently. An inability to corroborate an accurate alibi can occur for numerous reasons. An alibi might be of such a nature that there is no way
to verify it independently, despite its accuracy. For example, a person who accurately recalls watching television alone for the two hour period during which a crime was committed has no witness and no physical record of his behavior. Additionally, physical evidence that did exist at the time of the occurrence may have been lost in the interim, such as surveillance tapes that were erased and receipts that were lost. Analogous problems exist for alibi witnesses. When a witness to one’s alibi does exist, corroboration requires a match between the alibi provider’s account and the witness’s account. Because a witness may experience the same memory problems as an alibi provider (a weak or non-existent memory, memory conjunction errors, etc.), the witness’s account may be mistaken and will not corroborate the alibi provider’s accurate account. The inability to corroborate one’s alibi may be especially problematic to the extent people underestimate how difficult it is to prove their alibis (Turtle & Burke, 2001). Of course, these problems would be exacerbated by long delays between crime and alibi generation, as physical evidence is more likely to be lost and corroborators are more likely to forget their whereabouts.

Even when an innocent alibi provider can procure physical evidence and/or witnesses to corroborate the alibi, that evidence is often dismissed by evaluators. Olson and Wells (2004) demonstrated how the believability of an alibi varies as a function of the perceived ease of fabrication of physical evidence (such that the easier it is to fabricate, the more it is discounted) and person evidence (such that the greater the perceived motive the corroborator has to lie for the alibi provider, the more it is discounted). However, believable physical evidence (such as appearing on a timed and dated video surveillance tape) may be hard to obtain, even for innocent alibi providers. Additionally, because most people spend much of their time with friends and family, most innocent alibi providers’ alibis can only be corroborated by the very people who are most discounted by alibi evaluators. The likelihood of having a strong alibi that can be verified with person and physical evidence may be quite low.

If innocent people cannot reliably produce alibis that evaluators would consider strong, the legal system may wish to reassess the way it treats alibi evidence. The legal literature currently focuses largely on technical rules regarding the alibi defense (Epstein, 1964; Friedman, 1998; Gooderson, 1977), but not on the qualities of an alibi to which an evaluator should pay attention. This is in contrast to other types of legal evidence studied by researchers (e.g. eyewitness evidence; see Neil v Biggers, 1972; Wells et al., 1998). In general, however, evaluators seem to be highly skeptical of alibis, and tend to disbelieve alibis without solid corroborating evidence (Culhane & Hosch, 2004; Olson & Wells, 2004). In fact, anecdotal reports suggest evaluators may go to great lengths to try to undermine an alibi provider’s alibi (Olson & Wells, 2004). Real-world analyses also suggest that evaluators are highly skeptical of alibis. In an examination of the first 40 cases of wrongful conviction uncovered using DNA evidence, eight (20%) had been convicted in part on the basis of a weak alibi or the absence of an alibi, which was seen as incriminating evidence (Wells et al., 1998). An empirical finding that even known-innocent people cannot provide strong alibis might suggest to the legal system that weak alibis are not as diagnostic of guilt as previously assumed.

The plethora of theoretical reasons for why innocent alibi providers may fail to generate convincing alibis raises many questions: Are they able to remember their whereabouts for a specific time period in the past? Do they produce erroneous alibis
that subsequently change? How strong are their alibis? Are their alibis able to be corroborated? In the only other study to address at least some of these questions, Culhane et al. (2008) asked participants to generate alibis and to report the evidence they thought they could obtain to support them. Critically, however, the authors had no way to assess the accuracy of those reported alibis. Consequently, they were only able to assess errors occurring in the story phase – whether participants lacked a memory for their whereabouts (nearly 11% of their participants were unable to provide a report for their whereabouts for a time two days previous) – and were forced to neglect errors that could occur in the validation phase – whether participants generated mistaken alibis, and whether participants could produce the corroborating evidence necessary to support their alibis.

The current study, in contrast, is aimed at examining all of the potential errors that innocent alibi providers might make when generating their alibis, and is consequently the first study to examine both the story phase and the validation phase of alibi generation. To accomplish this, we had alibi providers generate alibis for time periods in the past, and they then investigated their own alibis and attempted to procure the evidence they claimed could support them. The distinction between the story phase and the validation phase is likely blurred in real criminal investigations – officers may allow suspects to consult a personal calendar while generating their alibi story, for example. Because the current study is more focused on the theoretical reasons why alibi providers may generate mistaken alibis as opposed to mimicking real world practices, we kept the two phases distinct by asking participants to rely solely on their memories for the initial alibi generation and giving them access to potential memory aids later in their investigations. This methodology affords two advantages: First, it allows us to disentangle memory failure (which occurs in the story phase) from a lack of corroborable evidence (which occurs in the validation phase). Allowing participants to look at day planners to generate their alibis, for instance, does not allow us to evaluate the participants’ memories per se of their whereabouts. Second, this methodology allows us to address another question: Is it beneficial to allow innocent alibi providers access to materials when reporting their alibis? An observation of a high proportion of weak alibis that need changing, for instance, would suggest that alibi providers should be allowed to consult external sources. This is especially important to the extent that investigators attempt to pressure suspects into giving alibis immediately, for instance if they believe additional time to investigate one’s alibi would only benefit guilty suspects.

Can innocent people generate strong and accurate alibis? Given the previously stated logic, we hypothesized that a substantial proportion of alibi providers would lack a memory of their whereabouts for the time periods in question and would thus (1) fail to report an alibi at Time 1, or (2) mistakenly provide erroneous narratives and offer non-existent evidence to support their alibis (which would be determined by a large proportion of participants needing to change their alibis and/or the evidence to support them), and that (3) most of the generated alibis would be weak, as assessed by Olson & Wells’s (2004) taxonomy of alibi strength. Given the long psychological history of the effects of latency on memory (Schacter, 1999), we also hypothesized that latency would exacerbate each of these effects.
Method

Participants
Participants were 259 undergraduate students at the initial lab session; four participants did not return for the second lab session, resulting in 255 participants (107 men, 146 women (two participants did not report their sex) at a large Midwestern university (although we did not collect demographic information beyond gender, the sample was similar to the demographic makeup of the university (88% White, mean age = 20 years). Participants were recruited for an experiment titled ‘Police Detective Reasoning Skills’ and agreed to participate in two lab sessions 48 hours apart. They earned extra credit in psychology classes for their participation.

Procedure
During the first lab session, participants received a questionnaire asking them to provide four alibi stories. Participants were asked to write a narrative alibi describing where they were and what they were doing for four two-hour time periods; these alibis are referred to as participants’ initial alibis. In accordance with the theoretical distinction between the story phase and validation phase of alibi generation, participants’ initial alibis were considered part of the story phase and thus participants were not allowed to consult a calendar or any other materials that would provide an external clue to their whereabouts – they were instructed to rely solely on their memories. Two time periods covered 10:00 am until 12:00 noon and 10:00 pm until 12:00 midnight on a fixed date (the Saturday prior to the beginning of the semester), ranging from 6 to 14 weeks prior to the first lab session (the distant past time-delay condition), and two time periods covered 10:00 am until 12:00 noon and 10:00 pm until 12:00 midnight on a date three days prior. (Testing days for the initial alibis were always on Tuesdays and Wednesdays; thus alibis in the near past time-delay condition were also for weekend days of Saturday or Sunday). Participants were asked to report who they had been with via checkboxes listing types of person evidence (relative, friend, acquaintance, and stranger) and to report any kind of physical evidence they thought they might be able to hypothetically produce to police to support each alibi via a fill-in-the-blank question. They were also asked to estimate, using a 0–100 scale, ‘How confident are you that you could produce some evidence (person or physical)?’ All participants were given the option (via a checkbox labeled ‘I don’t remember’) to report that they did not remember where they were or whom they were with.

In accordance with the validation phase of alibi generation, participants were instructed to investigate the four initial alibis they had given over the next 48 hours. They attempted to locate any physical evidence to support their alibis so they could confirm or disconfirm their original assertions about being able to provide physical evidence. They were also asked to contact the people they had reported previously who could corroborate their alibi to confirm or disconfirm their accuracy. They were instructed not to bring their physical or person evidence to the second lab session, but they were provided with a photocopy of their original alibi statements and the questions they would be asked upon their return.

Forty-eight hours after the first session, participants returned to the lab and reported (1) whether their original physical evidence required a change from what
they had reported and what kind of physical evidence they could currently offer, (2) whether their original person evidence required a change from what they had reported and how their corroborator(s) remembered the time in question, and, if needed, (3) a new alibi for the times in question. These are referred to as participants’ investigated alibis. Participants were then debriefed, thanked, and dismissed.

Results

Alibi story coding

Narrative alibi accounts were scored by the first author and a second independent rater, a trained undergraduate research assistant, according to a slightly modified version of Olson and Wells’ (2004) taxonomy of alibi strength (see Table 1). Note the original taxonomy and our modified version focus on the evidence offered to support the alibi rather than the surface characteristics – the narrative details – of the alibi story. An alibi may be supported by either person evidence, i.e. a corroborative account from another individual about the suspect’s whereabouts during the time in question, or physical evidence, i.e. actual physical items that speak to the suspect’s whereabouts during the time in question. The taxonomy distinguishes between four levels of person evidence that vary in terms of the perceived credibility of the corroborator (listed in order of increasing perceived strength): (1) no person evidence, (2) motivated other person evidence (such as a friend or relative), (3) non-motivated other person evidence (such as a stranger), and (4) non-motivated familiar other evidence (such as an acquaintance). The taxonomy used in the current study slightly modified Olson and Wells’ levels of physical evidence such that an alibi could have one of four levels (listed in order of increasing perceived strength): (1) no physical evidence, (2) weak physical evidence, which has no time or place information (such as a CD that was ostensibly burned on a computer during the time in

Table 1. Frequency of investigated alibis (proportions shown in parentheses) as a function of evidence and latency condition.

<table>
<thead>
<tr>
<th>Physical evidence</th>
<th>None</th>
<th>Motivated other</th>
<th>Non-motivated stranger</th>
<th>Non-motivated familiar other</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Distant-past</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>95 (0.19)</td>
<td>282 (0.55)</td>
<td>4 (&lt;0.01)</td>
<td>9 (0.02)</td>
<td>390 (0.76)</td>
</tr>
<tr>
<td>Weak</td>
<td>5 (0.01)</td>
<td>29 (0.06)</td>
<td>0</td>
<td>0</td>
<td>34 (0.07)</td>
</tr>
<tr>
<td>Moderate</td>
<td>2 (&lt;0.01)</td>
<td>39 (0.08)</td>
<td>1 (&lt;0.01)</td>
<td>6 (0.01)</td>
<td>48 (0.09)</td>
</tr>
<tr>
<td>Strong</td>
<td>6 (0.01)</td>
<td>25 (0.05)</td>
<td>1 (&lt;0.01)</td>
<td>6 (0.01)</td>
<td>38 (0.07)</td>
</tr>
<tr>
<td>Total</td>
<td>108 (0.21)</td>
<td>375 (0.74)</td>
<td>6 (0.01)</td>
<td>21 (0.04)</td>
<td>510</td>
</tr>
<tr>
<td><strong>Near-past</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>31 (0.06)</td>
<td>281 (0.55)</td>
<td>2 (&lt;0.01)</td>
<td>12 (0.02)</td>
<td>326 (0.64)</td>
</tr>
<tr>
<td>Weak</td>
<td>5 (0.01)</td>
<td>41 (0.08)</td>
<td>2 (&lt;0.01)</td>
<td>2 (&lt;0.01)</td>
<td>50 (0.10)</td>
</tr>
<tr>
<td>Moderate</td>
<td>3 (&lt;0.01)</td>
<td>73 (0.14)</td>
<td>2 (&lt;0.01)</td>
<td>1 (&lt;0.01)</td>
<td>79 (0.15)</td>
</tr>
<tr>
<td>Strong</td>
<td>4 (&lt;0.01)</td>
<td>39 (0.08)</td>
<td>2 (&lt;0.01)</td>
<td>10 (0.02)</td>
<td>55 (0.11)</td>
</tr>
<tr>
<td>Total</td>
<td>43 (0.08)</td>
<td>434 (0.85)</td>
<td>8 (0.02)</td>
<td>25 (0.05)</td>
<td>510</td>
</tr>
</tbody>
</table>
question), (3) moderate physical evidence, which has time or place information but could not be definitely linked to the participant (such as a phone record or cash receipt), and (4) strong physical evidence, which contains time and place information and could be linked to the specific participant (such as a time card from an employer, a credit card receipt, or an airline ticket stub).

Using the strongest piece of physical evidence and the strongest piece of person evidence offered in the alibi, raters categorized each alibi into one of the 16 cells of the taxonomy. This categorization represents a ‘best case scenario’ for a given alibi. Then each alibi was coded for change between the initial alibi story and the investigated alibi story: no change, narrative change, or one of three types of evidence change (stronger, weaker, or lateral). Alibis were coded as having undergone no change if the story recounted and the evidence offered was consistent from the initial to the investigated version of the story. In accordance with our aforementioned ‘best case scenario’ categorization scheme, we included in this no change category those alibis for which participants were unable to speak to their corroborators since in the best case scenario the corroborators would have agreed with the alibi providers (thus resulting in no change). Thirty-seven participants (15%) reported they could not speak to one of their corroborators for at least one of their alibis. Alibis were coded as having undergone narrative change if the participant reported a different narrative story for their investigated alibis. For example, a participant may have initially reported she had been out to a movie with friends, but after investigation she reported she was driving to campus with her parents. Alibis were coded as having undergone evidence change if the basic narration of their alibis remained consistent, but the evidence reported to support the alibi changed. Alibis that underwent evidence change were coded as having gotten stronger, having gotten weaker, or having changed laterally.

Stronger evidence change could occur in two ways: (1) the participant added a piece of person or physical evidence (e.g. the initial alibi story included no person evidence, but the investigated alibi story included a corroborator) or (2) if the evidence the participant offered improved upon investigation according to the taxonomy (e.g. the initial story included an unspecified receipt, but the investigated story included a credit card receipt). Weaker evidence change could occur in three ways: (1) the participant lost evidence from the initial story to the investigated story (e.g. the initial story included a receipt, but the investigated story indicates the receipt could not be found); (2) if the evidence the participant offered failed to address the critical time frame (e.g. the initial story included an employee time card, but the investigated story reported the time card showed the participant started work after the time he was asked for); (3) if the person corroborator did not remember the event the same way the participant did. Lastly, alibis could undergo a lateral evidence change – some aspect of the evidence offered changed, but the alibi did not move in the taxonomy (e.g. the participant initially reported she had been shopping with a good friend, but on the investigated report stated she had been shopping with a sister). Reliability between the two raters was relatively high for the strength classification of the initial alibi (Cohen’s kappa = 0.85), the strength classification of the investigated alibi (Cohen’s kappa = 0.81), and the determination of how the alibi changed (Cohen’s kappa = 0.76). Disputes between raters were resolved in a discussion session.
Alibi accuracy and strength

Each participant provided four alibis: Two near-past alibis and two distant-past alibis. For each of these latency conditions, participants provided a morning alibi from 10 am to 12 pm and an evening alibi from 10 pm to 12 am. The main analyses treat the time period in question as a stimulus sampling variable (Wells & Windschitl, 1999), and thus collapse across it. (Later analyses explore differences between distant-past and near-past conditions.) The main units of analyses were scores from 0 to 4 calculated for each participant, which represented the number of alibis provided, the number of alibis changed, etc. Results are organized into three sections corresponding to the three ways in which alibi generation may put innocent suspects at risk (i.e. lack of memory, erroneous memory, and lack of corroboration).

Did participants provide an alibi?

Overall, participants were relatively willing to provide an initial alibi for their whereabouts; on average, people reported an initial alibi 88% of the time ($M = 3.54$, $SD = 0.78$). After investigation of their alibis, most participants returned with an alibi story; on average, people reported an alibi 92% of the time ($M = 3.69$, $SD = 0.66$).

Did participants provide mistaken alibi reports?

Because there is no objective way to know whether the participants’ alibis were, in fact, accurate, the accuracy of the alibis must be estimated. This was accomplished by looking at whether participants changed their alibis from the initial alibi to the investigated alibi. Table 2 displays frequencies of alibi change. 11.5% of participants’ alibis ($M = 0.46$, $SD = 0.74$) were objectively wrong and required a narrative change. An additional 24.9% of participants’ alibis ($M = 0.99$, $SD = 0.96$) did not require a narrative change but did require an evidence change (see Table 2).

As previously discussed, alibis that required an evidence change could have gotten stronger, gotten weaker, or changed laterally. It was much more common for alibis to become weaker rather than stronger after investigation – on average, 16.4% of participants’ alibis ($M = 0.65$, $SD = 0.81$) became weaker after investigation, while only 4.5% of participants’ alibis ($M = 0.18$, $SD = 0.41$) became stronger and 4% of participants’ alibis ($M = 0.16$, $SD = 0.41$) changed laterally. Interestingly, alibis became stronger equally often through gaining person evidence (2.3%, $M = 0.09$, $SD = 0.29$) as through gaining physical evidence (2.5%, $M = 0.10$, $SD = 0.31$). On the other hand, alibis became weaker mostly due to a loss of

<table>
<thead>
<tr>
<th>Evidence change</th>
<th>Unchanged</th>
<th>Narrative change</th>
<th>Weaker</th>
<th>Stronger</th>
<th>Lateral</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Latency</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Distant-past</td>
<td>294 (0.58)</td>
<td>94 (0.18)</td>
<td>85 (0.16)</td>
<td>19 (0.04)</td>
<td>18 (0.04)</td>
<td>510</td>
</tr>
<tr>
<td>Near-past</td>
<td>355 (0.70)</td>
<td>23 (0.04)</td>
<td>82 (0.16)</td>
<td>27 (0.06)</td>
<td>23 (0.04)</td>
<td>510</td>
</tr>
<tr>
<td>Total</td>
<td>649 (0.64)</td>
<td>117 (0.11)</td>
<td>167 (0.16)</td>
<td>46 (0.05)</td>
<td>41 (0.04)</td>
<td>1020</td>
</tr>
</tbody>
</table>
person, as opposed to physical, evidence: 11.9% of participants’ alibis ($M = 0.47$, $SD = 0.71$) lost person evidence following investigation, whereas only 5.5% of participants’ alibis ($M = 0.22$, $SD = 0.52$) lost physical evidence.

Overall, then, participants tended to initially overestimate the strength of their alibis, especially with respect to person evidence.

Alibis that become stronger over time should tend to allow investigators to uncover more corroborating evidence, and consequently, they may not be as much of a problem for innocent alibi providers. Alibis that become weaker over time, however, may appear especially suspicious to investigators, as corroborating evidence seems to disappear. Thus, we would argue the most dangerous mistaken alibis for innocent suspects are those that undergo a narrative change after investigation and those that undergo an evidence change specifically with evidence that becomes weaker after investigation. Adding those two types of errors reveals that on average 28% of participants’ alibis ($M = 1.11$, $SD = 0.99$) were ‘dangerously’ mistaken.

**Were alibis able to be corroborated?**

Overall, participants’ investigated alibis were quite weak according to the modified alibi taxonomy. Table 1 displays the absolute number of distant-past and near-past alibis that fell into each of the taxonomy’s categories. The vast majority of participants’ alibis fell into the upper-left quadrant of the taxonomy, indicating they had no or weak physical evidence and no or weak person evidence. Looking at only the two strongest forms of evidence within each category, participants reported that either moderate or strong physical evidence could be found for only 21% of their alibis, and that a non-motivated familiar other or a non-motivated stranger could only corroborate 6% of their alibi. After generating their initial alibi, participants were asked how confident they were that they would be able to find the evidence they reported having for their alibi. Overall, confidence was relatively high ($M = 80.0$, $SD = 36.0$).

**Effects of latency**

Consistent with the general notion that memories deteriorate over time, alibis for distant-past time periods were worse than alibis for near-past time periods. This is evidenced by the fact that participants (1) produced an initial memory for 79% ($M = 1.57$, $SD = 0.75$) of their distant-past alibis but produced an initial memory for 98% ($M = 1.96$, $SD = 0.22$) of their near-past alibis; and (2) provided a narratively mistaken alibi for 18% ($M = 0.37$, $SD = 0.66$) of their distant-past alibis but for only 4.5% ($M = 0.09$, $SD = 0.30$) of their near-past alibis. Participants were not, however, less likely to produce strong corroborative evidence – alibis in the lower right quadrant of the taxonomy – for distant-past (2.7%), as opposed to near-past (2.9%), alibis. This is likely due to floor effects, as there was a very low rate of such alibis even in the near-past condition. Participants were significantly less confident they could produce corroborative evidence for their distant-past alibis ($M = 67.2$, $SD = 29.8$) than they were for their near-past alibis ($M = 88.3$, $SD = 17.7$), $t(236) = 10.57$, $p < 0.001$, $d = 0.89$. 


Discussion

The perception in the legal system that alibis have diagnostic value is critically dependent on the assumption that innocent people should be able to produce relatively strong and accurate alibis; otherwise, alibis would be useless as tools to differentiate the innocent from the guilty. A psychological understanding of autobiographical memory and the variables that might influence it, however, suggest this assumption may be faulty.

We initially identified three ways innocent alibi providers might be at risk: they might lack a memory for their whereabouts, they might provide mistaken narratives or evidence to support those narratives, or they might have an inability to corroborate their accurate memories of their whereabouts. Based on this analysis we predicted that participants, despite being ‘innocent,’ would tend to (1) fail to report an alibi, (2) report mistaken narratives and evidence to support those narratives, or (3) report alibis considered by evaluators to be weak.

Although the latter two hypotheses were supported, the first was not, as evidenced by the relatively low rate at which participants reported they could not remember where they were during the specified times, and virtually all of them returned with an alibi story after investigation. Clearly, participants were willing to report an alibi. However, just because participants reported an alibi does not mean it was correct. Indeed, inaccurate memory seemed to permeate participants’ alibis – in support of the second hypothesis, a substantial percentage (36.4%) of participants’ alibis were mistaken. Some of this inaccuracy can be attributed to mistaken memory on the part of the alibi provider, as evidenced by the number of alibis that had changed their narrative details. Some of the inaccuracy can be attributed to miscalculating the amount of supporting evidence for the alibi; commonly, person corroborators did not remember the event the same as the alibi providers or physical evidence was lost. In support of the third hypothesis, the reported alibis generally suffered from a lack of corroboration. Most alibis had relatively weak physical and person evidence to support them, despite participants’ initial confidence they could adequately corroborate their alibis. Finally, latency harmed participants’ abilities to provide strong alibis by leading to fewer alibis overall, and a greater number of mistaken alibis. In all, the current results do not bode well for the ability of innocent alibi providers to generate accurate alibis, or for the assumption that alibis have high diagnostic value.

A quantity–accuracy trade-off

In a criminal investigation, innocent alibi providers can be at risk in several ways: through the time and hassle of being a suspect in a case, legal fees, and ultimately, incarceration. Without a strong alibi, innocent alibi providers might remain suspects, and alibis discovered to be inaccurate might arouse or increase suspicion on the part of the investigator. But in attempting to generate a strong alibi, innocent alibi providers are faced with a dilemma: should they report as much information as possible, so detectives are maximally able to corroborate their story but be at risk of providing faulty information, or should they report only information they are highly confident is accurate, but at the risk of not providing important details that may
exonerate them? How actual alibi providers choose to resolve this dilemma may have serious consequences.

That very few participants in the current study were unable or unwilling to provide alibi stories, and that a substantial percentage of those that were initially generated turned out to be mistaken, suggests participants were primarily focused on providing a large quantity of information, as opposed to strictly accurate information. In research borrowing from signal-detection theory (e.g. Klätzy & Erdelyi, 1985; Koriat & Goldsmith, 1996b) it has been noted that in autobiographical recall there tends to be a trade-off between the quantity of information provided and the accuracy of the information, such that participants can either provide a larger quantity of information or more accurate information, but not both. Koriat and Goldsmith (1996a,b; Koriat et al., 2000) have expanded upon this, suggesting the quantity-accuracy trade-off is modulated by several other processes. In their monitoring and control framework, Koriat and Goldsmith propose that when people attempt to recount a past event, they (1) engage in a monitoring process, whereby they judge the likelihood that a remembered event is correct, and then (2) engage in a control process, whereby they apply a control threshold, deciding whether or not to report the event by comparing the likelihood of being correct to a criterion. Thus, raising the criterion will lead to less reported information (but what is reported will tend to be of relatively high accuracy), whereas lowering the criterion will lead to more reported information (but much of what is reported will tend to be of relatively low accuracy).

Innocent alibi providers undergo a similar situation to that of participants in memory research: they are asked to free report everything they can remember about events in a certain time and place. Unlike participants in memory research, though, innocent alibi providers may struggle with competing impulses to provide a large quantity of information (because a detailed and complete alibi would seem less suspicious) and to provide only information they are certain is accurate (because inaccuracies would seem more suspicious). To the extent that innocent alibi providers feel pressure to provide a large quantity of information, or even to provide any alibi story at all, they may employ a relatively low control threshold, including information in their alibi that they are less confident is correct. On the other hand, being a suspect in a criminal investigation carries with it the potential for serious consequences, and these high stakes could lead to high accuracy motivation and an increased control threshold. Under these conditions, innocent alibi providers would thus report relatively less information but would only include in the alibi information that they are certain is correct.

In retrospect, then, our failure to support our first hypothesis (that there would be a high rate of failures to report an alibi) was not surprising given support for our second hypothesis (that there would be a high rate of mistaken alibis reported): a high rate of inaccurate alibis is suggestive of a lax control threshold, which would mean people are willing to provide a memory report, even if unsure about its accuracy. That our participants were so willing to provide initial alibi stories suggests that they were operating with a low control threshold, including information in their alibi stories that was potentially incorrect. The sizeable percentage of mistaken alibis also supports the idea that participants were operating with a low control threshold, as changed alibis indicate low initial accuracy. To the extent that real world alibi providers act similarly, the decision to adopt a low control threshold when asked to
provide an alibi may be strategic – innocent alibi providers are probably more concerned with being exonerated as quickly as possible, which can be accomplished by providing large quantities of information the investigator could subsequently verify. Furthermore, it is likely that due to their knowledge that they did not in fact commit the crime in question, innocent alibi providers do not consider the possibility that their alibis could be inaccurate, and do not fully appreciate how suspicious accidental inaccuracies would look. Thus, adopting a low criterion may seem like a desirable strategy, since it may be perceived as having many benefits and few costs.

It is also possible that the large quantity of information provided by alibi providers, at the cost of relatively low accuracy, is due to a natural motivation of innocent alibi providers to produce a coherent story of one’s whereabouts in order to move suspicion away from them. This motivation may result in an overestimation of the accuracy of their memories, by, for example, leading innocent alibi providers to interpret vague memories as being compatible with their reported story. In the language of memory models, this motivation might lead alibi providers to overestimate the likelihood of accuracy of a number of memories, some of which would then surpass the alibi provider’s criterion threshold, leading to a higher quantity of information reported at the cost of relatively low accuracy.

It is worth noting that even alibis that were not shown to be mistaken (i.e. that did not change) tended to lack convincing corroborable evidence to support them. Even with a generous scoring procedure giving participants the ‘best case’ score on their alibis, a majority of participants’ near-past and distant-past alibis were alibi evaluators would consider weak according to Olson and Wells’s (2004) taxonomy of alibi strength. The majority of participants’ alibi corroborators were people who, in the context of a criminal investigation, are unlikely to be believed – family members and friends, an unsurprising finding (see Martin, 1967). Alibi providers may simply not realize how dismissive evaluators are towards this type of person evidence. Physical evidence offered by participants was rare, and most of it was evidence that would not serve as credible proof of participants’ whereabouts. Much like participants in Turtle and Burke’s (2001) study, several participants attempted to offer a day planner, which merely tells where they were supposed to have been, not whether they were actually there. Another participant stated he had been collecting bottle caps in a parking lot, and the bottle caps collected would serve as his physical evidence. These examples demonstrate a general tendency for participants to offer physical evidence without realizing the evidence did not tie them to either the place they had been or the time they had been there. This is consistent with previous findings that alibi providers tend to overestimate the strength of their alibis and the meaningfulness of the corroborative evidence (e.g. Turtle & Burke, 2001). This apparent naivety on the part of alibi providers may be dangerous, as it should tend to lead them to overestimate how believable they are to an alibi evaluator.

**Limitations**

Certainly, actual detectives have greater access to certain types of evidence, such as phone records and surveillance footage, than the average college student. Furthermore, it is possible our participants may not have been sufficiently motivated to put in the effort needed to uncover corroborating evidence. Consequently, real-world suspects’ alibis may be stronger than our participants’ alibis. There are, however, at
least three counter-arguments to be made. First, the rarity with which participants reported alibis that even contained any potentially strong physical evidence suggests the lack of physical evidence is more a function of unavailability rather than inaccessibility. Second, our participants did seem at least somewhat motivated to investigate their alibis. We had a nearly-perfect return rate (only four participants did not return for the second day of the study), and anecdotal reports from participants indicated that many of them were enthusiastic about their investigations. Third, although increased resources may strengthen alibis, it may also weaken them, as it allows for the uncovering of not just corroborative evidence, but also of inconsistent information. In fact, the natural human tendency to look for evidence confirming one's alibi, whereas innocent suspects are especially likely to look for corroboration of one's alibi. Real-world investigation, then, may actually decrease the strength of innocent suspects' alibis compared to the type of investigation in which our participants engaged. Consequently, it is unclear to us whether average alibi strength would become stronger or weaker overall if actual detectives investigated innocent people's alibis.

Nonetheless, the differences in motivation between a research participant and a real life suspect could produce different results. Unfortunately this is, to a certain extent, an inherent limitation of alibi generation research – it is impossible, for ethical reasons, to have participants generate alibis under any sort of threat of the same magnitude as being incarcerated for years. Instead of dismissing laboratory-based alibi research, however, we offer three suggestions. First, the effects of motivation on alibi providers' reports is an open empirical question; future research could manipulate motivation (at least within a limited range), and study its effects on accuracy. Knowing the effects of motivation on alibi reports within a controlled laboratory context may hint at how the alibis of real world suspects differ from the alibis of participants in lab studies. Second, we advocate a position that promotes convergence of methodologies to study alibis, and to that end recommend researchers examine alibis of real-world suspects to estimate their accuracy, and to use those data in conjunction with results from laboratory-based studies. Third, one can acknowledge the shortcomings of laboratory-based alibi research while simultaneously using it to generate theoretical models of alibi generation. Although it is unclear whether the percentages of correct alibis will generalize from this (or any other alibi study) to the real world, the interpretation of the pattern of results – for instance, the trade-off between quantity and accuracy – offers insight into the cognitive processes of actual alibi providers.

In order to understand the cognitive processes of alibi providers, our methodology required some deviations from standard police procedure. For example, we did not allow participants to consult day planners or other evidence when providing their initial alibi stories, a restriction that real-world investigators may not impose on alibi providers. Although this restriction was implemented to maintain the integrity of the theoretical distinction between the story phase and the validation phase of alibi generation, it also means that the inaccuracy rates observed among our alibi providers' initial alibis may not generalize to real alibi providers who would have access to such materials. Although this limitation does not undermine any of the current study's main findings, the alibi literature in general would benefit from research methodologies that more closely mimic real-world police practices.
This research is also a somewhat limited perspective into the kinds of alibis we could expect from the average innocent alibi provider. College students represent one limited demographic group, and future research should address the strength of alibis produced by other innocent individuals drawn from other subpopulations. However, college students should be able to produce relatively good alibis compared to the general population: as a group, students have structured days and weeks, punctuated by semester breaks and a change in schedule, which should provide an excellent framework for autobiographical recall (see Pillemer, Picariello, Law, & Reichman, 1995). Furthermore, research has found memories from one’s early adulthood years to be especially well-remembered throughout the lifespan (the reminiscence bump; Glück & Bluck, 2007), suggesting college students should have strong autobiographical memories relative to the rest of the population. Thus, our student sample might be an above-average demographic group in terms of providing accurate alibis, and we might expect non-college-aged innocent alibi providers would offer even weaker alibis than seen in our sample.

Implications and future directions

The current study is the first empirical psychological study to examine the entire process of alibi generation, and provides the first data to suggest that even innocent people have difficulty generating convincing, accurate alibis. These data undermine the implicit assumption among many members of the legal system that innocent people should be able to produce accurate, corroborable alibis. Rather, innocent people willingly and readily report a significant number of erroneous alibis, but in so doing tend to overestimate the evidence they have to support them, and, in many instances, are forced to change them upon further investigation. From a practical point of view, a weak, non-existent, or changing alibi may not be as incriminating as often thought, and the diagnostic value of alibis should be regarded with some skepticism.

Furthermore, this study provides a theoretical structure regarding innocent alibi generation. First, the distinction between memory failure, mistaken alibi generation, and weak alibi corroborability is a useful conceptualization of alibi generation mistakes; any intervention aimed at increasing the ability of innocent people to produce believable alibis should take all three into consideration. This is important because our data suggest the main problem with alibi providers is not in an inability to produce a story per se, but rather in the accuracy and corroborability of the story. Second, the observed results are consistent with a quantity–accuracy trade-off model, suggesting its usefulness in conceptualizing alibis. No theoretical model has yet been proposed for conceptualizing alibis; the proposed adoption of a quantity–accuracy model should help to organize alibi generation findings and guide future alibi research.

For example, the current study suggests innocent alibi providers may have an especially lax control threshold when providing alibis. This is likely a normal response, as the primary motivation of innocent suspects is to be exonerated as soon as possible, and providing copious amounts of information provides the greatest opportunity for detectives to verify at least some part of their story. Unfortunately, innocent alibi providers’ knowledge of their own innocence likely leads them to fail to recognize the possibility and danger of inaccurate alibis. The downside of
adopting such a lax threshold, consequently, is that the information provided may be inaccurate, thus leading to further suspicion. Although this trade-off between quantity and accuracy of information may be unavoidable, it is probable that various manipulations will be able to sway alibi providers towards one side or the other. For example, the relative weight alibi providers give to information quantity versus information accuracy may be affected by their beliefs about the severity of the consequences that may occur as a result of an erroneous alibi, their perceptions of alibi evaluators’ beliefs in their guilt, the stated importance of providing a lot of information versus accurate information, and a host of other variables. Knowledge of how alibi providers’ strategies and motivations change in response to variables such as these would allow alibi evaluators to more appropriately tailor a situation in which they are attempting to elicit an alibi to suit their purposes (e.g. using alibis from multiple people as an investigative tool to find suspects, or using an alibi from a single suspect as a diagnostic tool to assess the likelihood of that suspect’s guilt). These possibilities await further research.

Results from the current research suggest it should no longer be assumed that innocent alibi providers will be able to generate unchanging, accurate alibis. This unwarranted assumption contradicts both empirical findings and our current theoretical understanding of memory, and as such, legal policy based upon such an assumption would be faulty. That there are a number of people being released from prisons who were serving time for crimes they did not commit is well known (The Innocence Project, 2010), but what may be less well known is that a significant percentage of them were convicted in part on the basis of having had a ‘weak alibi’ (Wells et al., 1998). As both real-world cases and the current results attest, even known-innocent people have trouble generating accurate alibis. Consequently, the diagnostic value of a weak alibi may be minimal, and a proper evaluation of alibis among real-world alibi evaluators should reflect this.

Notes

1. The large amount of variability in the distant-past condition is due to our desire to keep the date of the time period for which participants had to account constant. Thus, participants who participated later in the study had a longer time delay than participants who participated earlier in the study. Because this would confound time delay in the distant past with participants if we examined differences in latency among distant-past alibis (i.e. participants who signed up later in the semester may have differed from participants who signed up earlier in the semester), we examined latency by simply comparing near-past alibis with distant-past alibis, which is a non-confounded comparison due to the within-subjects design.

2. This very likely overestimates the actual accuracy rate of alibis, but is consistent with our overall ‘best case scenario’ treatment of alibis.

3. The percentages equal more than 16.4% because some alibis grew weaker due to a loss of both person and physical evidence.

4. It is likely, however, that college students are less aware than actual investigators of the types of strong physical evidence that could exist to support their alibis. For example, students may not realize just how often they are captured on videotape.

5. It should be noted that if the reader is interested in the alibis provided by people who are able to consult external sources, we do have data that speak to that issue – our participants’ investigated alibis are, essentially, their alibis after having consulted such sources (see Table 1).
References


