2006

Using counterfactuals to assess eyewitnesses' abilities to estimate the effects of external influences on their lineup identifications

Steve Douglas Charman
Iowa State University

Using counterfactuals to assess eyewitnesses' abilities to estimate the effects of external influences on their lineup identifications

by

Steve Douglas Charman

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Psychology

Program of Study Committee:
Gary L. Wells, Major Professor
Veronica Dark
Craig Anderson
Roy Teas
Doug Bonett

Iowa State University

Ames, Iowa

2006

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Steve Douglas Charman

has meet the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

TABLE OF CONTENTS

ABSTRACT

Eyewitnesses are often asked in real-world cases whether their lineup responses were affected by various external influences, but it is unknown whether they can accurately answer these questions. Two experiments used a new paradigm to examine eyewitnesses' abilities to estimate the influence of lineup manipulations. Eyewitnesses were administered either confirming feedback or no feedback (Experiment 1, n = 103), or a cautionary instruction or no cautionary instruction (Experiment 2, n = 114). Eyewitnesses then gave actual responses (retrospective confidence, view, and attention measures in Experiment 1; identification decision in Experiment 2) as well as counterfactual responses stating how they would have responded in the alternative condition. Results across both studies showed an asymmetric estimation of influence pattern: Eyewitnesses who received an influencing manipulation estimated significantly less of a change in their responses than eyewitnesses who did not receive an influencing manipulation. A 48-hr delay between actual and counterfactual responses did not moderate any effects.

INTRODUCTION

Imagine that you are an eyewitness to a crime, brought into a police station to attempt a lineup identification. You identify lineup member #3, after which the lineup administrator tells you "Good! That's our suspect!" You then tell the administrator that you are 90% confident that #3 is, in fact, the criminal. Based on your identification, formal charges are brought against #3. Months later, at trial, the defense attorney for the suspect argues to the judge that, because the lineup administrator told you that you had identified the suspect before you gave a confidence statement, you were unduly influenced and thus your confidence in your identification should be suppressed. The judge, not wanting to completely dismiss your confidence statement, tells the defense attorney that if there is concern over whether you were influenced, that he should simply ask you at trial whether the fact that you were told that you identified the suspect influenced your confidence statement. You are asked this question and you begin to think. What confidence statement would you have given had you not been told that you identified the suspect? Would you have still been 90% confident? If this statement boosted your confidence, by how much did it boost it? Can you accurately report on whether, and how, the lineup administrator's statement influenced you?

Most eyewitness experts, and other psychologists familiar with the shortcomings of human memory and its reconstructive processes, would probably express skepticism about eyewitnesses' abilities to accurately answer these questions. It is, in fact, largely this skepticism that has tended to lead eyewitness researchers to focus on preventing eyewitness errors from occurring in the first place. Relatively little research has been devoted towards developing techniques to separate accurate from inaccurate witnesses after the identification has been made, as such techniques are difficult to come by. Witness variables, for example,

such as confidence generally are not highly related to accuracy, thus limiting their usefulness as criteria for separating accurate from inaccurate witnesses (Sporer, Penrod, Read, & Cutler, 1995). Although decision time has been shown to be a stronger and more reliable postdictor of eyewitness accuracy than confidence (i.e., quick decisions are more likely to be accurate than slower decisions), it is nonetheless not reliable enough to use in the real world (Sporer, 1992; Weber, Brewer, Wells, Semmler, & Yeast, 2003). The lack of criteria that are highly diagnostic of a witness's accuracy means that individual jurors tend not to be able to differentiate accurate from inaccurate witnesses (Wells, Lindsay, & Ferguson, 1979). Thus, to obviate the problem of separating accurate from inaccurate witnesses, most eyewitness research has been aimed at developing techniques that minimize the absolute number of inaccurate witnesses, an approach known as the system-variable approach (Wells, 1978).

System-variable research has been largely successful, identifying many lineup practices that create inaccurate witnesses through unnecessary influence. Ideally, a lineup identification should be based solely on the witness's memory of the criminal. However, many real-world lineup techniques lead (or at least allow) witnesses to incorporate extra-memorial factors in their identifications. For example, lineup administrator-induced expectations about the presence or absence of the criminal within a lineup can dramatically change the likelihood that witnesses will identify someone from a lineup (Malpass & Devine, 1981; Phillips, McAuliff, Kovera, & Cutler, 1999). In order to counter such undue influence, techniques such as the double-blind lineup (in which the administrator does not know the identity of the suspect) have been recommended (Wells et al., 1998). Despite this emphasis among eyewitness researchers to develop lineup techniques that avoid unduly influencing the witness, however, the legal system has been reluctant to incorporate many of these ideas into

its practices. Although this reluctance probably stems from a number of factors, one such cause may be the implicit assumption inherent within the legal system that witnesses can accurately report on how certain variables influenced them. If this assumption is true, then proper lineup techniques that minimize influence on witnesses are not as important; the court can always ask witnesses whether (and how) they were influenced by various variables and thus 'undo' the effects of influence. However, if this assumption is false, then the legal system's trust in witnesses' self-reports of influence is misplaced and, to the extent that it reduces the perceived need for proper lineup reforms, counterproductive.

The legal system's reliance on witnesses' self-reports is pervasive and often implicit. In an important case, *Neil v. Biggers* (1972), the Supreme Court outlined five variables that should be taken into consideration when determining the accuracy of an eyewitness. Three of those variables – how confident the witness is in the identification, how good a view the witness had of the criminal, and how much attention the witness paid to the criminal – rely exclusively on self-reports by the witness. All three of these self-reports have been shown to be affected by various aspects of the lineup procedure (notably, post-identification feedback; Wells & Bradfield, 1998). Thus, the implicit assumption of the legal system is that witnesses can give accurate self-reports by 'undoing' any potential influence created by various aspects of the lineup procedure. To the extent that witnesses cannot undo any such effects, the criteria handed down in *Neil v. Biggers* are inadequate in determining the accuracy of witnesses.

In fact, the general assumption that witnesses can accurately report on how they were influenced (henceforth called the *accurate report assumption*) is actually a collection of more specific assumptions. First, it assumes that witnesses consciously *notice* the influencing

variable in the first place. A variable that is shown to influence a witness may be subtle enough to not even be consciously perceived by witnesses. For example, although experimenter beliefs about the identity of a criminal have been shown to influence witnesses' confidence in their identifications, witnesses often tend to be unaware of having perceived any influencing information (Garrioch & Brimacombe, 2001).

Second, given that witnesses consciously notice the influencing variable, the accurate report assumption assumes that witnesses can *remember*, at a later time, the presence of the influencing variable. Because real-world eyewitnesses are generally asked at trial (usually many months following the identification itself) about the lineup procedure, memory for specific lineup procedures is likely to have degraded. It has been shown that witnesses forget various aspects of a lineup procedure (Rose, Bull, Vrij, 2005) and thus the assumption that eyewitnesses can remember specific lineup procedures over time is questionable.

Third, given that witnesses both consciously noticed and remembered a specific influencing variable, the accurate report assumption assumes that witnesses can *determine* whether a variable had an influence on them. Witnesses may claim that an influencing variable did not influence them, or may claim that a noninfluencing variable did influence them. For example, Wells, Olson, and Charman (2003) either gave witnesses confirming feedback ("Good. You identified the suspect.") or did not give witnesses this feedback following a lineup identification. Although confidence was significantly higher for those witnesses who received confirming feedback, those who later claimed that their confidence was influenced by the feedback were in reality no more influenced than witnesses who later claimed that their confidence was not influenced by the feedback. Thus, witnesses showed no ability to accurately determine whether post-identification feedback influenced them.

Witnesses' determinations about influence may not be based on direct knowledge of the impact of the variable, but rather on implicit theories: their beliefs (accurate or not) about how certain variables affect people in general (Nisbett & Wilson, 1977; Wilson, Centerbar, & Brekke, 2002). Whether witnesses show any ability to determine the consequences of influencing variables apart from these implicit theories is unknown.

Fourth, given that witnesses consciously notice the influencing variable, remember it having been present, and determine that the variable had an influence on them, the accurate report assumption assumes that witnesses can accurately *estimate* the effect of the influencing variable by assessing its direction and magnitude of influence. Although witnesses may accurately report that a certain variable influenced them, they may tend to overestimate or underestimate its influence, or even estimate that it influenced their response in a direction opposite its actual influence. For example, it is commonly thought that giving witnesses the pre-lineup instruction that "the criminal may have changed appearance since the time of the crime" will increase eyewitness accuracy, an assumption that led to its inclusion in a formal set of guidelines concerning the collection of eyewitness evidence (Technical Working Group, 1999). However, this instruction has been shown to reduce eyewitness accuracy (Charman & Wells, in press). To the extent that witnesses have the same implicit theory about the effects of the appearance-change instruction, their estimates of the influence of the instruction will be in a direction opposite that of its actual influence.

A substantial body of literature casts doubt on people's abilities to determine whether their judgments, attitudes, or behaviors were influenced by an extraneous variable, as well as their abilities to correct for any such influence (Petty & Wegener, 1993; Stapel, Martin, & Schwarz, 1998; Wilson & Brekke, 1994). According to many of these researchers, when

people think that they have been influenced (oftentimes referred to as having been "biased" or "contaminated") and attempt to correct for that influence, they tend to use their implicit theories (oftentimes referred to as "lay theories" or "naïve theories") about how the influencing variable would have affected them (Petty & Wegener, 1993; Wegener & Petty, 1997; Wilson et al., 2002; for a different view, see Martin & Stapel, 1998). If these implicit theories do not reflect the actual influence of a variable, people will often undercorrect, overcorrect, or unnecessarily correct (i.e., 'correct' for influence that never existed in the first place) for the perceived influence (Petty & Wegener, 1993; Stapel et al., 1998). Given the general consensus within the literature that people have little ability to accurately correct for influence, there is little reason to believe that eyewitnesses will be able to do so.

Nonetheless, given the lack of research within the eyewitness literature on correctly estimating the influence of lineup manipulations, the current experiments test the validity of the four assumptions that make up the accurate report assumption in order to determine whether witnesses can accurately report on whether, and to what extent, they were influenced by specific variables. In order to test these assumptions, two different influencing variable manipulations were used. The influencing variables were chosen according to two criteria. To determine whether witnesses could accurately report on how they were influenced, it was of course necessary to choose variables that reliably influence witnesses' responses. Thus, the first criterion was that the influencing variables had to have been shown repeatedly in previous literature to have demonstrable influencing effects. Also, for the sake of generalizability, it was important in these studies to assess whether the accuracy of witnesses' reports depended on the type of response that was influenced. It may be, for example, that witnesses can accurately report on how their confidence in their identifications

is influenced by a given variable, but cannot accurately report on how their actual identification choices are influenced. Thus, the second criterion in the selection of influencing variables was that the influencing variables chosen should have their primary effects on different types of responses. The two influencing variables chosen that meet both of these criteria are post-identification feedback and pre-lineup instructions. These are discussed in turn.

It has been shown repeatedly that witnesses who receive confirming feedback following an identification later erroneously report having made their identification with higher confidence than witnesses who did not receive confirming feedback (Bradfield, Wells, & Olson, 2002; Wells & Bradfield, 1998; Wells et al., 2003). The effects of confirming feedback on retrospective confidence (i.e., confidence at the time of the identification) tend to be quite strong, with effect sizes typically around $d = 1$ or greater. However, when asked at the end of these types of studies whether they were influenced by the feedback, those witnesses who said that they were influenced were, on average, no more confident than those witnesses who said that they were not influenced (Wells et al., 2003). Upon first glance, this appears to demonstrate that witnesses have no ability to accurately report on influencing variables. However, there is another interpretation. Due to the between-subjects design of the experiment, differences in confidence were only assessed at the group level (as opposed to the individual witness level). It is possible, then, that although everyone was influenced by the feedback, only those witnesses who said that they were influenced were aware of it. Thus, it is possible that at least some witnesses could accurately report on the influence of confirming feedback. In order to more fully disambiguate witnesses' accuracy in their reports of influencing variables, Experiment 1 examined witnesses' abilities to report on the

influence of confirming feedback on their retrospective confidence. Because other self-reports, such as how good a view witnesses had of the criminal and how much attention they paid to the criminal's face, have also been shown to be affected by feedback (Wells & Bradfield, 1998), these measures were also assessed in Experiment 1.

By definition, post-identification feedback is given to a witness following an identification. It is possible, however, that witnesses may not be aware that manipulations administered after an identification can influence them, but are aware of influencing manipulations that occur during or prior to their identification. Experiment 2 used a manipulation that occurs prior to a witness being shown a lineup and that exerts its main effects on the probability of an identification attempt instead of confidence. It has been shown repeatedly that instructions that imply that the criminal is in the lineup significantly increase the probability of an identification attempt, and instructions that suggest that the criminal may not be in the lineup significantly decrease the probability of an identification attempt (Malpass & Devine, 1981; Steblay, 1997). Although there are a number of variations of such instructions that have been shown to affect the probability of identification attempts, Experiment 2 used the most commonly recommended instruction. Specifically, half the witnesses were given the instruction that "the criminal may or may not be in the lineup" (a cautionary instruction) and the other half were not given this instruction, suggesting that the actual criminal was in the lineup.

Because the specific conditions under which witnesses will be more or less accurate in their reports of influence is unknown, a general test of witnesses' abilities necessitates the strategic sampling of multiple influencing manipulations that vary along multiple dimensions. Note that the influencing manipulations for the two current experiments differ

substantially from each other. First, the two manipulations exert their main effects on different measures: retrospective confidence for Experiment 1, and probability of an identification attempt for Experiment 2. Second, the timing of the two manipulations differs between studies. Administration of post-identification feedback occurs *after* a lineup identification task in Experiment 1, whereas administration of the cautionary instruction occurs *before* a lineup identification task in Experiment 2. Third, the influencing manipulations differ in terms of their effect on the accuracy of witnesses' reports. Whereas the presence of confirming feedback makes witnesses *less* accurate in their confidence assessments, the presence of the cautionary instruction makes witnesses *more* accurate in their identification decisions.

Using various influencing manipulations that differ substantially from one another on important dimensions provides a strong test of whether eyewitnesses can accurately report on how they were influenced. Although the first two specific assumptions of the general *accurate report* assumption (i.e., that witness consciously *notice* and *remember* receiving the influencing variable) are easily tested, the second two assumptions (i.e., that witnesses can *determine* whether a variable influenced them, and can *accurately estimate* the effects of influence) are somewhat more difficult to test. Because the latter two assumptions had never before been adequately addressed in the eyewitness literature, it was necessary to develop a novel paradigm to test them.

The technique employed to test these two assumptions was to have all witnesses give not only an *actual* response to the influence (i.e., retrospective confidence for Experiment 1 and lineup identification decision for Experiment 2), but also a *counterfactual* response. The term 'counterfactual' refers to an alternate version of the past or present that could have

occurred, but did not (Roese & Olson, 1995). Assessing a counterfactual version of the present requires a mutable antecedent that, had it been different, could conceivably have led to an alternate outcome. Mentally mutating an antecedent can therefore affect judgments of causality (Wells & Gavanski, 1989). Interestingly, a number of asymmetries concerning the creation and consequences of counterfactuals have arisen in the counterfactual literature. For instance, negative outcomes are more likely to lead to the generation of counterfactuals than positive outcomes (Landman, 1987; but see also Roese & Olson, 1993); upward counterfactuals (describing an alternate state of events that is better than actuality) are generated more often following failure than following success outcomes (Markman, Gavanski, Sherman, & McMullen, 1995); salient antecedent events are more likely to be mentally mutated than non-salient antecedent events when generating a counterfactual (Landman, 1987); and exceptional events are more likely to be mentally mutated than routine events when generating counterfactuals (Kahneman & Miller, 1986). The preponderance of asymmetries in the counterfactual literature suggests the possibility of asymmetric counterfactual responses in the present experiments.

In the context of the present experiments, a counterfactual response refers to witnesses' reports of how they would have responded had they been in the alternate condition. For example, witnesses who received confirming feedback following an identification and then reported their retrospective confidence gave a counterfactual response of how they think they would have responded to the confidence question had they not received the confirming feedback. Similarly, witnesses who did not receive confirming feedback following an identification and then reported their retrospective confidence gave a counterfactual response of how they think they would have responded to the confidence

question had they received the specific confirming feedback that was given to other witnesses. Thus, all witnesses gave an *actual* response to influence (or noninfluence) and a *counterfactual* response to noninfluence (or influence)[1].

The only previously-used paradigm for assessing witnesses' abilities to accurately report on the effects of influencing variables involved simply comparing the degree of actual influence between those witnesses who claimed they were influenced and those witnesses who claimed they were not influenced (Wells & Bradfield, 1998; Wells et al., 2003). The advantages of the actual/counterfactual paradigm introduced in this work instead of this previous paradigm are threefold. First, by collecting both actual and counterfactual responses from witnesses who receive an influencing manipulation and witnesses who do not receive an influencing manipulation, the paradigm provides an objective way to compare the accuracy of witnesses' counterfactual responses. Logically, if witnesses who receive an influencing manipulation can accurately report on the effect of influence, their mean counterfactual response should match the mean actual response of witnesses who do not receive an influencing manipulation, and vice versa. Although any interpretation of such a comparison can only be made at a group level, and not an individual witness level, this design at least provides an objective standard against which to compare group performance.

Second, instead of using a simple "Yes, I was influenced"/"No, I was not influenced" judgment, the actual/counterfactual paradigm uses continuous, quantitative outcome

---

[1] For economy of language, witnesses in the current experiments will be referred to as having received an influencing manipulation or as having not received an influencing manipulation throughout the manuscript. This may be technically incorrect, as even witnesses who did not receive the influencing manipulation in question nonetheless may have been influenced by the *absence* of that manipulation. What is more precisely meant is that witnesses who received an influencing manipulation received additional information about the identification, which served to reduce witnesses' uncertainty about their responses, whereas witnesses who did not receive an influencing manipulation did not receive this additional information.

measures, and can therefore measure the *degree* to which witnesses claim they were influenced. Thus, this paradigm is a more sensitive measure of witnesses' beliefs about influence. Additionally, although witnesses in the previous paradigm might have accurately reported that they were influenced by a given variable, they might nonetheless have over- or underestimated the degree to which they were influenced. Whereas these inaccuracies are necessarily obscured within the previous paradigm, they are brought to light using the actual/counterfactual paradigm. In fact, the actual/counterfactual paradigm can quantitatively measure at a group level exactly how much witnesses over- or underestimate the effects of influence.

Third, asking witnesses directly whether they were influenced, as was done in the previous paradigm, may have inadvertently biased the witnesses to respond in a certain way. The simple act of asking whether they were influenced may imply to witnesses that the experimenter is expecting them to have been influenced and that they should answer that they were in fact influenced. Alternatively, witnesses may want to resist feeling as if they were manipulated, and consequently may be more likely to respond that they were not influenced. Under the actual/counterfactual paradigm, beliefs of influence are assessed by simply asking witnesses to re-respond to questions under a different antecedent condition. This technique should be less likely to imply that the influencing manipulation should have influenced them, and witnesses should give more accurate responses as a result.

The actual/counterfactual paradigm should provide sensitive comparisons between the actual responses of witnesses who receive an influencing manipulation and the counterfactual responses of witnesses who do not receive an influencing manipulation (and vice versa) to determine the extent to which witnesses can accurately estimate the effect of

influence. There were a number of distinct patterns that such data could take; however, there were four distinct possibilities that seemed most likely given their theoretical grounding. These possibilities are illustrated in the four panels of Figure 1.

First, at one extreme, if witnesses can accurately estimate the effects of influence, we would expect to see a complete cross-over interaction (Panel A: Accurate estimation of influence). Witnesses who receive the influencing variable should be able to completely 'erase' the influence from their minds, and witnesses who do not receive the influencing variable should be able to 'add' the influence to their response. Thus the counterfactual responses of witnesses who receive the influencing manipulation should exactly match the actual responses of witnesses who do not receive the influencing manipulation, and vice versa. This complete cross-over interaction is the strong version of the argument that witnesses can accurately estimate the effects of influence.

Second, at the opposite extreme, witnesses may deny that a variable influenced (or would have influenced) their responses (Panel B: No estimation of influence). This would result in a main effect of the influencing variable, with no main effect of actual versus counterfactual response and with no interaction between the two variables. This possibility has both theoretical grounding and empirical support. Witnesses may perceive attempts to influence them as threats against their freedoms, which, according to psychological reactance theory (Brehm, 1966), should lead them to attempt to re-exert their freedoms. One way to perceive that they have re-exerted their freedom is by denying the effect of influence. Additionally, people generally like to believe that their choices are under their own control, and may consequently tend to dismiss the idea that external influences affected their behavior (Langer, 1975; Taylor & Brown, 1988; Wegner, 2003).

More generally, as previously discussed, people seem to have little or no direct introspective access to higher order cognitive processes (Nisbett & Wilson, 1977). In their classic studies, Nisbett and Wilson found that participants would often claim to have been influenced by variables that did not influence them, and would often claim to have not been influenced by variables that did influence them. It was argued that the participants' claims of influence were based on implicit causal theories about how variables produce certain responses. Participants would thus only accurately estimate the effect of influence to the extent that their implicit theories were correct. However, because these implicit theories are often incorrect, people are generally poor at accurately estimating the effects of influence.

Empirical research from the eyewitness area in particular has suggested that witnesses are unable to estimate the effects of influence. Specifically, witnesses whose identification confidence was influenced by postidentification feedback show little ability to undo the effect of influence (Semmler, Brewer, & Wells, 2004). This conclusion was reached by asking witnesses who had just made an identification and subsequently received experimenter feedback to state their *current* identification confidence (their confidence at that moment) and their *retrospective* identification confidence (their confidence at the time of the identification). Because the influencing variable (experimenter feedback) was given after the lineup task, witnesses whose confidence was inflated by the feedback and who could accurately report on the inflation should have given significantly lower retrospective confidence estimates than current confidence estimates. However, in Study 1 of the Semmler et al. experiment, witnesses' retrospective confidence reports were not significantly different from their current retrospective reports. In fact, witnesses' reports of retrospective confidence were highly correlated with their reports of current confidence ($r = .83$ in Study 1; $r = .80$ in

Study 2), suggesting that witnesses based their retrospective confidence on their current confidence. Thus, it seems that witnesses were not able to undo the influence of the experimenter feedback.

Somewhere between the two extremes (accurate estimation of influence and no estimation of influence) is a third possibility: Witnesses may underestimate the effects of influence (Panel C: Underestimation of influence). In this case, we would expect to see an interaction that is not a complete cross-over. In other words, witnesses who receive the influencing variable should be able to only partially (but not fully) 'undo' the effects of the influencing variable. Similarly, witnesses who do not receive the influencing variable should be able to only partially (but not fully) 'add' the effects of the influencing variable. This possibility is based in part on the anchoring and adjustment heuristic (Epley & Gilovich, 2001; Kahnemann & Tversky, 1972). According to this heuristic, when people estimate a value, they tend to begin with a known value (an 'anchor'), and adjust from that anchor. However, the adjustment tends to be insufficient. Thus, witnesses in the current study may anchor on their actual response and adjust insufficiently, leading to an underestimation of influence. Some evidence for an underestimation of influence was found in the results of Semmler et al. (2004). Although there were no significant differences between witnesses' current and retrospective confidence reports following post-identification feedback in Study 1, witnesses' retrospective confidence reports were slightly (but significantly) lower than their current confidence reports following post-identification feedback in Study 2, suggesting that witnesses postulated that feedback affected their responses, but underestimated the magnitude of that effect. Because an underestimation of influence at least indicates that

witnesses are correctly postulating some effect of the influencing manipulation, it is the weak version of the argument that witnesses can accurately estimate the effects of influence.

The most complex interaction pattern is the fourth possibility, which is that the ability of witnesses to accurately estimate the effects of influence may depend on the presence or absence of the manipulated variable (Panel D: Asymmetric estimation of influence). Specifically, witnesses may more readily 'add' the effects of an influencing variable than 'erase' those effects. This would lead to an asymmetric interaction pattern. This possibility has support from two distinct literatures. First, an asymmetric pattern is consistent with research on the hindsight bias, which states that once an outcome is known, people tend to overestimate the likelihood that they knew that outcome would occur all along (Fischhoff, 1975; Guilbault, Bryant, Brockway, & Posavac, 2004). Because both of the current experiments involve a manipulation that gives more information to witnesses who receive the influencing manipulation (i.e., that they identified the actual criminal in Experiment 1; that the actual criminal may not be in the lineup in Experiment 2), these witnesses may overestimate the likelihood that they would have known that information anyway when generating counterfactual responses. Thus, these witnesses may tend to think that even if they had not received confirming feedback, they would have 'known' that they identified the criminal, or even if they had not received the cautionary instruction, they would have known that the criminal may not be in the lineup. This hindsight bias should mitigate witnesses' ability to accurately 'erase' the effects of the influencing variable, and consequently they should tend to underestimate the influence. Witnesses who do not receive the influencing manipulation, on the other hand, should not exhibit any hindsight bias because they do not receive any extra information until they are asked to generate their counterfactual responses.

According to this possibility, witnesses who do not receive the influencing manipulation should therefore estimate a greater, and more accurate, effect of influence than witnesses who do receive the influencing manipulation.

Second, the asymmetric estimation of influence pattern is also consistent with research on counterfactual thinking. Specifically, it has been found that framing counterfactuals in terms of mental addition, in which people must assess whether an alternative antecedent would produce a greater outcome, has a greater impact than framing counterfactuals in terms of mental subtraction, in which people must assess whether an alternative antecedent would produce a lesser outcome (Dunning & Parpal, 1989). Although the current experiments do not explicitly frame counterfactuals in terms of either mental addition or mental subtraction, it is possible that witnesses may interpret the counterfactual questions in such a way. Witnesses who do not receive the influencing manipulation may estimate how they would have responded had they instead received the influencing manipulation by self-generating an additive frame (e.g., "Had I gotten confirming feedback, how much *more* confident would I be?"). Conversely, witnesses who receive the influencing manipulation may estimate how they would have responded had they not received the influencing manipulation by self-generating a subtractive frame (e.g., "Had I not gotten confirming feedback, how much *less* confident would I be?"). If witnesses do in fact self-generate these frames, witnesses who do not receive an influencing manipulation should estimate a greater effect of influence than witnesses who do receive an influencing manipulation. Thus, although the mechanisms driving the asymmetry are different, research into both the hindsight bias and counterfactual thinking predict that witnesses who do not

receive the influencing manipulation should give greater estimates of the effect of influence than witnesses who do receive the influencing manipulation.

Note that regardless of the witness response measure in question, the logic behind these possible patterns of results holds. Thus, the different dependent variables that were used in the current experiments could each show any of these four patterns. Although patterns of data other than the four listed here are possible, these have the strongest theoretical grounding. It should also be noted that although the reasoning behind each of these four possibilities is independent of the specific type of influence that is manipulated, it is also possible that different patterns of results could occur for different influence manipulations. For example, it is possible to find that witnesses underestimate the influence of confirming feedback but asymmetrically estimate the influence of the cautionary instruction. These potential differences in results could be due to differences in witnesses' implicit theories about how different influencing variables should influence them (Nisbett & Wilson, 1977; Wegener & Petty, 1997). For example, if witnesses have an implicit theory that post-identification feedback influences confidence judgments, then they may be more likely to report that they would have responded differently had they received (or not received) such feedback. However, if witnesses do not have an implicit theory that the cautionary instruction influences the probability of attempting an identification, they may be less likely to report that they would have responded differently. Thus, similar patterns of results across influencing manipulations should not necessarily be expected.

Witnesses' abilities to accurately report on the effects of influencing variables rely heavily on their memories of the effects of the influencing variable. Although memory for a particular response may persist for a substantial period of time, memory for the cognitive

processes that led to that particular response may not. If the accuracy of witnesses' counterfactual responses relies on their abilities to 'retrace' their cognitive processes, time delay should harm the accuracy of their counterfactual responses. Thus, a time delay manipulation between witnesses' actual responses and their counterfactual responses was also included in the design of the current experiments. Some witnesses gave their counterfactual responses immediately following their actual responses whereas other witnesses gave their counterfactual responses 48 hours after their actual responses. The time delay only affected witnesses' counterfactual responses; their actual responses were always given immediately following the mock crime. Including a delay condition allows analyses of witnesses' memories for the influencing variable (which were assessed through recall and recognition questions) as well as of witnesses' memories for their original responses to the lineup. However, because what is of interest is the effect of time delay on witnesses' abilities to retrace their cognitive processes that led to their actual responses, and not the memories for the actual responses themselves, all participants in the current experiments, regardless of delay condition, were shown their actual responses prior to generating their counterfactual responses. Any effect of time delay on counterfactual responses was therefore not confounded with witnesses' memories for their actual responses. If such a delay does, in fact, harm witnesses' abilities to retrace their cognitive processes and consequently their abilities to accurately estimate the effects of influence, then no-delay witnesses should estimate greater effects of influence (as assessed through the interaction) than delay witnesses.

EXPERIMENT 1: POST-IDENTIFICATION FEEDBACK

Method

Participants

Participants were drawn from the research pool of undergraduate psychology students at Iowa State University. All participants received course credit for their participation. One hundred and seventeen participants served as mock witnesses. Fourteen participants' data could not be used for various reasons (eight delay participants did not return 48 hours later to complete the study; data from two participants had to be thrown out due to experimenter error, four participants expressed previous knowledge about the study), leaving 103 participants in the experiment.

Design and Overview

Experiment 1 was a 2 (feedback: confirming vs. none) X 2 (response: actual vs. counterfactual) X 2 (timing of counterfactual response: immediate vs. delayed) mixed factorial design, with feedback and timing of counterfactual response being between-subjects variables, and response being a within-subjects variable. All participants engaged in the following order of tasks: They witnessed a mock crime, attempted an identification of the criminal from a lineup, received either confirming feedback or no feedback about their decision, responded to testimony-relevant questions (concerning their retrospective confidence, their view of the criminal, and the amount of attention paid to the criminal; these are participants' *actual* responses), answered questions about their memory for the feedback they were given (if any) and for their actual responses, and reported how they would have answered the confidence, view, and attention questions had they been in the alternative feedback condition (participants' *counterfactual* responses). Participants in the delay

condition experienced a 48-hour delay before answering questions about their memory for influence and giving their counterfactual responses.

Materials

Mock crime. Participants viewed a mock crime approximately 45 seconds long. The video showed a male fiddling with a bomb on the roof of a campus building. The video focuses on the criminal for a few seconds, after which the criminal realizes he is being watched and escapes down a hallway. The criminal's face is visible on at least three occasions.

Lineups. All participants viewed the same lineup that did not contain the criminal. The lineup was composed of six pictures of males who were similar in appearance to the criminal. Lineup members were shown straight-on from the shoulders up in a simultaneous 2 (down) X 3 (across) display. Each picture measured approximately 2 inches by 3 inches and the lineup was shown to each participant at a distance of approximately two feet.

Procedure

Participants were brought into the lab either individually or in groups of two, but viewed the mock crime and lineups individually. Participants were randomly assigned to both feedback and delay conditions. Upon entering the lab, participants signed a consent form and were told that they would be watching a short video of some people and that we were interested in their impressions of these people. They were led individually to private cubicles and were seated in front of a computer. The experimenter began the mock crime video and left the room. Immediately following the video, the experimenter returned and informed participants (henceforth called 'witnesses') that they were now witnesses to a crime and asked them to attempt to make an identification from a photo lineup. The experimenter

showed witnesses a six-person lineup that did not contain the actual criminal. Criminal-absent lineups were used for two reasons: First, to ensure that all witnesses were inaccurate and thus to eliminate any noise associated with having a mixture of accurate and inaccurate witnesses, and second, to maximize the effect of post-identification feedback, which has been shown to have stronger effects for inaccurate witnesses than accurate witnesses (Bradfield et al., 2002). Because it was important to have all witnesses make an identification, the experimenter instructed witnesses to "select the person who you think planted the bomb on the roof." This instruction has been shown to cause almost all witnesses to identify someone from a lineup (Wells & Bradfield, 1998). If a witness failed to make an identification, the experimenter told the witness "I need you to try and identify the bomber."

Following the lineup identification, the experimenter gave either confirming feedback ("Good, you identified the suspect") or no feedback to witnesses, depending on the specific witness's condition. Witnesses then responded to questions about their retrospective confidence in their identification, how good a view they had of the criminal, and how much attention they paid to the criminal's face (see Appendix A for exact wording of the questions that witnesses received). Confirming feedback has been shown to affect witnesses' responses on all three of these measures (Wells & Bradfield, 1998); these were witnesses' *actual* responses. Witnesses in the no-delay condition immediately proceeded with the rest of the experiment; witnesses in the delay condition were excused at this point and reminded to return 48 hours later. When delay witnesses returned, they followed the exact same procedure as no-delay witnesses.

Witnesses were then given questions concerning their memory for the influence and for their actual responses (see Appendix B for the exact wording of the questions that were

asked of witnesses who received confirming feedback and who did not receive confirming feedback). Specifically, witnesses were asked 1) to give an open-ended response as to what the experimenter had told them immediately following their identification (the cued recall question); 2) to choose, from a list of five options, which statement the experimenter had given them immediately following the identification (the recognition question); 3) to select whether the statement that the experimenter gave them (if they were given a statement) occurred before or after they responded to the confidence question; and 4) to re-respond from memory how they had originally responded to the confidence, view, and attention questions. Once witnesses responded to each question, they were told what the correct response was, and were only then allowed to see the next question.

Participants were then shown both the lineup and their original responses to the confidence, view, and attention questions, were instructed to imagine that they were in the alternative feedback condition, and were asked to respond to those questions as if they had been in the alternative condition (see Appendix C for the exact wording of the instructions that were given to participants who were given confirming feedback and who were not given confirming feedback). These were witnesses' *counterfactual* responses. Following completion of these questions, all participants were debriefed and excused.

Results

Overview

Only two participants (1.9%) did not immediately make an identification from the lineup. However, both participants made an identification after the additional prompt by the experimenter (to "try and identify the bomber"), and thus their data are included in all analyses. To answer the question of whether witnesses are able to accurately report on the effects of confirming post-identification feedback, data analyses are broken down into two main sections. The first section examines witnesses' memories for the feedback and for their actual responses. This is ascertained by examining 1) whether witnesses can recall what sort of feedback, if any, they received (as ascertained through their responses to open-ended questions); 2) whether witnesses can recognize what the experimenter told them immediately following their identification (as ascertained through their responses to multiple-choice questions); and 3) whether witnesses can remember how they originally answered the actual confidence, view, and attention questions. Time delay was also examined as a possible moderator of any effects.

The second section examines witnesses' abilities to determine whether they were influenced by, and to estimate the effects of, confirming feedback. This was done by comparing the counterfactual responses given by witnesses who received confirming feedback to the actual responses made by witnesses who did not receive feedback (and vice versa). These comparisons were used to determine 1) whether witnesses show any ability to accurately estimate the effects of influence, 2) whether the estimation of influence is equal for witnesses who received confirming feedback and witnesses who did not receive confirming feedback, and 3) whether witnesses tended to overestimate, underestimate, or

accurately estimate the effects of confirming feedback. Time delay was also examined as a possible moderator of estimates of influence.

Did witnesses notice and remember the confirming feedback and their actual responses?

Open-ended responses. Witnesses' responses to open-ended questions concerning what they were told immediately following their identification were content-analyzed for whether they reported having received some sort of confirmation about their identification decision. Two raters independently analyzed each witness's response; the two raters agreed on 100% of the responses. Of witnesses who received confirming feedback, 80.8% correctly reported in the open-ended responses that they received some sort of experimenter confirmation of their identification. Witnesses in the delay condition, who answered this question 48-hours after having received feedback, did not significantly differ in their likelihood of reporting having received confirming feedback from witnesses in the no-delay condition (80% vs. 81%, respectively), $\chi^2(1) = .02$, $p = .89$, $-.37 < d < .41$[2]. Two witnesses (3.9%) in the no-feedback condition reported having received some sort of confirmation of their identification from the experimenter.

Multiple choice recognition responses. Overall, 90.3% of witnesses correctly selected from a list of five options what the experimenter had told them immediately following their identification. This did not vary significantly as a function of whether witnesses received feedback, $\chi^2(1) = 1.69$, $p = .19$, $-.13 < d < .65$, or as a function of delay, $\chi^2(1) = 3.38$, $p = .07$, $-.02 < d < .76$, although the latter approached significance with a somewhat greater

[2] In order to use a common metric, all effect sizes throughout the manuscript have been converted to Cohen's *d*, which was defined as a small, medium, and large effect size for $d = .20$, .50, and .80, respectively (Cohen, 1988). In addition, 95% confidence intervals were constructed around all effect sizes.

proportion of witnesses in the no-delay condition (14.8%) selecting an incorrect option than witnesses in the delay condition (4.0%).

Of those witnesses who received confirming feedback, 65.4% recalled having received the feedback *before* answering the question about their confidence. The remaining 34.6% of witnesses incorrectly reported having received the feedback *after* answering the question about their confidence. This did not vary significantly as a function of delay, $\chi^2(1) = .15$, $p = .70$, $-.46 < d < .62$.

Remembered responses. Although not a main focus of the study, the possibility that delay or feedback influenced witnesses' memories for their initial responses was also examined. If, indeed, this is the case, then witnesses' abilities to estimate the effects of confirming feedback may depend on their memories for their own initial responses. However, paired samples t-tests indicated that witnesses' remembered reports did not differ significantly from their actual reports on any of the three variables: $t(102) = 1.00$, $p = .32$, $-.07 < d < .47$ for confidence; $t(102) = .21$, $p = .83$, $-.23 < d < .31$ for view; $t(102) = 1.09$, $p = .28$, $-.05 < d < .49$ for attention. Difference scores were then calculated by subtracting witnesses' actual responses to the confidence, view, and attention measures from their respective remembered responses to these measures. These differences scores equaled the degree to which witnesses misremembered their initial responses. A 2 (delay vs. no delay) by 2 (confirming feedback vs. no confirming feedback) ANOVA was conducted on each one of these difference scores. Delay did not affect the degree to which witnesses misremembered their responses to any of the three measures: $F(1, 99) = .03$, $p = .87$, $-.39 < d < .39$ for confidence; $F(1, 99) = .38$, $p = .54$, $-.27 < d < .51$ for view; $F(1, 99) = .33$, $p = .57$, $-.29 < d < .49$ for attention. Feedback also did not affect the degree to which witnesses misremembered

their responses to any of the three measures: $F(1, 99) = 1.95$, $p = .17$, $-.11 < d < .67$ for confidence; $F(1, 99) = .04$, $p = .84$, $-.39 < d < .39$ for view; $F(1, 99) = 1.35$, $p = .25$, $-.17 < d < .61$ for attention. Likewise, no interaction between delay and feedback existed for any of the three measures: $F(1, 99) = .07$, $p = .80$, $-.33 < d < .45$ for confidence; $F(1, 99) = .04$, $p = .85$, $-.39 < d < .39$ for view; $F(1, 99) = 2.83$, $p = .10$, $-.04 < d < .73$ for attention. Thus, witnesses tended to be quite accurate when remembering their initial responses, a finding that was not moderated by either time delay or feedback.

Did witnesses accurately estimate the effects of confirming feedback?

Confidence, view, and attention scores were all significantly correlated with one another on both the actual measures and the counterfactual measures (see Table 1). Because the feedback effect refers to the distorting effect of feedback across multiple measures, scores on these three feedback measures were averaged to create a composite score. Because confidence could range from 0 - 100 in 10-point increments, the confidence scale was converted to a 10-point scale to match the view and attention scales before the composite scores were calculated. All analyses concerning the feedback effect were thus done on this composite score as well as on each of the individual measures. Means of all scores are displayed in Table 2. Patterns of data are displayed in Figure 2. Because preliminary analyses showed no effect of delay, data was initially collapsed across this variable.

Feedback effect. In order to test whether witnesses can accurately estimate the effects of confirming feedback, it must first be shown that feedback indeed had a significant effect on witnesses' confidence. Based on the wealth of previous research concerning the feedback effect, including a recent meta-analysis (Douglass & Steblay, in press), as well as its demonstrated consistency and magnitude, one-tailed tests were used to assess the feedback

effect. Witnesses who received confirming feedback showed significantly higher composite scores than witnesses who did not receive confirming feedback, $t(101) = 2.68$, $p < .01$, $.14 < d < .92$. Looking at the feedback effect for each individual measure revealed that compared to witnesses who did not receive confirming feedback, witnesses who received confirming feedback reported being significantly more confident, $t(101) = 1.86$, $p = .03$, $.04 < d < .70$; having paid significantly more attention to the criminal, $t(101) = 2.01$, $p = .02$, $.07 < d < .73$; and having had a better view of the criminal, $t(101) = 1.95$, $p = .03$, $.06 < d < .72$. The typical feedback effect was thus replicated in the present study.

Test of estimation of influence. If witnesses show any ability to accurately estimate the effects of feedback, then those witnesses who received feedback should *lower* their counterfactual reports of confidence, view, and attention. Conversely, those witnesses who did not receive feedback should *increase* their counterfactual reports of confidence, view, and attention. Thus, any ability of witnesses to accurately estimate for the effects of influence should appear as a significant response (actual vs. counterfactual) X feedback (confirming vs. none) interaction. A 2 X 2 mixed ANOVA on the composite measure revealed a significant interaction, $F(1, 101) = 187.08$, $p < .001$, $2.38 < d < 3.14$. A significant interaction was also present for each of the measures individually: $F(1, 101) = 137.11$, $p < .001$, $1.98 < d < 2.69$ for confidence; $F(1, 101) = 84.39$, $p < .001$, $1.48 < d < 2.13$ for view; $F(1, 101) = 76.06$, $p < .001$, $1.44 < d < 2.08$ for attention. The presence of significant interactions across all three measures suggests that witnesses estimated that confirming feedback had (or would have had) at least some influence on their responses.

Test of equality of estimation of influence. The previous analysis does not, however, indicate whether witnesses who received feedback estimated the effects of feedback to the

same degree as witnesses who did not receive feedback. The asymmetric estimation of influence hypothesis predicted that witnesses who received the influencing variable (in this experiment, confirming feedback) would estimate less of an effect for influence than witnesses who did not receive the influencing variable. To test this prediction, difference scores were calculated for the composite measure as well as for each of the individual measures. This difference score equaled the magnitude of the estimation of influence that witnesses made with their counterfactual responses. For the purposes of testing the asymmetric estimation of influence hypothesis, what is of interest is not the direction of the estimation of influence, but simply its magnitude. Thus, for witnesses who received feedback, this difference score was calculated as their actual measure score minus their counterfactual measure score. For witnesses who did not receive feedback, this difference score was calculated as their counterfactual measure score minus their actual measure score. Calculating difference scores in this way generated an *absolute difference score* that simply represents the magnitude of the estimation of influence, regardless of whether that estimation was an increase (i.e., for witnesses in the no feedback condition) or a decrease (i.e., for witnesses in the feedback condition) in scores. These scores are displayed in Table 3.

Examining the absolute difference composite scores indicated that, consistent with the predicted asymmetric estimation of influence hypothesis, witnesses who received confirming feedback estimated significantly less of an effect of feedback than did witnesses who did not receive confirming feedback, $t(101) = 4.27$, $p < .001$, $.45 < d < 1.25$. Looking at individual measures, witnesses who received confirming feedback estimated significantly less of an effect of feedback than witnesses who did not receive confirming feedback on the

confidence measure, $t(101) = 4.26$, $p < .001$, $.45 < d < 1.25$; the view measure, $t(101) = 2.79$, $p < .01$, $.17 < d < .95$; and the attention measure, $t(101) = 1.96$, $p = .05$, $.00 < d < .78$.

There are two possible ways in which witnesses who received confirming feedback could have estimated a smaller effect of feedback than witnesses who did not receive confirming feedback. First, witnesses who received confirming feedback might simply have been less likely to acknowledge any effect at all of feedback (i.e., their counterfactual responses would be the same as their actual responses). Alternatively, among those witnesses who did acknowledge an effect of feedback, the estimate of the magnitude of the feedback effect may have been less among witnesses who received feedback than among witnesses who did not receive feedback. Data for these two possibilities are displayed in Table 4 and Table 5, respectively. Consistent with the former interpretation, witnesses who received confirming feedback were significantly less likely to acknowledge any effect of feedback than witnesses who did not receive confirming feedback on the confidence measure, $\chi^2(1) = 3.78$, $p = .05$, $.00 < d < .78$; the view measure, $\chi^2(1) = 9.78$, $p = .002$, $.26 < d < 1.05$; and the attention measure, $\chi^2(1) = 5.51$, $p = .02$, $.08 < d < .86$. Partial support was found for the latter interpretation: Looking only at witnesses who acknowledged an effect of feedback revealed that witnesses who received confirming feedback estimated the magnitude of the effect to be significantly less than witnesses who did not receive confirming feedback on the confidence measure, $t(81) = 3.70$, $p < .001$, $.37 < d < 1.27$, but not on either the view measure, $t(59) = .65$, $p = .52$, $-.35 < d < .69$, or the attention measure, $t(61) = .47$, $p = .64$, $-.38 < d < .62$.

Recall that a substantial proportion of witnesses who received confirming feedback (34.6%) incorrectly reported having received confirming feedback *after* having answered the confidence question. Because there is no way feedback can influence responses that have

already been made, these witnesses should have estimated less of an influence of feedback than witnesses who correctly reported having received confirming feedback *before* having answered the confidence question. In fact, witnesses' estimates of the influence of confirming feedback on their confidence reports did not significantly differ as a function of whether they reported having received the feedback before (M = 13.2) versus after (M = 10.6) having answered the confidence question, $t(50) = .84$, $p = .41$, $-.33 < d < .81$.

Test of accuracy of estimation. The data indicate that witnesses who received confirming feedback estimated the influence of feedback to be significantly less than did witnesses who did not receive confirming feedback, but were either of those estimates accurate? To assess the accuracy of witnesses' estimates of influence, the counterfactual scores of witnesses who received feedback were compared to the actual scores of witnesses who did not receive feedback. Similarly, the counterfactual scores of witnesses who did not receive feedback were compared to the actual scores of witnesses who did receive feedback. If witnesses were accurately estimating for the effects of influence, then these differences should be non-significant.

Using the composite scores as dependent measures revealed that witnesses who received feedback did not significantly over- or underestimate the influence of feedback; their counterfactual composite score did not differ significantly from the actual composite score of witnesses who did not receive feedback, $t(101) = .20$, $p = .84$, $-.35 < d < .43$. Looking at measures individually revealed that witnesses who received feedback did not significantly over- or underestimate the influence of feedback on any of the three measures: $t(101) = .61$, $p = .54$, $-.27 < d < .51$ for confidence; $t(101) = .08$, $p = .94$, $-.37 < d < .41$ for

view; $t(101) = .12$, $p = .90$, $-.37 < d < .41$ for attention. Thus, witnesses who received feedback were relatively accurate in their estimates of the influence of feedback.

In contrast, witnesses who did not receive feedback overestimated the influence of feedback on the composite measure; their counterfactual composite score was significantly higher than the actual composite score of witnesses who did receive feedback, $t(101) = 2.98$, $p < .01$, $.20 < d < .98$. Looking at measures individually revealed that witnesses who did not receive feedback overestimated the influence of feedback on confidence, $t(101) = 3.72$, $p < .001$, $.34 < d < 1.14$. Witnesses did not significantly over- or underestimate the influence of feedback on either the view measure, $t(101) = 1.60$, $p = .11$, $-.07 < d < .71$, or the attention measure, $t(101) = 1.01$, $p = .31$, $-.19 < d < .59$.

Test of a moderating influence of time delay. Because all witnesses, regardless of delay condition, responded to the confidence, view, and attention questions immediately following the administration of feedback, delay could only have affected witnesses' counterfactual responses. Thus, to test the moderating effect of time delay, 2 (timing of counterfactual responses: delayed vs. immediate) X 2 (feedback: confirming vs. none) between-subjects ANOVAs were conducted on the counterfactual responses. Delay had no significant main effects on witnesses' counterfactual composite scores, $F(1, 99) = .74$, $p = .39$, $-.23 < d < .55$, nor on any of the individual measures: $F(1, 99) = .44$, $p = .51$, $-.27 < d < .51$ for confidence; $F(1, 99) = 1.30$, $p = .26$, $-.17 < d < .61$ for view; $F(1, 99) = .03$, $p = .85$, $-.39 < d < .39$ for attention. Delay also did not significantly interact with feedback on witnesses' counterfactual composite scores, $F(1, 99) = .78$, $p = .38$, $-.21 < d < .57$, nor on any of the individual measures: $F(1, 99) = 1.23$, $p = .27$, $-.17 < d < .61$ for confidence; $F(1, 99) = .41$, $p = .52$, $-.27 < d < .51$ for view; $F(1, 99) = .07$, $p = .79$, $-.33 < d < .45$ for attention.

However, the previous analyses did not control for witnesses' initial scores on these measures; in other words, although they demonstrate that delay did not moderate witnesses' counterfactual responses, they do not speak as to whether time delay moderated witnesses' estimates of influence. To address this question, the previous analyses were repeated using witnesses' absolute difference scores as the dependent variables. Similar to the previous results, delay had no significant main effects on witnesses' estimates of the influence of feedback on composite scores, $F(1, 99) = .92$, $p = .34$, $-.19 < d < .59$, nor on any of the individual measures: $F(1, 99) = 1.55$, $p = .22$, $-.15 < d < .63$ for confidence; $F(1, 99) = .04$, $p = .84$, $-.39 < d < .39$ for view; $F(1, 99) = .21$, $p = .65$, $-.29 < d < .49$ for attention. Delay also did not significantly interact with feedback on witnesses' estimates of the influence of feedback on composite scores, $F(1, 99) = .30$, $p = .59$, $-.29 < d < .49$, nor on any of the individual measures: $F(1, 99) = .14$, $p = .71$, $-.33 < d < .45$ for confidence; $F(1, 99) = .03$, $p = .86$, $-.39 < d < .39$ for view; $F(1, 99) = 1.09$, $p = .30$, $-.17 < d < .61$ for attention. Therefore, delay did not significantly moderate any of the witnesses' responses.

Discussion

The purpose of Experiment 1 was to determine whether witnesses could accurately report on the influence of confirming feedback. This was broken down into two sections. The first section tested the assumptions that witnesses noticed, and could remember, the confirming feedback. Results supported these assumptions: The majority of witnesses who received confirming feedback accurately reported having received confirming feedback, effects that were not moderated by a 48-hour time delay.

The second section tested the assumptions that witnesses could determine whether confirming feedback influenced them, and could accurately estimate the influence of confirming feedback. Four patterns of data (accurate estimation, no estimation, underestimation, asymmetric estimation) were thought to be possible. Results of Experiment 1 are clearly consistent with the asymmetric estimation of influence hypothesis. Witnesses who did not receive confirming feedback estimated the influence of confirming feedback to be significantly greater than witnesses who did receive confirming feedback. In other words, generating a counterfactual response in which witnesses *added* the effects of confirming feedback had a greater impact than generating a counterfactual response in which witnesses *subtracted* the effects of confirming feedback. This asymmetry on the confidence measure resulted from the fact that, compared to witnesses who did not receive confirming feedback, witnesses who did receive confirming feedback 1) were less likely to report that confirming feedback had (or would have had) any effect whatsoever, and 2) estimated less of an impact of confirming feedback among those who did make a correction. The asymmetry on the view and attention measures resulted only from the fact that fewer witnesses who received confirming feedback reported that the feedback influenced their responses. A 48-hour time

delay between providing an actual response and providing a counterfactual response had no significant effect on this pattern.

It had been predicted that such an asymmetry would lead witnesses who received confirming feedback to underestimate the impact that confirming feedback had on their responses, and witnesses who did not receive feedback to more accurately estimate the impact that receiving confirming feedback would have had on their responses. Interestingly, the observed results were somewhat different. Witnesses who received confirming feedback more accurately estimated the effect of feedback than witnesses who did not receive confirming feedback. In fact, as a group, the counterfactual responses of witnesses who received confirming feedback were not significantly different from the actual responses of witnesses who did not receive feedback, suggesting that they were objectively accurate in their estimates of the influence of confirming feedback. In contrast, witnesses who did not receive feedback tended, as a group, to overestimate the influence feedback would have had on their retrospective confidence responses (but not on their view or attention responses). Thus, it appears that the ability of witnesses to accurately estimate the influence of confirming feedback is dependent on whether those witnesses actually received feedback or not.

EXPERIMENT 2: PRE-LINEUP INSTRUCTIONS

To test whether this observed asymmetric estimation of influence would generalize across a different influencing manipulation, Experiment 2 used the same paradigm as Experiment 1 using a different influencing variable, namely pre-lineup instructions.

Method

Participants

Participants were drawn from the research pool of undergraduate psychology students at Iowa State University. All participants received course credit for their participation. One hundred and forty-one participants served as mock witnesses. Data from eighteen participants could not be used for various reasons (sixteen delay participants did not return 48 hours later to complete the study; two participants expressed previous knowledge about the study), leaving 123 participants in the experiment.

Design and Overview

Experiment 2 was a 2 (cautionary instruction: given vs. not given) X 2 (response: actual vs. counterfactual) X 2 (timing of counterfactual response: immediate vs. delayed) mixed factorial design, with feedback and timing of counterfactual response being between-subjects variables, and response being a within-subjects variable. All participants engaged in the following order of tasks: They witnessed a mock crime, were given either the cautionary instruction by the experimenter or not, were asked to attempt an identification of the criminal from a lineup (their *actual* identification response), answered questions about their memory for the influence and for their actual responses, and reported how they would have responded to the lineup had they been in the alternative instruction condition (their *counterfactual* identification response). Participants in the delay condition experienced a 48-hour delay before answering questions about their memory for influence and giving their counterfactual responses.

Materials

Mock crime and lineup. Both the video of the mock crime and the lineup were

identical to those used in Experiment 1.

Procedure

Participants were brought into the lab either individually or in groups of two, but viewed the mock crime and lineups individually. Participants were randomly assigned to both instruction and delay conditions. Participants followed the same general procedure as participants in Experiment 1. Following the mock crime, however, participants (now called 'witnesses') were given one of two instructions. Before viewing the lineup, witnesses assigned to the cautionary instruction condition were told to "keep in mind that the actual criminal may or may not be in the lineup. Look at the people in the photo lineup and tell me if you see the criminal." Witnesses in the no cautionary instruction condition were told to "look at the people in the photo lineup and try to identify the criminal you saw in the video." Witnesses could then identify a lineup member or choose to identify no one. They were then asked about their confidence in their decision. Their responses were recorded by the experimenter. No feedback was given to the witnesses following the identification decision. Witnesses in the no-delay condition immediately proceeded with the rest of the experiment; witnesses in the delay condition were excused at this point and reminded to return 48 hours later. When delay witnesses returned, they followed the exact same procedure as no-delay witnesses.

Witnesses were given questions concerning their memory for the influence and for their actual responses (see Appendix D for the exact wording of the questions that were asked of witnesses who received the cautionary instruction and witnesses who did not receive the cautionary instruction). Specifically, witnesses were asked 1) to give an open-ended response as to what the experimenter had instructed them immediately before showing them

the lineup (the recall question); 2) to choose, from a list of five options, which instruction, if any, the experimenter had given them immediately before showing them the lineup (the recognition question); and 3) to re-respond from memory how they had originally responded to the lineup. After witnesses responded to each question, they were told what the correct response was, and were only then allowed to see the next question.

Witnesses were then shown both the lineup and the sheet on which the experimenter had recorded their lineup decision, were instructed to imagine that they were in the alternative instruction condition, and were asked to respond to the lineup as if they had been in the alternative instruction condition (see Appendix E for the exact wording of the instructions that were given to participants who were given the cautionary instruction and who were not given the cautionary instruction). These were witnesses' *counterfactual* responses. Following completion of these questions, all participants were debriefed and excused.

Results

Overview

Results are broken down into two sections and are organized in the same format as they were for Experiment 1. The first section examines witnesses' memories for the pre-lineup instruction and for their actual responses. This is ascertained by examining 1) whether witnesses can recall what pre-lineup instructions, if any, they received (as ascertained through their responses to open-ended questions); 2) whether witnesses can recognize what the experimenter told them immediately prior to the identification attempt (as ascertained through their responses to multiple-choice questions); and 3) whether witnesses can remember how they originally responded to the lineup. Time delay was also examined as a possible moderator of any effects.

The second section examines witnesses' abilities to determine whether they were influenced by, and whether they could accurately estimate, the influence of the cautionary instruction. This was done by comparing the actual responses of witnesses who received the cautionary instruction to the counterfactual responses of witnesses who did not receive the cautionary instruction (and vice versa). The comparisons were used to determine 1) whether witnesses show any ability to accurately estimate the effects of influence, 2) whether estimates of the effect of influence are equal for witnesses who received the cautionary instruction and witnesses who did not receive the cautionary instruction, and 3) whether witnesses tended to overestimate, underestimate, or accurately estimate the influence of pre-lineup instructions. Finally, time delay was examined as a possible moderator of any estimations of influence.

Did witnesses notice and remember the cautionary instruction and their actual responses?

Open-ended responses. Witnesses' responses to open-ended questions concerning what instructions they were given immediately before being shown the lineup were content-analyzed for whether they reported receiving a cautionary instruction that the criminal may not be in the lineup. Two raters independently analyzed each witness's response; the two raters agreed on 100% of the responses. Of witnesses who received the cautionary instruction, 82.5% correctly reported having been instructed that the criminal may not be in the lineup. Witnesses in the delay condition, who answered this question 48-hours after having received the instruction were as likely to report having received a cautionary instruction (74%) as witnesses in the no-delay condition (89%), $\chi^2(1) = 2.35$, $p = .13$, $-.08 < d < .64$. One person (1.7%) in the no instruction condition reported having received an instruction that the criminal may not be in the lineup.

Multiple choice recognition responses. Overall, 84.6% of witnesses correctly selected from a list of five options the instruction that was given to them. This did not vary significantly as a function of either instruction, $\chi^2(1) = .02$, $p = .89$, $-.33 < d < .37$, or delay, $\chi^2(1) = 1.21$, $p = .27$, $-.16 < d < .56$.

Remembered responses. Overall, 98.4% of witnesses correctly remembered their identification decision. This did not vary significantly as a function of delay, $\chi^2(1) = .04$, $p = .84$, $-.32 < d < .40$. Difference scores, which equaled the degree to which witnesses misremembered their confidence, were calculated by subtracting witnesses' reported confidence from their remembered confidence. These difference scores did not differ significantly from zero, indicating that witnesses' reported confidence in their identification did not differ significantly from their subsequent remembered confidence, $t(122) = 1.47$, $p =$

.15, $.22 < d < .32$. In fact, actual confidence and remembered confidence were highly correlated, $r = .98$, $p < .001$. The difference scores did not differ significantly as a function of delay, $t(121) = .85$, $p = .40$, $-.21 < d < .51$ or instructions, $t(121) = .66$, $p = .51$, $-.23 < d < .47$. Thus, witnesses tended to be quite accurate when remembering their initial responses, a finding that was not moderated by time delay or instructions.

Did witnesses accurately estimate the effects of instructions?

Proportions of correct lineup rejections (i.e., "not there" responses) as a function of condition are displayed in Table 6. Patterns of data are displayed in Figure 3.

Instruction effect. In order to test whether witnesses can accurately estimate the effects of the cautionary instruction, it must first be shown that the instruction indeed had a significant effect on witnesses' identification decisions. Consistent with past research, the cautionary instruction to "keep in mind that the criminal may or may not be present in the lineup" led to more correct "not there" responses (47.6%) than an instruction that did not include this phrase (6.6%), $\chi^2(1) = 25.77$, $p < .001$, $.66 < d < 1.41$.

Test of estimation of influence. Given that the previous analysis demonstrated that instructions indeed influenced identification decisions, if witnesses show any ability to accurately estimate the effects of the cautionary instruction, then some proportion of witnesses who received the cautionary instruction and subsequently rejected the lineup (i.e., said "not there") should report that had they instead not received the cautionary instruction they would have made an identification. Conversely, some proportion of witnesses who did not receive the cautionary instruction and made an identification should report that had they instead received the cautionary instruction they would have rejected the lineup. Therefore, any tendency to accurately estimate the effects of the cautionary instruction should appear as

a significant response (actual vs. counterfactual) X instruction (cautionary vs. no cautionary) interaction. Because all identifications were incorrect, no distinction was made between witnesses who identified different lineup members; all witnesses were classified as either having made an identification or having rejected the lineup. Consistent with the idea that witnesses were at least partially estimating for the effects of the cautionary instruction, this interaction was significant, $z = 8.09$, $p < .001$.

Test of equality of estimation of influence. The asymmetric estimation of influence hypothesis predicted that witnesses who receive a manipulation (in this experiment, a cautionary instruction) should estimate less of an influence than should witnesses who do not receive a manipulation. To test this prediction, a chi-square test was performed on the percentage of witnesses who gave a counterfactual response that was different from their actual response and in the direction of actual influence. Consistent with the predicted asymmetric estimation of influence hypothesis, witnesses who received the cautionary instruction were less likely to change their response (30.2%) than witnesses who did not receive the cautionary instruction (46.7%), although this effect fell slightly short of attaining statistical significance, $\chi^2(1) = 3.55$, $p = .06$, $-.01 < d < .70$.

Test of accuracy of estimation. If witnesses can accurately estimate the effects of the cautionary instruction, then the proportion of witnesses who received the cautionary instruction and rejected the lineup should not differ significantly from the proportion of witnesses who did not receive the cautionary instruction, but report that they would have rejected the lineup had they received the cautionary instruction. Similarly, the proportion of witnesses who did not receive the cautionary instruction and rejected the lineup should not differ significantly from the proportion of people who received the cautionary instruction, but

report that they would have rejected the lineup had they not received the cautionary instruction. In fact, witnesses who received the cautionary instruction underestimated the influence of the instruction; they were more likely than witnesses who did not receive the cautionary instruction to report that they would have rejected the lineup (20.6% vs. 6.7%, respectively), $\chi^2(1) = 5.03$, $p = .03$, $.05 < d < .77$. In contrast, those witnesses who did not receive the cautionary instruction did not significantly over- or underestimate the influence of the cautionary instruction; they were as likely as witnesses who actually received the cautionary instruction to report that they would have rejected the lineup (53.3% vs. 47.6%, respectively), $\chi^2(1) = .40$, $p = .53$, $-.23 < d < .47$.

Test of a moderating influence of time delay. Because all witnesses, regardless of delay condition, responded to the lineup immediately on Day 1, delay could only have affected witnesses' counterfactual responses. Thus, to determine whether delay significantly affected witnesses' counterfactual responses, a 2 (counterfactual responses: delayed vs. not delayed) X 2 (cautionary instruction: given vs. not given) chi-square test was conducted on the counterfactual responses. Delay did not significantly affect the proportion of witnesses who reported that they would have rejected the lineup, $\chi^2(1) = .37$, $p = .54$, $-.26 < d < .46$, nor did it significantly interact with instruction, $z = 1.29$, $p = .10$. Delay also had no significant effect on the likelihood that witnesses would correctly change their responses, $\chi^2(1) = .01$, $p = .93$, $-.34 < d < .38$. Therefore, time delay did not significantly moderate any of the witnesses' responses.

Discussion

The purpose of Experiment 2 was to determine whether witnesses could accurately report on the influence of the cautionary instruction, and more specifically, whether the results of Experiment 1 – in particular the asymmetric estimation of influence pattern – would generalize to a pre-lineup instruction manipulation. This question was broken down into two sections. The first section tested the assumptions that witnesses noticed, and could remember, the cautionary instruction and their initial lineup responses. Results supported these assumptions; the majority of witnesses who received the cautionary instruction accurately reported having received an instruction that the criminal may or may not be in the lineup, an effect that was not moderated by a 48-hour time delay.

The second section tested the assumptions that witnesses could determine whether the cautionary instruction influenced them, and could accurately estimate the influence of the cautionary instruction. Results of Experiment 2 are consistent with the asymmetric estimation of influence hypothesis. Witnesses who received the cautionary instruction estimated significantly less of an effect of the instruction than did witnesses who did not receive the cautionary instruction. In other words, generating a counterfactual response in which witnesses *added* the influence of the cautionary instruction had a greater impact than generating a counterfactual response in which witnesses *subtracted* the influence of the cautionary instruction. A 48-hour time delay between providing an actual identification response and providing a counterfactual identification response had no significant effect on this pattern.

Again, it had been predicted that an asymmetric estimation would lead witnesses who received the cautionary instruction to underestimate the impact that the cautionary instruction

had on their lineup decision, and witnesses who did not receive the cautionary instruction to more accurately estimate the impact that the cautionary instruction would have had on their lineup decision. Results were consistent with this prediction. Comparisons of actual to counterfactual responses suggest that many witnesses who received the cautionary instruction and rejected the lineup failed to recognize that their rejection was a consequence of the cautionary instruction. In contrast, these comparisons also suggest that witnesses who did not receive the cautionary instruction and consequently made an identification correctly tended to recognize that they would have rejected the lineup had they instead received the cautionary instruction. Again, it appears that the ability of witnesses to accurately estimate the influence of the cautionary instruction is dependent on whether they received that instruction or not.

GENERAL DISCUSSION

The legal system often implicitly assumes that witnesses can accurately report on the effects of influencing variables (the *accurate report* assumption). This general assumption is composed of four more specific assumptions: that witnesses consciously *notice* an influencing variable, that they *remember* that influencing variable at a later time, that they can *determine* whether that variable had an influence on them, and that they can accurately *estimate* the effects of the influence by assessing the influencing variable's direction and magnitude of influence. The purpose of the current experiments was primarily to assess the validity of these assumptions, and secondarily to determine whether the accuracy of witnesses' reports of influence was moderated by time delay.

Results clearly indicate that witnesses in both experiments tended to consciously notice and remember the influencing variable. When asked to provide an account of what the experimenter had told them immediately after (Experiment 1) or before (Experiment 2) the lineup identification, the majority of witnesses (approximately 80 - 85% across both experiments) who had received the influencing variable correctly recalled having received it. The majority of witnesses (approximately 85 - 90% across both experiments) who had received the influencing variable also correctly recognized the specific feedback statement (Experiment 1) or instruction (Experiment 2) that was given to them from a list of five options. This cannot be attributed to any sort of response bias because almost all witnesses who did not receive the influencing variable correctly reported not having received any confirming feedback (Experiment 1) or cautionary instruction (Experiment 2). Thus, witnesses across both experiments, as a whole, tended to notice and remember the influencing variable that was presented to them. The fact that rates of recalling/recognizing

the influencing variable were not moderated by a 48-hour time delay in either experiment indicates that witnesses tended to remember having received the influencing variable regardless of delay condition.

One interesting exception to the general finding that witnesses tended to remember the influencing variable was found in Experiment 1. Although most witnesses in this experiment correctly reported receiving post-identification feedback *before* answering questions about their confidence, view, and attention, over one third incorrectly reported that they received the feedback *after* answering those questions. This finding is especially intriguing in light of the fact that such a temporal order of events would have made it impossible for feedback to have influenced witnesses' responses, and yet these witnesses nonetheless estimated having been just as influenced by the feedback as witnesses who reported having received feedback before giving their responses. It should be noted that all witnesses were told the correct temporal order of events prior to generating their counterfactual responses, and this reminder may have led witnesses to assume that feedback could have influenced them. The current data do not allow an assessment of whether a similar finding would occur among witnesses who are not told the correct temporal order before generating counterfactual responses. But if this finding is replicable, it would be more than simply an intellectual curiosity, as misremembering when one was presented with an influencing variable has important real-world implications. A witness who incorrectly claims that she did not receive feedback until after giving a confidence statement might lead her defense attorney to forgo the possibility that the feedback influenced her confidence statement. This finding may therefore be important, and is certainly worthy of future investigation.

Results concerning witnesses' abilities to determine whether they were influenced, and to accurately estimate the effects of the influence, were somewhat more interesting. After giving their actual lineup responses (reports of retrospective confidence, view, and attention in Experiment 1; lineup identification decisions in Experiment 2), all witnesses in both experiments were asked to give a counterfactual lineup response. These counterfactual responses were witnesses' estimates of how they would have responded to the key questions had they been in the alternative experimental condition (confirming feedback vs. no feedback for Experiment 1; cautionary instruction vs. no cautionary instruction for Experiment 2). This technique allowed an analysis of whether witnesses could accurately estimate the effects of the influence by comparing the counterfactual responses of witnesses who received the influencing manipulation to the actual responses of witnesses who did not receive the influencing manipulation, and vice versa.

Using this analysis, Experiment 1 demonstrated that witnesses were quite accurate when 'subtracting' the influence of confirming feedback, but were not accurate when 'adding' the influence of confirming feedback. These latter witnesses tended to overestimate the influence confirming feedback would have had. Conversely, Experiment 2 demonstrated that witnesses were quite accurate when 'adding' the influence of a cautionary instruction, but were not accurate when 'subtracting' the influence of a cautionary instruction. These latter witnesses tended to underestimate the influence the cautionary instruction had. None of these results from either study were significantly moderated by a 48-hour time delay. Thus, for each experiment, some witnesses were able to accurately determine the magnitude of influence, whereas others were not. Additionally, the accurate witnesses in Experiment 1 tended to be those who received the influencing variable (confirming feedback) and had to

subtract its influence whereas the accurate witnesses in Experiment 2 tended to be those who did not receive the influencing variable (the cautionary instruction) and had to add its influence. Can these apparently conflicting results be integrated into a larger theoretical framework?

Despite the seemingly opposing findings of the two experiments, results from both experiments exhibited an asymmetric estimation of influence pattern, such that witnesses who received the influencing manipulation estimated significantly less of an effect of the influencing variable than witnesses who did not receive the influencing manipulation. This asymmetric estimation of influence pattern had been predicted a priori on the basis of two independent literatures, one concerning the hindsight bias and the other concerning counterfactual thinking. According to research on the hindsight bias, people have a difficult time 'erasing' information from their minds, as they tend to think that they would have known that information anyway. In the current experiments, one can think of witnesses who received the influencing manipulation as having also received additional information about the identification process (i.e., that they identified the correct person in Experiment 1; that the actual criminal may not be in the lineup in Experiment 2). When generating counterfactual responses about how they would have responded had they not received that information, these witnesses were likely to have thought that despite not having received that information explicitly from the lineup administrator, they would have known it anyway. Consequently, they would have reasoned that whether or not they had received that information, their responses would not have been very different. In contrast, witnesses who did not receive the influencing manipulation gave their actual responses without having had that extra information. When later asked how they would have responded had they been

given that information, these witnesses could reason how that information would have influenced them, and consequently their counterfactual responses would differ from their actual responses. Thus, the asymmetric estimation of influence pattern may simply reflect the fact that the hindsight bias affected witnesses who received the influencing manipulation, but not witnesses who did not receive the influencing manipulation.

According to research on counterfactual thinking, additive frames have a greater impact than subtractive frames. It is possible that witnesses who did not receive the influencing manipulation self-generated additive frames when generating their counterfactual responses (e.g., "How much *more* confident would I have been had I received confirming feedback?") whereas witnesses who did receive the influencing manipulation self-generated subtractive frames when generating their counterfactuals (e.g., "How much *less* confident would I have been had I not received confirming feedback?"), leading to the asymmetric pattern of data seen in Experiment 1. It is thought that additive versus subtractive frames lead to an asymmetry because people tend to give more weight to factors that *produce* an outcome than to factors that *inhibit* an outcome (Dunning & Parpal, 1989). Thus, witnesses who did not receive the influencing manipulation and who self-generated an additive frame when giving their counterfactual responses are likely to focus on factors that would have produced the outcome of high confidence (i.e., the confirming feedback). In contrast, witnesses who did receive the influencing manipulation and who self-generated a subtractive frame when giving their counterfactual responses are nonetheless still likely to focus on factors that would produce, rather than inhibit, the outcome of high confidence. For example, they might tend to focus on the sense of familiarity they experienced while looking at that lineup

member while ignoring the perceived discrepancies between their memory of the criminal and the identified lineup member's appearance.

The explanations for the asymmetric estimation of influence pattern offered by the hindsight bias and the counterfactual research on framing are similar, though not identical. Both postulate that counterfactual responses generated by witnesses who received the influencing manipulation should show less of a change (compared to actual responses) than those generated by witnesses who did not receive the influencing manipulation. However, the hindsight bias approach explains the asymmetry by claiming that witnesses who received the influencing manipulation cannot completely 'erase' the influencing information from their minds when generating counterfactuals. The counterfactual approach, in contrast, explains the asymmetry by claiming that witnesses who received the influencing manipulation cannot 'erase' their high confidence from their minds when generating counterfactuals, which leads to a biased weighting of factors that might produce high confidence. The difference between the two approaches therefore largely comes down to whether it is knowledge of the antecedent condition (i.e., the influencing information) or knowledge of the outcome (i.e., the high confidence) that biases witnesses who receive an influencing manipulation. Future research is needed to disentangle these two explanations for the asymmetry. For example, imagine that witnesses were given disconfirming feedback instead of confirming feedback. The hindsight bias approach would predict that because disconfirming feedback gives witnesses information that they did not identify the actual criminal, those given disconfirming feedback should estimate less of an influence of feedback than those given no feedback. However, the counterfactual approach would predict the opposite. Because witnesses given disconfirming feedback should self-generate an additive frame when

generating their counterfactual response (i.e., "How much *more* confident would I have been had I not received disconfirming feedback?"), they should estimate more of an influence of feedback than witnesses not given feedback, who should self-generate a subtractive frame (i.e., "How much *less* confident would I have been had I received disconfirming feedback?"). Although the exact source of the asymmetry awaits future research, note that both explanations place the cause of the asymmetry at the feet of the witnesses who received the influencing manipulation. Both explanations stress that it is the misperceptions of these witnesses, not witnesses who did not receive the influencing manipulation, that drive the asymmetric pattern. Unfortunately, in the real world, it is typically witnesses who received an influencing variable who are asked to estimate the effects of the influencing variable. The aforementioned hypotheses for the asymmetry pattern suggest that it is precisely these witnesses who are most biased when generating counterfactuals.

Interestingly, however, the asymmetric estimations of influence led to different results for Experiments 1 and 2. In Experiment 1, the asymmetry was a result of witnesses who received the influencing manipulation accurately estimating the effects of feedback and witnesses who did not receive the influencing manipulation overestimating the effects of feedback. In Experiment 2, the asymmetry was a result of witnesses who received the influencing manipulation underestimating the effects of the cautionary instruction and witnesses who did not receive the influencing manipulation accurately estimating the effects of the cautionary instruction. Thus, although the overall asymmetric pattern of data was remarkably consistent between experiments, the specifics of the pattern differed. Perhaps there are certain qualities of the specific influencing variable under study that affect the

absolute degree to which witnesses can accurately estimate the effects of influence, but the estimation nonetheless tends to occur asymmetrically.

But what qualities of the influencing variable would affect the absolute degree of estimation? Although lacking relevant data that speak to this question, speculation is possible. It is almost certainly true that witnesses have different implicit theories of the effects of different influencing manipulations. Differences in these theories may originate from how easily they can imagine the specific variable having an effect, the salience of the variable, the perceived malleability of the response, etc. Witnesses may, for example, readily assume that post-identification feedback would affect their retrospective confidence reports because they perceive retrospective confidence to be malleable. They may not think that feedback could affect their reports of view and attention, however, because they perceive these variables to be more objective and hence less malleable. Similarly, witnesses may be more skeptical of the influence that pre-lineup instructions can have on the likelihood of making a lineup identification because they assume identifications to be purely a function of perceived familiarity. Depending on how accurate these implicit theories about influence are compared to the actual influence exerted by such manipulations, different patterns of data will result. Given that an asymmetry exists, if witnesses tend to have implicit theories that overestimate the strength of a manipulation, then one would expect to see data as found on the confidence measure in Experiment 1: Witnesses who do not receive an influencing manipulation should overestimate the influence whereas witnesses who do receive an influencing manipulation should more accurately estimate the influence. If witnesses tend to have accurate implicit theories about the impact of a manipulation, then one would expect to see data as found in Experiment 2: Witnesses who do not receive an influencing

manipulation should accurately estimate the influence whereas witnesses who do receive an influencing manipulation should underestimate the influence. If witnesses tend to have implicit theories that underestimate the strength of a manipulation, then one would expect a third pattern: Witnesses who receive an influencing manipulation as well as witnesses who do not receive an influencing manipulation should both underestimate the influence, but those who received the influencing manipulation will underestimate the influence more. Thus, differences in witnesses' implicit theories about a manipulation may affect the absolute degree of estimation that they exhibit. This coupled with a general asymmetric estimation of influence pattern would lead to different patterns of results, depending on the specific influencing variable in question. Future research would be needed to test this hypothesis.

Strengths and Limitations.

No previously-used paradigm could adequately answer the question of whether witnesses can accurately estimate the effects of influence, and the question had, until now, gone largely unasked. The current experiments introduce a novel experimental paradigm to the eyewitness literature that allows us to examine this question. The paradigm is a powerful one for a number of reasons. First, it assesses not simply whether witnesses believe they were influenced or not (e.g., Wells et al., 2003), but also the magnitude of the perceived influence. This is important because although witnesses may accurately claim to have been influenced by a given variable, they might not accurately assess the magnitude of that influence, a point obscured by previous attempts to address this issue. Second, assessing both the actual as well as the counterfactual responses of witnesses allows a direct comparison between the counterfactual responses of witnesses who received the influencing manipulation and the actual responses of witnesses who did not receive the influencing manipulation, and vice

versa, to assess whether witnesses' tend to overestimate, underestimate, or accurately estimate the effects of influence. Third, the paradigm is modeled after certain real-world practices in which witnesses are asked how they would have responded had they received or not received some potentially influencing manipulation, thus maintaining a high degree of external validity.

Nonetheless, there are a number of limitations of the current research. Estimation of influence was assessed at the group level, not the individual witness level. Any conclusions that are drawn therefore do not necessarily reflect individual witnesses. For example, witnesses in Experiment 2 who received the cautionary instruction tended, as a group, to underestimate its influence. This simply means that fewer witnesses switched from a "not there" response (in their actual report) to a lineup identification response (in their counterfactual report) than was expected. But of course many witnesses did switch responses. Thus, although as a group witnesses underestimated the influence of the cautionary instruction, some individual witnesses might have accurately estimated its influence and others might not have. Additionally, although the 'correct' response for most of those witnesses was to switch to a lineup identification (i.e., they actually would have made an identification had they not received the cautionary instruction), the 'correct' response for some was to maintain a 'not there' response (i.e., they actually would have still made a 'not there' response, even without the cautionary instruction). There is no way to tell which of the individual witnesses would have made a 'not there' response had they not received the cautionary instruction, only that a certain proportion of them would have. Similar ambiguities arise from Experiment 1; although we know that on average confirming feedback boosted

witnesses' retrospective confidence about 9%, this does not tell us how any individual witness was influenced.

This restriction on the interpretation of these data may be an unfortunate but necessary limit imposed by the type of question asked in these experiments, a lament also noted by other researchers of people's abilities to estimate influence (Wilson et al., 2002). It seems difficult, if not impossible, to conceive of a paradigm in which one could measure individual witnesses' accuracy. That would require a knowledge of what the witness would actually have done in the alternative experimental condition, which would require each witness to be run through the experiment twice – one time receiving the influencing variable and the other time not receiving the influencing variable. That is the only way to compare an individual witness's counterfactual response to his or her own actual response in the alternative experimental condition. Unfortunately, the problems inherent with this type of within-subjects design would render the interpretation of any such data meaningless. Thus, the between-subjects, group-level assessment of data may be a necessary component of this type of research.

Caution should also be taken when generalizing from these experiments. Although the asymmetric estimation of influence pattern was observed in the predicted direction across both experiments, and although it has a solid theoretical grounding in both the hindsight bias and the counterfactual literature, it would be premature to conclusively state that a similar asymmetric pattern would exist across most other potential manipulations. Perhaps, for example, it is not the presence or absence of the influencing variable that causes the asymmetry, but some other underlying factor that coincidentally co-varied with the presence/absence factor. Although no such underlying factor is immediately apparent,

replication with other manipulations is nonetheless required to validate that the presence or absence of the influencing variable is truly the cause of the asymmetry.

This research also involved a somewhat limited operationalization of time delay. Although time delay was found to have no significant moderating effects on any variable for either of the two studies, this lack of result may simply have occurred because the time delay used – 48 hours – was not long enough to produce any deleterious effects on witnesses' memories for the influencing variable. Perhaps a time delay of a week, a month, or longer could interact with witnesses' estimates of influence, but given the limitations of laboratory experiments, these delay manipulations would be increasingly more difficult to test. Unfortunately, in real-world cases, delays of months or even years between a lineup identification and a trial are not uncommon. In a recent DNA exoneration case, a witness was asked in court about lineup procedures that occurred fifteen years prior (Newsome v. McCabe, 2002).

It is also unknown, on the basis of the two experiments described here, whether any estimates of influence made by witnesses were the result of introspective awareness of the influence or implicit theories about the effects of influence. Could those witnesses who accurately estimated the effects of influence actually retrace their cognitive processes and accurately 'undo' or 'add' the effects of influence? Or did they simply have an implicit theory about how they think they would have been affected, and happened to be accurate? Although these are difficult questions to definitively answer, data suggest that witnesses were responding on the basis of implicit theories rather than introspection, for a number of reasons. For instance, 27% of witnesses who received confirming feedback in Experiment 1 gave the same counterfactual response as actual response on the confidence measure,

meaning that they felt that they were not influenced by the confirming feedback. Wells et al. (2003), however, found that 65% of witnesses who received confirming feedback in their study reported that they were not affected by the feedback. This large discrepancy is most easily explained in one of two ways.

First, witnesses in the current experiment were reminded of *when* they received feedback (i.e., that it was prior to generating their counterfactual responses), whereas witnesses in the Wells et al. (2003) study were not reminded of when they received confirming feedback. If some witnesses in the latter study erroneously thought they received feedback *after* giving their confidence reports (as did 34% in the current study) and they were never disabused of this notion, then they may have reasoned that feedback could logically not have influenced them. In contrast, witnesses in the current experiment *were* disabused of this notion, and thus should have been more likely to indicate that they could have been influenced by the feedback. Note that this explanation for the discrepancy between the two studies almost necessitates the use of implicit theories; witnesses who would have said that they were not influenced (because they mistakenly thought feedback followed their responses) reported having been influenced when they were disabused. How could they suddenly now report the effects of influence other than with an implicit theory?

Second, the discrepancy between the two studies may be explained by the way the question concerning influence was asked in each study. Witnesses in the Wells et al. (2003) study were asked whether their response to the confidence question had been influenced by the feedback, and they simply responded yes or no. In contrast, witnesses in Experiment 1 of the present study were asked how they would have responded to the confidence question had they not received feedback, and were given the confidence scale to indicate their response.

Although technically the same question, the phrasing of the question in the current study may have implied to the witnesses that their counterfactual response should be different from their actual response, leading to a lower rate of 'no change' responses. If this is indeed the case, then it indicates that witnesses' responses were affected by factors external to their own cognitive processes (e.g., wording of the question, beliefs about what the correct response is), suggesting a lack of introspective awareness. In particular, the phrasing of the question used in Experiment 1 may have lead witnesses to think about implicit theories about how confirming feedback would have influenced them, and these implicit theories may have driven their responses.

It is also difficult to reconcile the fact that witnesses overestimated and underestimated the effects of influence with the idea that they used introspective awareness to generate their counterfactual responses. If witnesses truly had direct introspective access, how could they possibly think that confirming feedback would have had a stronger impact than it in fact would have? How could they think that receiving the cautionary instruction did not influence them when it in fact did? In contrast, both of these findings are perfectly consistent with an implicit theories perspective – overestimation and underestimation of influence occur when witnesses' implicit theories overestimate and underestimate, respectively, the actual influence of a variable. If it is true, as argued previously, that differences in the absolute degree of estimation observed across influencing variables are the result of differences in implicit theories about those influencing variables, little room is left for introspection.

Even if introspection could somehow account for systematic over- or underestimation of influence, there is still no theoretical reason why this would vary as a function of response

or influencing manipulation. But witnesses who did not receive feedback tended to overestimate the effect that feedback would have had on their confidence whereas they more accurately estimated the effect that feedback would have had on their reports of view and attention. If witnesses were introspecting, what mechanism would have led them to overestimate the effect of feedback when making one response but accurately estimate the effect of feedback on two other responses? Once again, the implicit theories perspective offers an easy explanation. If witnesses have implicit theories that self-reported retrospective confidence is highly malleable, but self-reported view of the criminal and attention paid to the criminal are not, then their estimates of the influence of feedback on their reports of confidence should be greater than their estimates of the influence of feedback on their reports of view and attention. In fact, consistent with the idea that confidence responses are perceived as being more malleable than view and attention responses, previous research has found that witnesses more readily admit that confirming feedback could have influenced their responses to retrospective confidence questions rather than their responses to view and attention questions (Wells & Bradfield, 1998). Thus, whereas an introspective account of witnesses' abilities would require ad hoc additions to the theory in order to explain the aforementioned anomalies, an implicit theories perspective offers a parsimonious and powerful approach to understanding witnesses' estimates of influence. Until more research uncovers witnesses' implicit theories, however, this perspective necessarily remains speculative.

Future research

In a practical sense, it does not matter much in the real world how witnesses arrive at their final counterfactual responses, be it introspection or implicit theories, as long as those

responses are accurate. Knowledge about the introspective capabilities of witnesses does matter though when trying to determine how to improve witnesses' performance when they are not accurate. If witnesses' responses are guided by implicit theories, and not introspective awareness, then an understanding of how witnesses generate implicit theories, and the impact those implicit theories have on witnesses' subsequent responses, is of paramount importance. Future research might wish to more fully examine the implicit theories witnesses have about various lineup manipulations.

The willingness of witnesses to give different counterfactual responses than actual responses suggests that regardless of whether through introspection or implicit theories, witnesses are willing to report that they were (or could be) influenced. But if witnesses were aware of the influence of confirming feedback or of the cautionary instruction, why would witnesses have let these variables influence them in the first place? Why would they not correct for them at the time they gave their actual responses? Perhaps witnesses are not aware of the influence of variables at the time they are experiencing them, but only become aware retrospectively at a later time when they are asked about them. Perhaps when witnesses are receiving the influencing variable, they lack an alternative version of events, and it is only when such an alternative is presented to them that they can contrast the two to determine influence. This would suggest that witnesses' abilities to accurately correct for influence depend on the salience of a counterfactual occurrence. For example, maybe witnesses who are given a biased lineup instruction suggesting that the criminal is present in the lineup could 'undo' the influence of that instruction in real time, as opposed to retrospectively, if they were also aware of the possibility that they could instead have received a cautionary instruction. This possibility is partially supported by research that found that people's

theories about the causal relationships between variables are influenced by the relative availability and salience of various causal explanations (Anderson & Sechler, 1986). Thus, if witnesses are aware that they could have received a cautionary instruction, for example, the salience of that knowledge alone should affect their implicit theories about the effects of pre-lineup instructions. Not only could these implicit theories then moderate the effects of pre-lineup instructions, but they might also affect witnesses' estimates of influence. Anderson and Sechler found that simply generating possible reasons why two variables might be related increased people's belief in that relationship. Similarly, they also found that subsequently generating possible reasons why the two variables might not be related eliminated that effect. Thus, an awareness of an alternative pre-lineup instruction might lead witnesses to generate causal explanations about the effects of that variable, which could affect witnesses' implicit theories and subsequently their estimates of influence. Such limits and moderating factors of witnesses' abilities to accurately estimate the effects of influence should be another area at which future research is aimed.

The previous discussion suggests that in cases where witnesses who received some influencing variable tend to underestimate the influence of that variable (as in Experiment 2), it may be beneficial to first allow witnesses to generate causal explanations about how that influencing variable might generally affect lineup responses. If this increases the salience of that causal relationship, this should lead witnesses to estimate a greater effect of the influencing variable on their own responses, and hence to be more accurate when generating their counterfactual responses. This idea is consistent with the concept of belief perseverance, which describes the tendency of people to maintain their pre-existing beliefs about the relationship between two variables even when data supporting that belief are completely

discredited (Anderson & Kellam, 1992; Ross, Lepper, & Hubbard, 1975). The asymmetric estimation of influence pattern observed in the current experiments may reflect a type of belief perseverance. Witnesses who received an influencing manipulation are given information, which they are supposed to in effect "discredit" when giving their counterfactual responses. The asymmetric estimation of influence pattern may reflect the inability of witnesses who received an influencing manipulation to properly "discredit" the information that was given to them. However, belief perseverance has been shown to be reduced when people are forced to generate counterexplanations that conflict with their pre-existing beliefs (Anderson, 1982). Thus, having witnesses generate possible explanations for how a lineup manipulation might influence people should have the effect of reducing belief perseverance, which could eliminate the asymmetric estimation of influence pattern. However, this tactic should not be used indiscriminately: If witnesses who receive some influencing variable tend to accurately estimate the influence of that variable (as in Experiment 1), this procedure might lead them to overestimate the effect of the variable. Again, a more complete understanding of witnesses' implicit theories about various lineup manipulations would be needed to accurately predict when this procedure might be beneficial.

The conclusion from these experiments is clear, albeit tentative: Witnesses tend to asymmetrically estimate the effects of influencing variables. This asymmetric pattern should give one pause in assuming witnesses can accurately report on the effects of influencing variables. Previous discussion suggests that witnesses' abilities to accurately estimate the effects of influence depend on 1) whether they received the influencing manipulation or not (which drives the asymmetric pattern); and somewhat more speculatively, 2) whether

witnesses hold accurate or inaccurate implicit theories about the strength of the influencing manipulation (which might drive the absolute degree of estimation). Very little is currently known about witnesses' implicit theories about the effects of various lineup manipulations. Although some studies touching on this issue have demonstrated that peoples' implicit theories about the effect of influencing lineup manipulations are often inaccurate (e.g., Kassin & Barndollar, 1992; Hasel, 2006), the question concerning what specific theories these people actually hold has been largely neglected. Without knowing witnesses' implicit theories, it is impossible to predict whether witnesses will tend to overestimate, underestimate, or accurately estimate the effects of the influence of a given manipulation. We can only predict that the correction will occur asymmetrically such that witnesses who receive an influencing manipulation will estimate less of an effect than will witnesses who do not receive an influencing manipulation. Research into witnesses' implicit theories is therefore strongly encouraged. However, despite our lack of understanding of the relationship between witnesses' implicit theories and their estimates of influence, the inconsistent estimates of influence seen in the current two experiments lead to one unequivocal recommendation: The best way to avoid influenced responses among eyewitnesses is to avoid influencing them in the first place.

The current experiments introduced a novel paradigm to the eyewitness field in which witnesses' counterfactual responses as well as their actual responses are assessed. The strength of this paradigm lies in its ability to quantitatively assess the degree to which witnesses can estimate the effects of influence, an ability that is impossible with the previously-used paradigm. The data from the current experiments emphasize the promise of the actual/counterfactual paradigm by underscoring its sensitivity to finding different patterns

in the data. Across two experiments, underestimation, overestimation, and accurate estimation patterns were all shown; these estimation patterns were shown to occur asymmetrically as a function of whether witnesses received an influencing variable or not across two different influencing manipulations; and even within a single experiment (Experiment 1), estimation patterns were shown to vary as a function of outcome measure (confidence vs. view and attention). Although these findings in and of themselves do not prove the usefulness of the paradigm, they do suggest that it is sensitive enough to allow a multitude of data patterns to present themselves. In addition, the fact that some witnesses overcorrected in Experiment 1 suggests that concerns broached earlier about the previously-used paradigm – namely, that witnesses may refuse to admit that they were influenced due to psychological reactance or out of a desire to dismiss the idea that external factors may have influenced them – may not apply to the actual/counterfactual paradigm. It is hoped that this paradigm will allow further research into the question of whether witnesses can accurately report on the effect of influencing variables, for although the present experiments provide a promising experimental paradigm as well as preliminary data upon which to build subsequent hypotheses, they only begin to scratch the surface of a question to which an answer is not yet fully known.

FIGURES

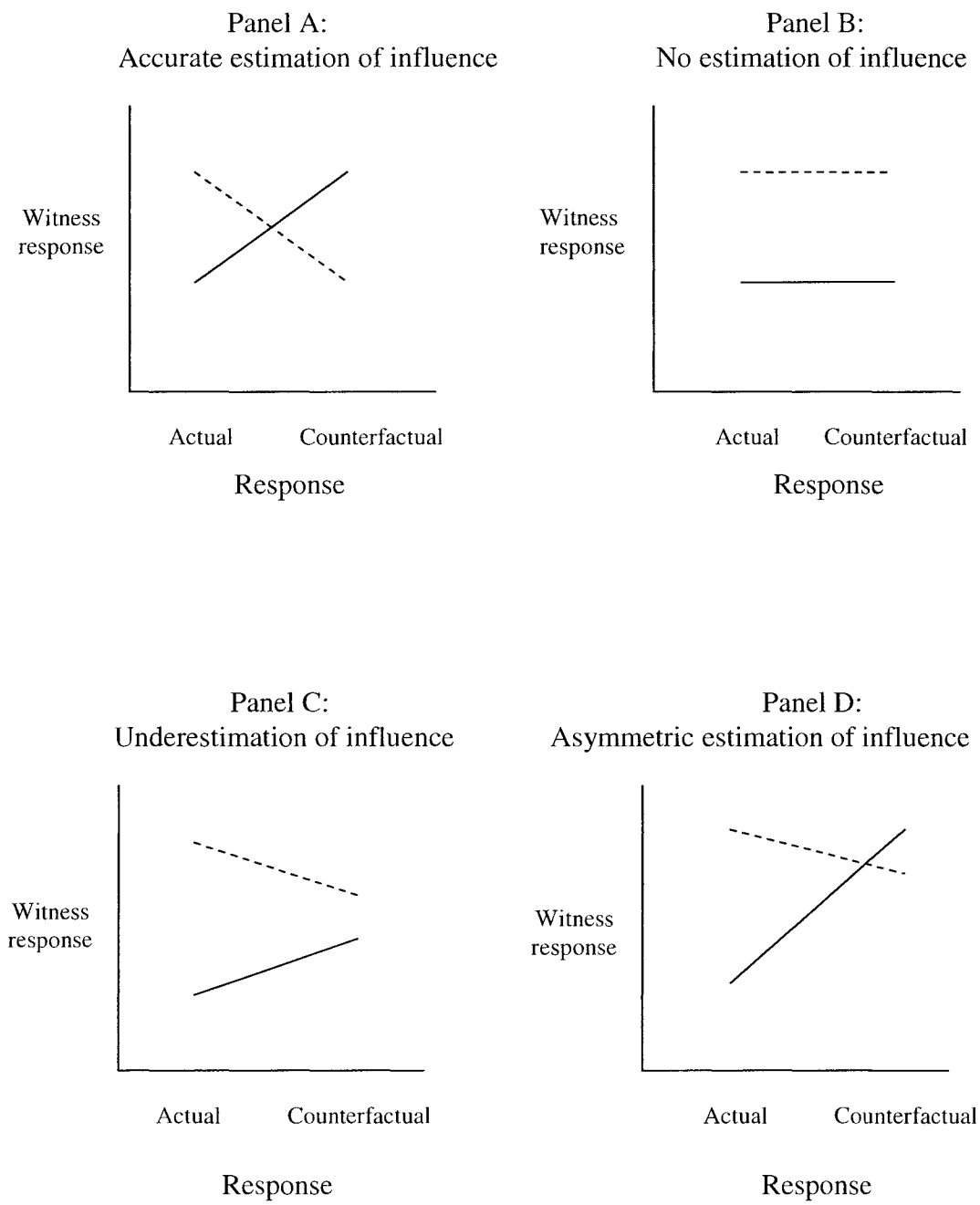

Figure 1. Possible patterns between witnesses' responses (actual vs. counterfactual) and experimental manipulation. Dotted and solid lines represent witnesses who received and did not receive the influencing variable, respectively. 'Witness response' is retrospective confidence for Study 1 and probability of a 'not there' response for Study 2.

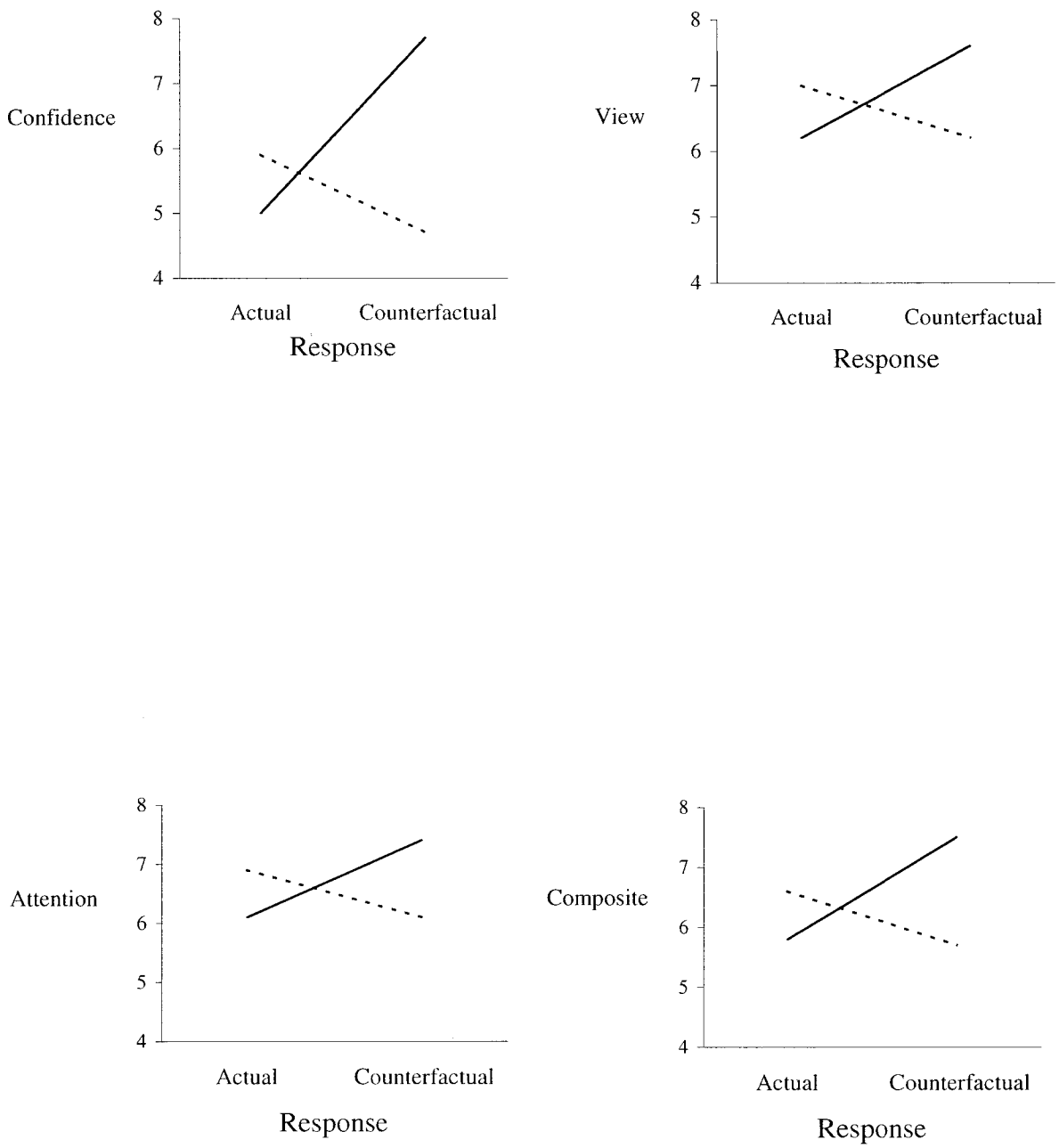

Figure 2. Actual and Counterfactual responses to confirming feedback. Dotted lines represent witnesses who received confirming feedback; solid lines represent witnesses who did not receive confirming feedback.

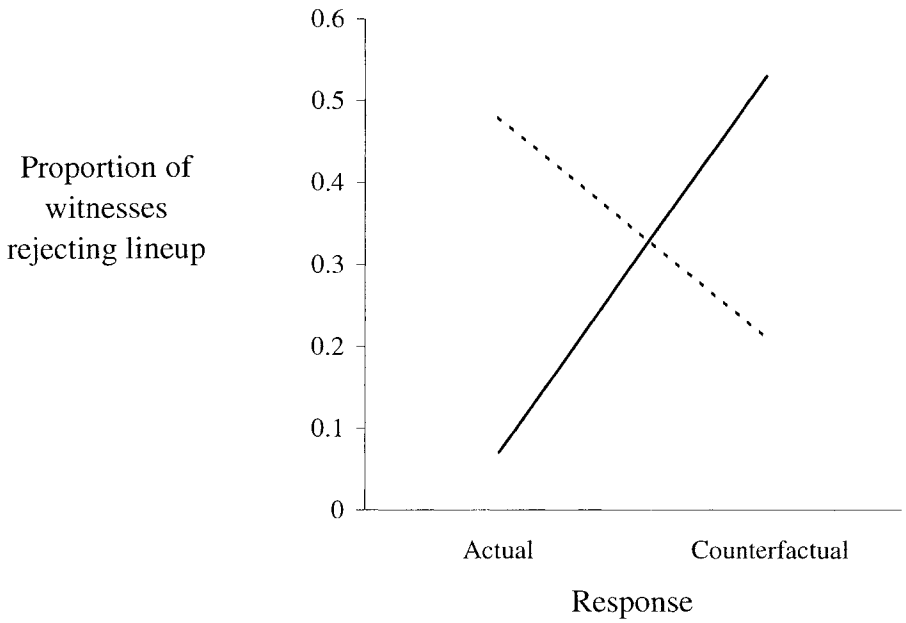

Figure 3. Proportion of actual and counterfactual lineup rejections. The dotted line represents witnesses who received the cautionary instruction; the solid line represents witnesses who did not receive the cautionary instruction.

TABLES

Table 1

Pearson correlation coefficients between confidence, view, and attention on actual and counterfactual responses

Measure	Confidence	View	Attention
	Actual responses		
Confidence	—	.38**	.28**
View		—	.25*
Attention			—
	Counterfactual responses		
Confidence	—	.54**	.36**
View		—	.42**
Attention			—

* $p < .05$ ** $p < .001$

Table 2

Means and standard deviations (in parentheses) of confidence, view, attention, and composite scores as functions of feedback, delay, and response

	Delay			No delay			Overall		
Measure	Actual	Remembered	Counterfactual	Actual	Remembered	Counterfactual	Actual	Remembered	Counterfactual
				Confirming feedback					
Confidence	6.2 (2.8)	6.1 (2.7)	5.1 (2.6)	5.7 (2.5)	5.6 (2.5)	4.3 (2.3)	5.9 (2.6)	5.8 (2.6)	4.7 (2.5)
View	7.3 (2.1)	7.4 (2.3)	6.6 (2.1)	6.7 (2.4)	6.6 (2.4)	5.6 (2.7)	7.0 (2.2)	6.9 (2.4)	6.2 (2.4)
Attention	7.1 (2.0)	6.8 (2.1)	6.2 (2.4)	6.8 (2.1)	6.8 (2.2)	6.0 (2.5)	6.9 (2.1)	6.8 (2.1)	6.1 (2.5)
Composite	6.9 (1.7)	–	6.0 (1.9)	6.4 (1.7)	–	5.4 (1.9)	6.6 (1.7)	–	5.7 (1.9)
				No feedback					
Confidence	5.2 (2.6)	5.2 (2.5)	7.5 (2.1)	4.9 (2.5)	4.9 (2.4)	7.7 (1.9)	5.0 (2.5)	5.0 (2.4)	7.7 (2.0)
View	6.3 (1.8)	6.3 (2.3)	7.7 (1.6)	6.1 (2.0)	6.0 (2.0)	7.5 (1.9)	6.2 (1.9)	6.2 (2.1)	7.6 (1.8)
Attention	6.3 (2.1)	6.3 (1.8)	7.3 (1.8)	5.9 (2.5)	5.9 (2.4)	7.4 (2.2)	6.1 (2.3)	6.1 (2.1)	7.4 (2.0)
Composite	5.9 (1.5)	–	7.5 (1.3)	5.6 (1.7)	–	7.5 (1.6)	7.5 (1.4)	–	7.5 (1.4)

Note. Confidence scores were converted to a 0 - 10 scale to match the other measures. Higher scores indicate greater confidence, a better view, and more attention paid to the criminal's face. Overall actual and counterfactual responses are also plotted in Figure 2.

Table 3

Means and standard deviations (in parentheses) of witnesses' estimates of the magnitude of the feedback effect as a function of feedback condition and measure

Measure	Confirming feedback	No feedback
Confidence	1.23 (1.10)	2.64 (2.11)
View	.77 (1.04)	1.44 (1.38)
Attention	.81 (1.07)	1.27 (1.34)
Composite	.94 (.84)	1.78 (1.16)

Note. Confidence scores were converted to a 0 - 10 scale to match the other measures.

Table 4

Percentage of witnesses who acknowledged an effect of feedback as a function of feedback condition and measure

Measure	Confirming feedback	No feedback
Confidence	73	88
View	44	75
Attention	50	73

Table 5

Means and standard deviations (in parentheses) of witnesses' estimates of the magnitude of the feedback effect, as a function of feedback condition and measure, among witnesses who acknowledged an effect of feedback

Measure	Confirming feedback	No feedback
Confidence	1.68 (.93)	2.99 (2.00)
View	1.74 (.86)	1.93 (1.26)
Attention	1.62 (.98)	1.76 (1.28)

Note. Confidence scores were converted to a 0 - 10 scale to match the other measures.

Table 6

Mean proportion of correct lineup rejections as a function of instruction condition, delay, and response

	Delay		No delay		Overall	
Instruction	Actual	Counterfactual	Actual	Counterfactual	Actual	Counterfactual
Cautionary	.52	.30	.44	.14	.48	.21
No cautionary	.04	.50	.09	.56	.07	.53

Note. Overall actual and counterfactual proportions are also plotted in Figure 2.

REFERENCES

Anderson, C. A. (1982). Inoculation and counterexplanation: Debiasing techniques in the perseverance of social theories. *Social Cognition, 1,* 126-139.

Anderson, C. A., & Kellam, K. L. (1992). Belief perseverance, biased assimilation, and covariation detection: The effects of hypothetical social theories and new data. *Personality and Social Psychology Bulletin, 18,* 555-565.

Anderson, C. A., & Sechler, E. S. (1986). Effect of explanation and counterexplanation on the development and use of social theories. *Journal of Personality and Social Psychology, 50,* 24-34.

Bradfield, A. L., Wells, G.L, & Olson, E.A. (2002). The damaging effect of confirming feedback on the relation between eyewitness certainty and identification accuracy. *Journal of Applied Psychology, 87,* 112-120.

Brehm, J. (1996). A theory of psychological reactance. Oxford, England: Academic Press.

Charman, S. D., & Wells, G. L. (in press). Eyewitness lineups: Is the appearance-change instruction a good idea? *Law and Human Behavior.*

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Douglass, A. L., & Steblay, N. (in press). Memory distortion in eyewitnesses: A meta-analysis of the post-identification feedback effect. *Applied Cognitive Psychology.*

Dunning, D., & Parpal, M. (1989). Mental addition versus subtraction in counterfactual reasoning: On assessing the impact of personal actions and life events. *Journal of Personality and Social Psychology, 57,* 5-15.

Epley, N., & Gilovich, T. (2001). Putting adjustment back in the anchoring and adjustment heuristic: Differential processing of self-generated and experimenter-provided anchors. *Psychological Science, 12,* 391-396.

Fischhoff, B. (1975). Hindsight ≠ foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1,* 288-299.

Garrioch, L., & Brimacombe, C.A.E. (2001). Lineup administrators' expectations: Their impact on eyewitness confidence. *Law and Human Behavior, 25*, 299-315.

Guilbault, R. L., Bryant, F. B., Brockway, J. H., & Posavac, E. J. (2004). A meta-analysis of research on hindsight bias. *Basic and Applied Social Psychology, 26,* 103-117.

Hasel, L. E. (2006). Do people discount testimony by eyewitnesses who received testimony-bolstering feedback? Unpublished master's thesis. Iowa State University, Ames.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction, *Psychological Review*, 80, 237-251.

Kassin, S. M., & Barndollar, K. A. (1992). On the psychology of eyewitness testimony: A comparison of experts and prospective jurors. *Journal of Applied Social Psychology, 22,* 1241-1249.

Landman, J. (1987). Regret and elation following action and inaction: Affective responses to positive versus negative outcomes. *Personality and Social Psychology Bulletin, 13,* 524-536.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32,* 322-328.

Malpass, R. S., & Devine, P. G. (1981). Eyewitness identification: Lineup instructions and the absence of the offender. *Journal of Applied Psychology, 66*, 482-489.

Markman, K. D., Gavanski, I., Sherman, S. J., & McMullen, M. N. (1995). The mental simulation of better and worse possible worlds. *Journal of Experimental Social Psychology, 29,* 87-109.

Martin, L. L., & Stapel, D. A. (1998). Correction and metacognition: Are people naïve dogmatists or naïve empiricists during social judgments? In Vincent Y. Yzerbyt, Guy Lories, & Benoit Dardenne (Eds.), *Metacognition: Cognitive and social dimensions* (pp. 228-247). Thousand Oaks, CA: Sage Publications, Inc.

Neil v. Biggers, 409 US 188 (1972).

Newsome v. McCabe (2002), 256 F.3d 747.

Nisbett, R. E., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review, 84,* 231-259.

Petty, R. E., & Wegener, D. T. (1993). Flexible correction processes in social judgment: Correcting for context-induced contrast. *Journal of Experimental Social Psychology, 29,* 137-165.

Phillips, M. R., McAuliff, B. D., Kovera, M. B., & Cutler, B. L. (1999). Double-blind photoarray administration as a safeguard against investigator bias. *Journal of Applied Psychology, 84,* 940-951.

Roese, N. J., & Olson, J. M. (1993). Self-esteem and counterfactual thinking. *Journal of Personality and Social Psychology, 65,* 199-206.

Roese, N. J., & Olson, J. M. (1995). What might have been. Mahwah, New Jersey: Lawrence Erlbaum Associates, Inc.

Rose, R. A., Bull, R., & Vrij, A. (2005). Non-biased lineup instructions do matter--A problem for older witnesses. *Psychology, Crime, & Law, 11,* 147-159.

Ross, L., Lepper, M. R., & Hubbard, M. (1975). Perseverance in self-perception and social perception: Biased attributional processes in the debriefing paradigm. *Journal of Personality and Social Psychology, 32,* 880-892.

Semmler, C., Brewer, N., & Wells, G. L. (2004). Effects of postidentification feedback on eyewitness identification and nonidentification confidence. *Journal of Applied Psychology, 89,* 334-346.

Sporer, S. L. (1992). Post-dicting eyewitness accuracy: Confidence, decision-times and person descriptions of choosers and non-choosers. *European Journal of Social Psychology, 22,* 157-180.

Sporer, S., Penrod, S., Read, D., & Cutler, B. L. (1995). Choosing, confidence, and accuracy: A meta-analysis of the confidence-accuracy relation in eyewitness identification studies. *Psychological Bulletin, 118,* 315-327.

Stapel, D. A., Martin, L. L., & Schwarz, N. (1998). The smell of bias: What instigates correction processes in social judgments? *Personality and Social Psychology Bulletin, 24,* 797-806.

Steblay, N. M. (1997). Social influence in eyewitness recall: A meta-analytic review of lineup instruction effects. *Law and Human Behavior, 21,* 283-297.

Taylor, S. E., & Brown, J. D. (1998). Illusion and well-being: A social psychological perspective on mental health. *Psychological Bulletin, 103,* 193-210.

Technical Working Group for Eyewitness Evidence (1999). *Eyewitness evidence: A guide for law enforcement*. Washington, DC: United States Department of Justice, Office of Justice Programs.

Weber, N., Brewer, N., Wells, G. L., Semmler, C., & Yeast, A. (2003). Eyewitness Identification Accuracy and Response Latency: The Unruly 10-12 Second Rule. *Journal of Experimental Psychology: Applied, 10,* 139-147.

Wegener, D. T., & Petty, R. E. (1997). The flexible correction model: The role of naïve theories of bias in bias correction. In M. Zanna (Ed.), *Advances in experimental social psychology* (pp. 141-208). San Diego, CA: Academic Press.

Wegner, D. M. (2003). The mind's best trick: How we experience conscious will. *Trends in Cognitive Sciences, 7,* 65-69.

Wells, G. L. (1978). Applied eyewitness testimony research: System variables and estimator variables. *Journal of Personality and Social Psychology, 36,* 1546-1557.

Wells, G. L., & Bradfield, A. L. (1998). "Good, you identified the suspect:" Feedback to eyewitnesses distorts their reports of the witnessing experience. *Journal of Applied Psychology, 83,* 360-376.

Wells, G. L., & Gavanski, I. (1989). Mental simulation of causality. *Journal of Personality and Social Psychology, 56,* 161-169.

Wells, G. L., Lindsay, R. C. L., & Ferguson, T. J. (1979). Accuracy, confidence, and juror perceptions in eyewitness identification. *Journal of Applied Psychology, 64,* 440-448.

Wells, G. L., Olson, E., & Charman, S. (2003). Distorted retrospective eyewitness reports as functions of feedback and delay. *Journal of Experimental Psychology: Applied, 9,* 42-52.

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.

Wilson, T. D., & Brekke, N. (1994). Mental contamination and mental correction: Unwanted influences on judgments and evaluations. *Psychological Bulletin, 116,* 117-142.

Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin, & D. Kahneman (Eds.). *Heuristics and biases: The psychology of intuitive judgment* (pp. 185-200). New York: Cambridge University Press.

APPENDIX A: EXPERIMENT 1 ACTUAL RESPONSE MEASURES

1) At the time you identified the person in the lineup, how certain were you that the person you identified from the lineup was the person you saw in the video?

0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
Not at all certain										Totally certain

2) How good of a view did you get of the person in the video?

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Very poor										Very good

3) How much attention were you paying to the person's face while viewing the video?

0	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
None										My total attention

APPENDIX B: EXPERIMENT 1 MEMORY MEASURES

Witnesses who received confirming feedback

---

*Immediately following your identification, what comments, if any, did the lineup administrator make to you regarding your identification?

Which of the following statements did the lineup administrator make to you following your lineup identification?

a) "Good. Other witnesses also identified him."
b) "Actually, you did not identify the suspect."
c) "Good. You identified the suspect."
d) "You actually identified the wrong lineup member."
e) The lineup administrator made no comment about your lineup identification.

The experimenter actually told you "Good. You identified the suspect" after your identification.

Did the lineup administrator make this statement before or after you answered the question about your certainty in your identification? (circle one)

Before After

---

*Note that each individual question was asked sequentially

APPENDIX B (continued)

The lineup administrator actually made that comment before you answered the question about your certainty in your identification.
After you made you identification, you responded to three questions. We are interested in whether you can remember your responses to those questions.

1) When you originally answered the following question about your certainty (after your identification), how did you respond?

At the time you identified the person in the lineup, how certain were you that the person you identified from the lineup was the person you saw in the video?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

2) When you originally answered the following question about your view of the criminal, how did you respond?

2) How good of a view did you get of the person in the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Very poor										*Very good*

3) When you originally answered the following question about the amount of attention you paid to the criminal, how did you respond?

3) How much attention were you paying to the person's face while viewing the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
None										*My total attention*

APPENDIX B (continued)

Witnesses who did not receive confirming feedback

Immediately following your identification, what comments, if any, did the lineup administrator make to you regarding your identification?

Which of the following statements did the lineup administrator make to you following your lineup identification?

a) "Good. Other witnesses also identified him."
b) "Actually, you did not identify the suspect."
c) "Good. You identified the suspect."
d) "You actually identified the wrong lineup member."
e) The lineup administrator made no comment about your lineup identification.

---

*Note that each individual question was asked sequentially

APPENDIX B (continued)

The lineup administrator actually made no comment about your identification.

After you made your identification, you responded to three questions. We are interested in whether you can remember your responses to those questions.

1) When you originally answered the following question about your certainty (after your identification), how did you respond?

At the time you identified the person in the lineup, how certain were you that the person you identified from the lineup was the person you saw in the video?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

2) When you originally answered the following question about your view of the criminal, how did you respond?

How good of a view did you get of the person in the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Very poor										*Very good*

3) When you originally answered the following question about the amount of attention you paid to the criminal, how did you respond?

How much attention were you paying to the person's face while viewing the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
None										*My total attention*

APPENDIX C: EXPERIMENT 1 COUNTERFACTUAL RESPONSE MEASURES

Witnesses who received confirming feedback

For the following questions, imagine that instead of saying "Good. You identified the suspect," the experimenter had not told you anything after you had made your identification.

1) If the experimenter had not told you anything after you had made your identification, how would you have answered the following question about your certainty?

At the time you identified the person in the lineup, how certain were you that the person you identified from the lineup was the person you saw in the video?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

2) If the experimenter had not told you anything after you had made your identification, how would you have answered the following question about your view of the criminal?

How good of a view did you get of the person in the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Very poor										*Very good*

3) If the experimenter had not told you anything after you had made your identification, how would you have answered the following question about how much attention you paid to the criminal?

How much attention were you paying to the person's face while viewing the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
None										*My total attention*

---

*Note that all witnesses were shown the original lineup as well as their actual responses when answering these questions

APPENDIX C (continued)

Witnesses who did not receive confirming feedback

For the following questions, imagine that instead of making no comment about your identification, the experimenter had told you "Good. You identified the suspect" immediately after making your identification.

1) If the experimenter had made this statement after you had made your identification, how would you have answered the following question about your certainty?

At the time you identified the person in the lineup, how certain were you that the person you identified from the lineup was the person you saw in the video?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

2) If the experimenter had made this statement after you had made your identification, how would you have answered the following question about your view of the criminal?

How good of a view did you get of the person in the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
Very poor										*Very good*

3) If the experimenter had made this statement after you had made your identification, how would you have answered the following question about how much attention you paid to the criminal?

How much attention were you paying to the person's face while viewing the video?

0	***1***	***2***	***3***	***4***	***5***	***6***	***7***	***8***	***9***	***10***
None										*My total attention*

---

*Note that all witnesses were shown the original lineup as well as their actual responses when answering these questions

APPENDIX D: EXPERIMENT 2 MEMORY MEASURES

Witnesses who received the cautionary instruction

How certain are you in your identification decision?

0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
Not at all certain										Totally certain

Immediately before you were shown a lineup, the lineup administrator said *"I have here a copy of a photo lineup."* What instructions, if any, did the lineup administrator give you after this statement?

After the lineup administrator said *"I have here a copy of a photo lineup,"* which of the following instructions did the lineup administrator then give to you?

a) "Keep in mind that your job is to identify the criminal. Please make an identification now."
b) "Keep in mind that it is as important to free the innocent as it is to identify the guilty. Please be cautious when making your identification."
c) "Keep in mind that the actual criminal may or may not be in the lineup. Look at the people in the photo lineup and tell me if you see the criminal."
d) "Keep in mind while you are making your identification that the criminal may have changed appearance."
e) "Look at the people in the photo lineup and try to identify the criminal you saw in the video."
f) None of the above

---

*Note that each individual question was asked sequentially

APPENDIX D (continued)

The lineup administrator actually instructed you to *"keep in mind that the actual criminal may or may not be in the lineup. Look at the people in the photo lineup and tell me if you see the criminal"* before showing you the lineup. We are now interested in whether you can remember the responses you made to the lineup.

1) Which lineup member did you identify immediately after seeing the mock crime? (Circle your response)

1) 1
2) 2
3) 3
4) 4
5) 5
6) 6
7) I did not identify any of the lineup members

2) After you made your identification, you responded to a question about your certainty. We are interested in whether you can remember your response to this question.

When you originally answered the following question about your certainty (after your identification), how did you respond?

How certain are you in your identification decision?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

APPENDIX D (continued)

Witnesses who did not receive the cautionary instruction

How certain are you in your identification decision?

0	**10**	**20**	**30**	**40**	**50**	**60**	**70**	**80**	**90**	**100**
Not at all certain										Totally certain

Immediately before you were shown a lineup, the lineup administrator said *"I have here a copy of a photo lineup."* What instructions, if any, did the lineup administrator give you after this statement?

After the lineup administrator said *"I have here a copy of a photo lineup,"* which of the following instructions did the lineup administrator then give to you?

a) "Keep in mind that your job is to identify the criminal. Please make an identification now."
b) "Keep in mind that it is as important to free the innocent as it is to identify the guilty. Please be cautious when making your identification."
c) "Keep in mind that the actual criminal may or may not be in the lineup. Look at the people in the photo lineup and tell me if you see the criminal."
d) "Keep in mind while you are making your identification that the criminal may have changed appearance."
e) "Look at the people in the photo lineup and try to identify the criminal you saw in the video."
f) None of the above

---

*Note that each individual question was asked sequentially

APPENDIX D (continued)

The lineup administrator actually instructed you to *"look at the people in the photo lineup and try to identify the criminal you saw in the video"* before showing you the lineup. We are now interested in whether you can remember the responses you made to the lineup.

Please look at the lineup that you were shown.

1) Which lineup member did you identify immediately after seeing the mock crime? (Circle your response)

1) 1
2) 2
3) 3
4) 4
5) 5
6) 6
7) I did not identify any of the lineup members

2) After you made your identification, you responded to a question about your certainty. We are interested in whether you can remember your response to this question.

When you originally answered the following question about your certainty (after your identification), how did you respond?

How certain are you in your identification decision?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

APPENDIX E: EXPERIMENT 2 COUNTERFACTUAL RESPONSE MEASURES

Witnesses who received the cautionary instruction

For the following questions, imagine that instead of telling you to *"keep in mind that the criminal may or may not be in the lineup. Look at the people in the photo lineup and tell me if you see the criminal"* the experimenter had told you to *"look at the people in the photo lineup and try to identify the criminal who you saw in the video."*

1) If the experimenter had given you this alternate instruction, how would you have responded to the lineup?

1) I would have identified lineup member #1
2) I would have identified lineup member #2
3) I would have identified lineup member #3
4) I would have identified lineup member #4
5) I would have identified lineup member #5
6) I would have identified lineup member #6
7) I would not have identified any of the lineup members.

2) If the experimenter had given you this alternate instruction, how would you have responded to the following question about your certainty?

How certain are you in your identification decision?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

---

*Note that all witnesses were shown the original lineup as well as their actual responses when answering these questions

APPENDIX E (continued)

Witnesses who did not receive the cautionary instruction

For the following questions, imagine that instead of telling you to *"look at the people in the photo lineup and try to identify the criminal who you saw in the video,"* the experimenter had told you to *"keep in mind that the criminal may or may not be in the lineup. Look at the people in the photo lineup and tell me if you see the criminal."*

1) If the experimenter had given you this alternate instruction, how would you have responded to the lineup?

1) I would have identified lineup member #1
2) I would have identified lineup member #2
3) I would have identified lineup member #3
4) I would have identified lineup member #4
5) I would have identified lineup member #5
6) I would have identified lineup member #6
7) I would not have identified any of the lineup members.

2) If the experimenter had given you this alternate instruction, how would you have responded to the following question about your certainty?

How certain are you in your identification decision?

0	***10***	***20***	***30***	***40***	***50***	***60***	***70***	***80***	***90***	***100***
Not at all certain										*Totally certain*

---

*Note that all witnesses were shown the original lineup as well as their actual responses when answering these questions