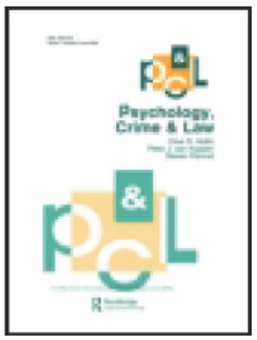

Psychology, Crime & Law

ISSN: 1068-316X (Print) 1477-2744 (Online) Journal homepage: http://www.tandfonline.com/loi/gpcl20

Judging guilt and accuracy: highly confident eyewitnesses are discounted when they provide featural justifications

Chad S. Dodson & David G. Dobolyi

To cite this article: Chad S. Dodson & David G. Dobolyi (2017): Judging guilt and accuracy: highly confident eyewitnesses are discounted when they provide featural justifications, Psychology, Crime & Law, DOI: 10.1080/1068316X.2017.1284220

To link to this article: http://dx.doi.org/10.1080/1068316X.2017.1284220

Accepted author version posted online: 18 Jan 2017.
Published online: 13 Feb 2017.

PSYCHOLOGY, CRIME & LAW, 2017
http://dx.doi.org/10.1080/1068316X.2017.1284220

Routledge
Taylor & Francis Group

Judging guilt and accuracy: highly confident eyewitnesses are discounted when they provide featural justifications

Chad S. Dodson[a] and David G. Dobolyi[b]

[a]Department of Psychology, University of Virginia, Charlottesville, VA, USA; [b]McIntire School of Commerce, University of Virginia, Charlottesville, VA, USA

ABSTRACT

Jurors are heavily swayed by confident eyewitnesses. Are they also influenced by how eyewitnesses justify their level of confidence? Here we document a counter-intuitive effect: when eyewitnesses identified a suspect from a lineup with absolute certainty ('I am completely confident') and justified their confidence by referring to a visible feature of the accused ('I remember his nose'), participants judged the suspect as *less* likely to be guilty than when eyewitnesses identified a suspect with absolute certainty but offered an unobservable justification ('I would never forget him') or no justification at all. Moreover, people perceive an eyewitness's identification as nearly 25% less accurate when the eyewitness has provided a featural justification than an unobservable justification or simply no justification. Even when an eyewitness's level of confidence is clear because s/he has expressed it numerically (e.g. 'I am 100% certain') participants perceive eyewitnesses as not credible (i.e. inaccurate) when the eyewitness has provided a featural justification. However, the effect of featural justifications – relative to a confidence statement only – is maximal when there is an accompanying lineup of faces, moderate when there is a single face and minimal when there is no face at all. The results support our Perceived-Diagnosticity account.

ARTICLE HISTORY
Received 27 October 2016
Accepted 21 December 2016

KEYWORDS
Eyewitness identification; confidence; juror decision-making; eyewitness testimony; credibility

Eyewitness confidence is likely the most important influence on jury decision-making (e.g. Bradfield & Wells, 2000). Not surprisingly, jurors find testimony from highly confident eyewitnesses more compelling than testimony from those who are less confident (e.g. Cutler, Penrod, & Dexter, 1990). But, does it matter how eyewitnesses explain and justify why they are so confident?

There is very little published research on what exactly eyewitnesses say when they see a lineup and make a decision. The only archival study on this this issue is one by Behrman and Richards (2005), who analyzed eyewitness identification statements from 183 actual criminal cases, most of which were armed robberies. The analysis was based on identification statements about a lineup that witnesses saw for the first time. Most witnesses provided verbal expressions of confidence and usually used one of (or variations on) 35

CONTACT Chad S. Dodson cdodson@virginia.edu Department of Psychology, University of Virginia, Charlottesville, VA 22904-4400, USA

different confidence phrases, such as 'I am positive', or 'I am sure he did it' or 'looks familiar' (see Behrman & Richards, 2005, Table 1). Moreover, roughly 30% of eyewitnesses referred to one or more physical features of the perpetrator as a justification for their identification decision.

So, how are people's judgments about the likely guilt of an accused individual influenced by the particular way in which eyewitnesses justify their level of confidence about their identification of a suspect from a lineup? For example, is the identical confidence statement, such as 'I am positive', weighted differently when an eyewitness justifies the confidence statement by referring to a visible feature about the suspect (e.g. 'I remember his chin') than when the eyewitness either refers to an unobservable element (e.g. 'He looks familiar') or provides a confidence statement without any additional justification? Three different outcomes are reasonable.

First, one might expect that the presence and/or kind of explanation for an eyewitness's level of confidence would have little effect on juror judgments of guilt. Much research shows that people are insensitive to a variety of variables that one would expect would influence judgments about guilt (e.g. Semmler, Brewer, & Douglass, 2011). For example, Cutler, Penrod, and Stuve (1988) and Cutler et al. (1990) manipulated 10 different factors that are known to affect eyewitness identification accuracy, such as the delay between the eyewitness's viewing the crime and making an identification. Only one factor influenced juror judgments about guilt: the eyewitness's level of confidence. Moreover, Brewer and Burke (2002) observed that juror judgments about guilt were unaffected when a highly confident eyewitness was shown to make either consistent or inconsistent testimony. That is, judgments about the likely guilt of a suspect were not affected by a highly confident eyewitness contradicting himself (e.g. reporting that the robber said that 'I'll shoot you' and then later reporting that the robber did not threaten me). In addition, Semmler et al. (2011) showed that expressions of eyewitness confidence continue to influence jurors even when they have been warned to evaluate those expressions cautiously. Overall, the foregoing evidence suggests that jurors will not be influenced by how an eyewitness explains why they are highly confident about an identification since they apparently are mainly sensitive to an eyewitness's level of confidence. We call this the Insensitivity account.

On the other hand, if jurors' judgments about guilt are influenced by how eyewitnesses justify their level of confidence then one might expect that they would be swayed more by an eyewitness who offered specific, observable details (e.g. 'I remember his chin') to justify his or her level of confidence than one who did not. After all, individuals do recognize faces more accurately and with more confidence when they can remember specific featural information than when they respond on the basis of mere familiarity (Reinitz, Seguin, Peria, & Loftus, 2012). If jurors are aware of this link, they should be more likely to judge a suspect as guilty if an eyewitness offers a justification that refers to a specific feature of the suspect than an unobservable feature or no justification. We call this the Memory account.

Finally, jurors may respond skeptically to a highly confident eyewitness who claims to have a memory for a specific feature of a suspect picked out of a lineup. This is because most individuals do not have particularly distinctive features. Moreover, ideal (i.e. 'fair') lineups are constructed so that none of the lineup members – including the suspect – has a distinctive feature that would cause that individual to stand out (e.g. Wells & Penrod, 2011).

According to the Perceived-Diagnosticity account, featural justifications (e.g. 'I remember his chin') orient people to consider the memorability of the feature and how well it discriminates between the lineup faces when they attempt to understand the intended meaning of an eyewitness's confidence statement (Dodson & Dobolyi, 2015). Consider an eyewitness who states that he is 'positive' about his identification because he remembers the suspect's chin. When jurors see the suspect, they may note that the feature is not particularly distinctive, which causes them to misinterpret the intended meaning of the confidence statement because they underestimate its level of confidence. By contrast, when an eyewitness justifies his confidence by referring to the suspect's familiarity or to some other unobservable detail, jurors may take this justification at face value because there is no way to evaluate its distinctiveness and how well this unobservable detail discriminates between the lineup faces. So, the Perceived-Diagnosticity account predicts that jurors will find an eyewitness's testimony more compelling when it makes reference to an unobservable than an observable detail (Dodson & Dobolyi, 2015).

Experiment 1

In Experiment 1, participants used eyewitness testimony to judge the likely guilt of an individual selected from a lineup. Participants saw a series of lineups – each consisting of six faces – that included an eyewitness's identification of one of the faces as the suspect. When identifying the suspect, the eyewitness stated his level of confidence and then offered either a featural justification, an unobservable justification, or no justification. Note that our confidence statements (e.g. 'I am pretty sure', 'I would never forget him') are nearly identical to the statements of actual eyewitnesses to an armed robbery (e.g. Behrman & Richards, 2005). The Insensitivity account predicts no effect of the presence or absence of a justification on estimates of guilt. By contrast, the Memory account predicts higher estimates of guilt in the Featural condition than in either the Confidence Only or the Unobservable conditions, whereas the Perceived-Diagnosticity account makes the opposite prediction.

Finally, half of the lineups consisted of black faces and half consisted of white lineups. However, we expected no effects of lineup race on perceived guilt judgments because Dodson and Dobolyi (2015) used these same lineups in their Experiments 2 and 3 and observed no effect of lineup race on participants' judgments about the intended numeric confidence of verbal expressions of confidence.

Method

Participants

Three hundred participants were randomly assigned to one of three conditions (i.e. Confidence Only, Featural Justification, Unobservable Justification) with 100 participants in each condition. Participants were Caucasian adults located in the United States who completed the task over the internet via Amazon's Mechanical Turk (mean age = 29.39, SD = 5.46, 52.00% female). Based on the effect sizes identified in Dodson and Dobolyi (2015), this sample size gives us over 99% power to detect moderately sized effects with an alpha level of 0.05, as determined by G*POWER (Faul, Erdfelder, Lang, & Buchner, 2007).

Materials and procedure

Participants saw eight lineups used previously by Dodson and Dobolyi (2015), each presented as a simultaneous array of six faces (2 row × 3 column), with four black lineups and four white lineups. Each lineup was validated as 'fair' in the sense that (a) no face in each lineup stood out and (b) each lineup face had a roughly equal chance of being selected in the absence of seeing the target (see Dobolyi & Dodson, 2013 for a full description of the lineup fairness methodology). One face within each lineup was randomly chosen as the culprit and was highlighted by a red box, indicating that this face was the eyewitness's choice. Four high confidence statements (e.g. 'I am positive', 'I am completely certain') and four moderately high confidence statements (e.g. 'I am pretty sure', 'I am fairly confident') were randomly assigned to the lineups, with two of each strength assigned to each lineup race.

In addition, eight featural justifications (e.g. 'I remember his chin', 'I remember his nose', 'I recognize his eyes') and eight unobservable justifications (e.g. 'He is very familiar', 'He looks like someone I know', 'I would never forget him') were matched on the average number of words in each justification (i.e. five words). The critical element about the unobservable justifications is that they refer to something that could only be known to the eyewitness. By contrast, the featural justifications refer to details (e.g. chin, nose, etc.) that an observer could evaluate. Each justification was randomly paired with one of the eight confidence statements and to one of the faces from one of the eight lineups. See Figure 2 for the full set of justifications.

Overall, then, there were three between-participant groups. Those in the Confidence Only group read only a confidence statement for each lineup. Those in the Featural Justification and Unobservable Justification groups read both a confidence statement and either a featural justification or an unobservable justification, respectively, for each lineup.

Participants were told that they would see a series of lineups that an eyewitness had seen earlier. They also were instructed that each lineup included both the eyewitness's identification – the highlighted face – and the eyewitness's written expression of certainty about the identification. Lastly, participants were told: 'Based on the eyewitness's above expression of certainty, please use the following numeric scale to provide an estimate of the likelihood that the accused is guilty'. Underneath this statement appeared a six-point scale, ranging from 50% to 100% that was anchored on one end with the label, 'Equally likely guilty or innocent' and on the other end with 'Certain that the accused is guilty'. We demonstrated the task with a practice lineup consisting of different colored smiley faces prior to the eight critical lineups.

Results and discussion

Our key hypothesis is that the content of the eyewitness's justification will affect participants' perception of the likely guilt of the accused individual. We examined this hypothesis with a 3 (Justification Condition: Confidence Only, Featural Justification, or Unobservable Justification) × 2 (Lineup Race: Black vs. White) × 2 (Statement Strength: High Confidence vs. Moderate Confidence) ANOVA of the guilt ratings. Unsurprisingly, there was a main effect of Statement Strength, $F(1, 297) = 458.94$, $MS_e = 117.52$, $p < .0001$, $\eta_p^2 = 0.61$, 95% CI [0.54, 0.66], with higher guilt judgments when eyewitnesses expressed high confidence

($M = 76.34$, SD = 15.18) than moderate confidence ($M = 62.93$, SD = 10.58). There was a main effect of Lineup Race, $F(1, 297) = 11.54$, $MS_e = 57.83$, $p < .001$, $\eta_p^2 = 0.04$, 95% CI [0.01, 0.09], such that suspects from white lineups ($M = 70.38$, SD = 14.66) were assigned slightly higher guilt estimates than were suspects from black lineups ($M = 68.89$, SD = 14.71). More interestingly, there was a main effect of Justification Condition, $F(2, 297) = 14.78$, $MS_e = 395.64$, $p < .0001$, $\eta_p^2 = 0.09$, 95% CI [0.04, 0.15], which was qualified by a Justification Condition × Statement Strength interaction, $F(2, 297) = 31.99$, $MS_e = 117.52$, $p < .0001$, $\eta_p^2 = 0.18$, 95% CI [0.10, 0.25]. There were no other significant effects, all Fs < 2.07.

The interaction between Justification Condition and Confidence Statement Strength is clear in Figure 1. For high confidence statements, participants were much *less* likely to perceive the accused as guilty when eyewitnesses explained their level of confidence by referring to a visible feature ($M = 68.67$, SD = 12.67) than when they referred to either an unobservable element ($M = 81.50$, SD = 13.73) or they provided a confidence statement without any justification ($M = 78.85$, SD = 12.85), $t(198) = 6.86$, $p < .0001$, Cohen's $d = 0.97$, 95% CI [0.61, 1.26] and $t(198) = 5.64$, $p < .0001$, Cohen's $d = 0.80$, 95% CI [0.51, 1.08], respectively. It is worth emphasizing that, as indicated by Cohen's d, this featural justification effect has a large effect size (Cohen, 1988). By contrast, for moderately strong confidence statements, perceptions of guilt were comparable for the Featural Justification ($M = 61.77$, SD = 9.34) and Unobservable Justification ($M = 62.42$, SD = 9.26) conditions, $t(198) = 0.49$, $p = .62$. Guilt judgments were lower in the Featural Justification condition than in the Confidence Only ($M = 64.6$, SD = 9.08) condition, $t(198) = 2.17$, $p = .03$, Cohen's $d = 0.31$, 95% CI [0.03, 0.58], but the size of this effect is relatively small.

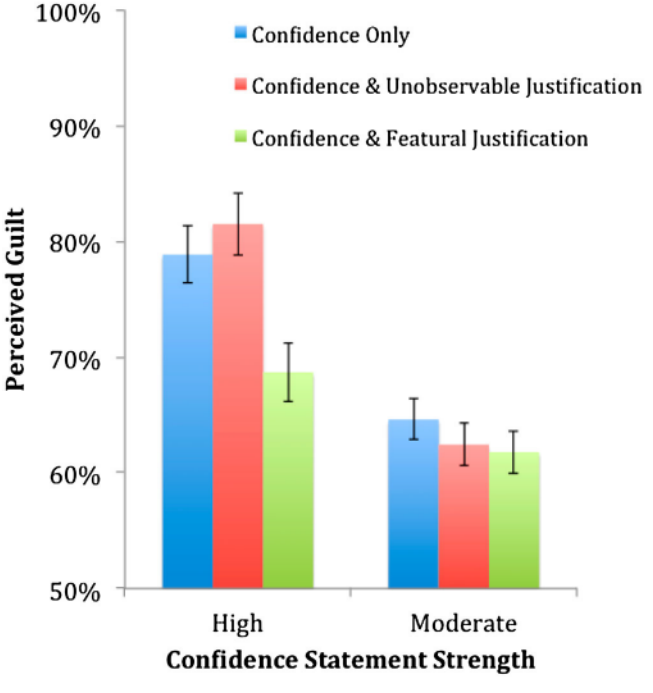

Figure 1. Perceived guilt ratings in Experiment 1 about an accused individual by participants who evaluated either an eyewitness's (a) Confidence Only, (b) Confidence and Unobservable Justification, or (c) Confidence and Featural Justification. Error bars are 95% CI.

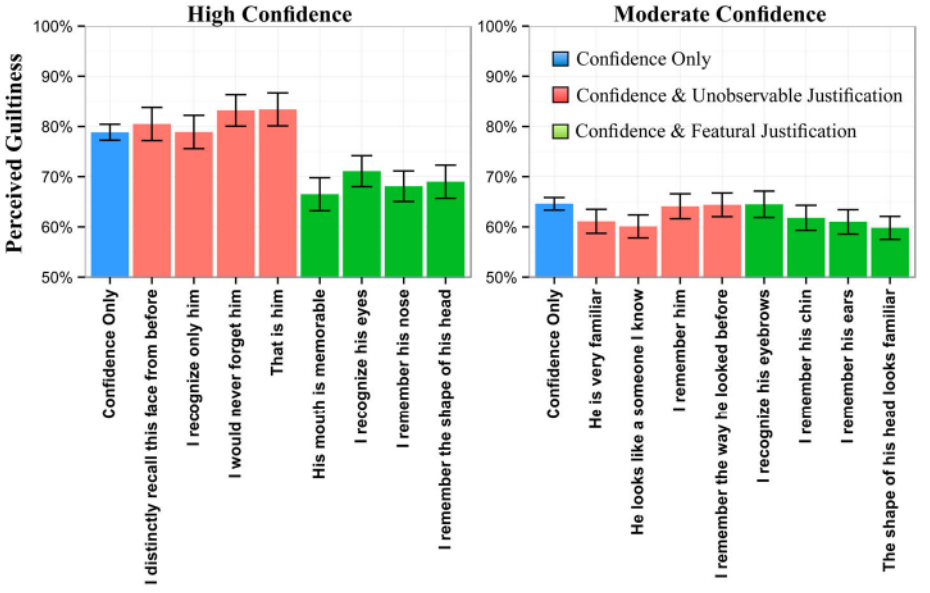

Figure 2. Perceived guilt ratings in Experiment 1 about an accused individual for each justification. Error bars are 95% CI.

Is it possible that the pattern of results is driven by a subset of the justifications? Figure 2 shows perceived guilt ratings for each justification in the Unobservable and the Featural Justification conditions. Note that both the Unobservable and Featural justifications also appeared with either a high (e.g. 'I am positive it's him') or moderate confidence (e.g. 'I am pretty sure it's him') statement but because of space constraints these confidence statements are not shown in the figure (i.e. only justifications are shown). Since the error bars are 95% confidence intervals, conditions with non-overlapping error bars are significantly different from each other. As is clear in the figure, for high confidence statements, every single unobservable justification produced a perceived guilt rating that was significantly higher than that produced by every single featural justification. By contrast, for moderate confidence statements, the ratings for every justification are similar to each other. Thus, the overall pattern of discounting high confidence eyewitnesses when they provide a featural justification – relative to when they either provide an unobservable justification or merely a confidence statement only – generalizes to every single featural and unobservable justification.

Overall, these data support a Perceived-Diagnosticity account of the interpretation of eyewitness statements of confidence. Moreover, these data are consistent with Dodson and Dobolyi (2015) who asked participants to provide a numeric translation for an eyewitness's verbal expression of confidence about a lineup identification. Highly confident eyewitnesses were perceived as less confident when they provided a featural justification than either an unobservable justification or no justification at all.

Experiment 2

We have interpreted the effect of featural justifications as caused by misinterpretation of an eyewitness's intended level of confidence (Dodson & Dobolyi, 2015). For example,

when an eyewitness states he is 'completely certain' and intends this verbal statement to have a numeric meaning of 100% certainty, people are more likely to underestimate this intended level of certainty when the eyewitness provides a featural justification (i.e. people may think he is only 80% certain) than when he provides a confidence statement only without a justification. This underestimation of an eyewitness's intended confidence explains why participants judge an accused as less likely to be guilty in Experiment 1 in the Featural Justification condition than in either the Confidence Only condition or the Unobservable Justification condition. Moreover, much research from other domains shows that people are more likely to misinterpret verbal than numeric expressions of confidence (e.g. Budescu, Broomell, & Por, 2009; Martire, Kemp, Sayle, & Newell, 2014). In fact, we were sufficiently confident in this misinterpretation account that we recommended in Dodson and Dobolyi (2015) that police collect numeric confidence statements from eyewitnesses in order to minimize interpretation errors.

However, there is an alternative explanation that has nothing to do with misunderstanding the eyewitness's intended level of confidence. Participants view eyewitnesses as inaccurate and not very credible when they provide a featural justification. Hearing that an eyewitness was completely certain because he remembered that the suspect's 'mouth was memorable' may indicate that the eyewitness's identification is inaccurate if the suspect's mouth does not appear to be distinctive and does not seem to discriminate the suspect from the other members of the lineup. Viewing the eyewitness's identification as inaccurate may in turn cause participants to view the eyewitness as less confident. Asking participants to estimate the intended numeric confidence of the eyewitness, as we did in Dodson and Dobolyi (2015), cannot distinguish between these two accounts of the featural justification effect: (1) misunderstanding of the eyewitness's intended level of confidence or (2) viewing the eyewitness as not credible and as inaccurate.

By contrast, asking participants to estimate perceived guilt can distinguish between the Misunderstanding and the Not-Credible accounts. To that end, we presented confidence statements either verbally (e.g. 'I am completely certain') or numerically (e.g. 'I am 100% certain') in Experiment 2. As in Experiment 1, we provided an eyewitness's statement of confidence either by itself or accompanied by either a featural justification or an unobservable justification.

If featural justifications cause participants to misunderstand the eyewitness's intended level of confidence then the featural justification effect on perceived guilt – that is, reduced guilt judgments to highly confident statements in the Featural Justification condition than in the Confidence Only condition – should be eliminated or greatly reduced when eyewitness statements of confidence are expressed numerically. Thus, the Misunderstanding account predicts no difference in perceived guilt ratings between the Confidence Only and the Featural Justification conditions in the numeric confidence statement condition because the eyewitness's intended level confidence should be obvious. There should be little chance of misunderstanding the eyewitness's level of confidence when he says he is 100% confident. By contrast, if featural justifications cause participants to view the witness as not credible and inaccurate then the featural justification effect with high confidence statements (i.e. lower guilt ratings in the Featural than in the Confidence Only conditions) should persist regardless of whether confidence is expressed numerically or verbally.

Method

Participants

Four hundred and eighty participants were randomly assigned in equal numbers to one of six between-participants conditions that were produced by two between-participants factors of (a) justification condition (i.e. Confidence Only, Featural Justification, Unobservable Justification) and (b) verbal vs. numeric presentation of the confidence statement. Participants were located in the United States and completed the task on either a desktop or laptop computer over the internet via Amazon's Mechanical Turk (mean age = 34.75, SD = 11.34, 42.71% female). Also, six participants were replaced because they responded '1' to all lineups, indicating a lack of seriousness. Based on the effect sizes from Experiment 1, this sample size gives us over 99% power to detect moderately sized effects with an alpha level of 0.05, as determined by G*POWER (Faul et al., 2007).

Materials and procedure

We used the identical materials, procedure and design from Experiment 1, except for the following changes. First, we manipulated how confidence statements were presented. In the numeric confidence conditions, we use 100% and 80%, respectively, to express high and moderate confidence (e.g. 'I am 100% certain', 'I am 80% confident'). By contrast, the verbal confidence conditions are a close replication of Experiment 1 in which we expressed high and moderate confidence with words (e.g. 'I am absolutely certain', 'I am fairly confident'). Second, we altered the perceived guilt scale. We did not want the six points on this scale to be labeled 50–100% because this would overlap with our use of 80% and 100% to express numeric certainty. So, instead, we labeled the perceived guilt scale from 1 to 6. Third, we replaced one of the featural justifications, 'The shape of his head looks familiar', with the statement, 'I remember his neck', so that the phrase, 'shape of his head', would no longer appear in two different justifications. All other details were identical to those in Experiment 1.

Results and discussion

Our central prediction involves guilt ratings in the Confidence Only and Featural Justification conditions to lineups where high confidence is expressed either numerically or verbally. We examined this prediction with a 2 (Number/Verbal Confidence) × 3 (Justification Condition: Confidence Only, Featural Justification, or Unobservable Justification) × 2 (Lineup Race: Black vs. White) × 2 (Confidence Statement Strength: High Confidence vs. Moderate Confidence) ANOVA of the guilt ratings. There were main effects of Number/Verbal Confidence, $F(1, 474) = 16.14$, $MS_e = 3.24$, $p < .0001$, $\eta_p^2 = 0.03$, 95% CI [0.01, 0.06], Confidence Statement Strength, $F(1, 474) = 614.66$, $MS_e = 0.91$, $p < .0001$, $\eta_p^2 = 0.56$, 95% CI [0.52, 0.60], Justification Condition, $F(2, 474) = 19.24$, $MS_e = 3.24$, $p < .0001$, $\eta_p^2 = 0.08$, 95% CI [0.04, 0.11], which were all qualified by a number of interactions.

With respect to our central predictions, the relevant interaction is one between Number/Verbal Confidence × Confidence Statement Strength × Justification Condition, $F(2, 474) = 3.39$, $MS_e = 0.91$, $p = .03$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.03], which moderates the two-way interactions between (a) Confidence Statement Strength × Justification

Condition, $F(2, 474) = 12.28$, $MS_e = 0.91$, $p < .0001$, $\eta_p^2 = 0.05$, 95% CI [0.02, 0.08], and (b) Confidence Statement Strength × Number/Verbal Confidence, $F(1, 474) = 10.61$, $MS_e = 0.91$, $p = .001$, $\eta_p^2 = 0.02$, 95% CI [0.01, 0.05].

Figure 3 shows that when high confidence is expressed verbally we replicated Experiment 1 and observed lower guilt ratings in the Featural Justification condition ($M = 4.14$, $SD = 1.29$) than in the Confidence Only condition ($M = 4.99$, $SD = 1.06$), $t(318) = 6.41$, $p < .0001$, Cohen's $d = 0.72$, 95% CI [0.49, 0.94]. But, in support of the Not-Credible account and contrary to the Misunderstanding account, when high confidence is expressed numerically, we also observed lower guilt ratings in the Featural than in the Confidence Only conditions ($M = 4.16$, $SD = 1.22$ vs. $M = 5.08$, $SD = 1.20$), $t(318) = 6.80$, $p < .0001$, Cohen's $d = 0.76$, 95% CI [0.53, 0.99]. The effect size for this featural justification effect is nearly identical in the numeric and verbal confidence statement conditions (Cohen's $d = 0.76$ vs. 0.72) – and both of these effect sizes are comparable to what we observed in Experiment 1 for this Featural vs. Confidence Only contrast (i.e. Cohen's $d = 0.80$). In addition, as in Experiment 1, guilt ratings were lower in the verbal condition to featural justifications than to unobservable justifications ($M = 4.14$, $SD = 1.29$ vs. $M =$

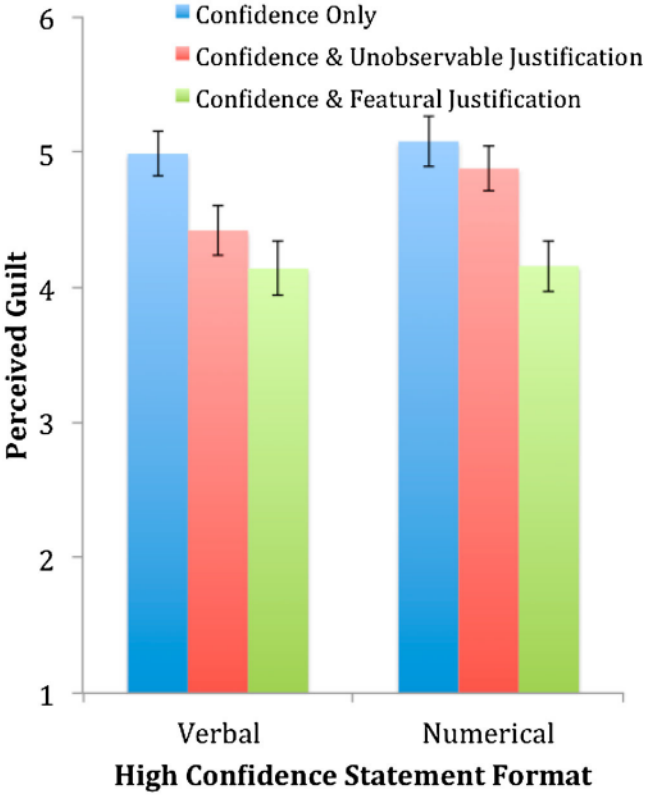

Figure 3. In the high confidence condition in Experiment 2, perceived guilt ratings about an accused individual by participants who evaluated either an eyewitness's (a) Confidence Statement Only, (b) Confidence and Unobservable Justification, or (c) Confidence and Featural Justification. Confidence was expressed either verbally (e.g. 'I am very confident') or numerically (e.g. 'I am 100% confident'). Error bars are 95% CI.

4.42, SD = 1.18), $t(318) = 2.03$, $p = .04$, Cohen's $d = 0.23$, 95% CI [0.01, 0.45], although inexplicably the effect size is smaller than what we observed for this contrast in Experiment 1. In the numeric condition, guilt ratings were also lower in the Featural condition than in the Unobservable condition, ($M = 4.16$, SD = 1.22 vs. $M = 4.88$, SD = 1.09), $t(318) = 5.56$, $p < .0001$, Cohen's $d = 0.62$, 95% CI [0.40, 0.85]. Overall, the featural justification effect on guilt judgments occurs when high confidence is expressed either numerically or verbally – supporting the Not-Credible account.

For the moderately strong confidence statements, the verbal and numerical conditions replicate what we observed in Experiment 1, although as is clear in Figure 4 guilt ratings are lower in all justification conditions for verbal statements than for numeric statements. For verbal statements of confidence, perceptions of guilt were lower in the Featural Justification ($M = 3.10$, SD = 1.04) condition than in the Confidence Only ($M = 3.49$, SD = 0.89) condition, $t(318) = 3.64$, $p = .0003$, Cohen's $d = 0.41$, 95% CI [0.18, 0.63]. Guilt judgments were comparable for the Featural Justification and Unobservable Justification ($M = 3.30$, SD = 1.10) conditions, $t(318) = 1.65$, $p = .10$. Likewise, for numeric statements of confidence, there were lower guilt ratings in the Featural Justification condition ($M = 3.62$,

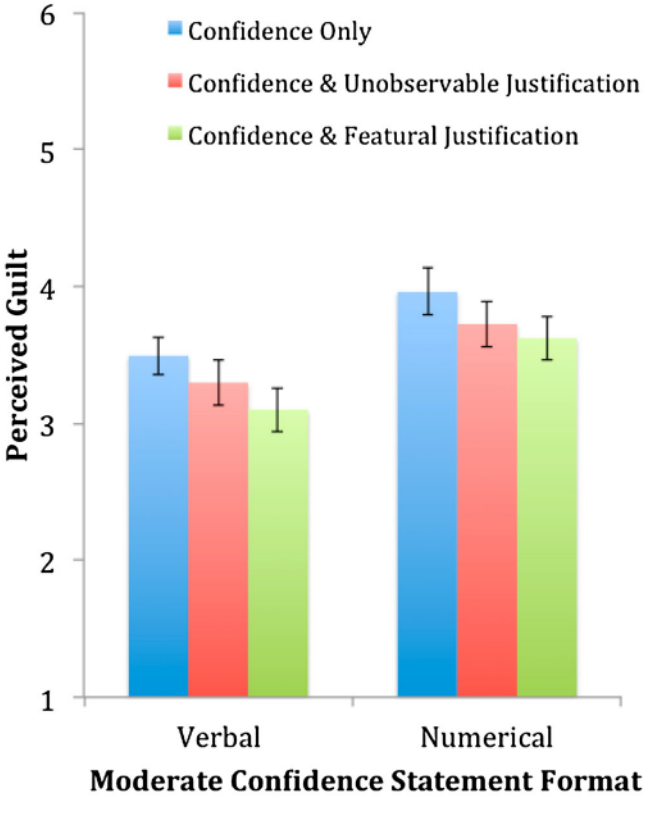

Figure 4. In the moderate confidence condition in Experiment 2, perceived guilt ratings about an accused individual by participants who evaluated either an eyewitness's (a) Confidence Only, (b) Confidence and Unobservable Justification, or (c) Confidence and Featural Justification. Confidence was expressed either verbally (e.g. 'I am very certain') or numerically (e.g. 'I am 100% certain'). Error bars are 95% CI.

SD = 1.02) than in the Confidence Only condition, (M = 3.96, SD = 1.11) condition, $t(318)$ = 2.82, p = .005, Cohen's d = 0.32, 95% CI [0.09, 0.54]. But, these ratings were similar for the Featural and Unobservable Justification conditions (M = 3.73, SD = 1.10), $t(318)$ = 0.87, p = .39.

The ANOVA also produced a main effect of Lineup Race, $F(1, 474) = 48.87$, $MS_e = 0.45$, $p < .0001$, $\eta_p^2 = 0.09$, 95% CI [0.06, 0.14], which was qualified by two interactions. The interaction between Lineup Race and Statement Confidence, $F(1, 474) = 57.37$, $MS_e = 0.29$, $p < .0001$, $\eta_p^2 = 0.11$, 95% CI [0.07, 0.15] shows that white suspects (M = 4.63, SD = 1.22) and black suspects (M = 4.60, SD = 1.25) were given comparable guilt estimates when the eyewitness expressed high confidence, $t(479) = 0.68$, p = .50. By contrast, when the eyewitness was moderately confident in the identification, white suspects (M = 3.73, SD = 1.03) were given higher guilt estimates than were black suspects (M = 3.33, SD = 1.10), $t(479) = 10.67$, $p < .0001$, Cohen's d = 0.49, 95% CI [0.39, 0.58]. There was also an interaction between Lineup Race and Number/Verbal Confidence, $F(1, 474) = 5.48$, $MS_e = 0.45$, p = .02, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.03]. As shown by the effect sizes, there was a larger difference in perceptions of guilt of the white and black suspects when confidence is expressed verbally (white M = 4.05, SD = 1.22 vs. black M = 3.76, SD = 1.34), $t(479) = 6.47$, $p < .0001$, Cohen's d = 0.22, 95% CI [0.14, 0.31] than when confidence is expressed numerically (white M = 4.31, SD = 1.20 vs. black M = 4.17, SD = 1.30), $t(479) = 3.99$, p = .0001, Cohen's d = 0.11, 95% CI [0.03, 0.19]. However, the effect-sizes are small in both of these contrasts. In the General Discussion, we discuss the effects of race on judgments of guilt, given the consistent effects of race in all of the experiments.

Overall, this experiment replicates the key findings from Experiment 1 and, importantly, it clearly disconfirms the Misunderstanding account and supports the Not-Credible account of the featural justification effect.

Experiment 3

Although Experiment 2 ruled out the Misunderstanding account, its support for the Not-Credible account is essentially just the default: Experiment 2 provides no direct evidence that participants actually perceive eyewitnesses as less credible and less accurate when their identification is based on a featural justification than either no justification or an unobservable justification. The purpose of Experiment 3 was to test the predictions of the Not-Credible account by asking participants to judge the likely accuracy of an eyewitness's identification. This experiment is a critical test of the mechanism underlying the effect of featural justifications on understanding confidence: do featural justifications cause people to misinterpret the intended meaning of the confidence statement or do featural justifications cause people to consider the identification to be inaccurate, which in turn causes people to discount confidence.

If the Not-Credible account is correct then participants should perceive an eyewitness's identification as less accurate when the identification is accompanied by a featural justification than by either no justification or an unobservable justification. Overall, besides instructing participants to estimate the likely accuracy of the eyewitness's identification, Experiment 3 is nearly identical in all other ways to the numeric confidence conditions in Experiment 2.

Method

Participants

Two hundred and forty participants were randomly assigned in equal numbers to one of three between-participants conditions: Confidence Only, Featural Justification, or Unobservable Justification. Participants were located in the United States and completed the task on either a desktop or laptop computer over the internet via Amazon's Mechanical Turk (mean age = 37.22, SD = 12.13, 48.75% female). This sample size gave us over 99% power to detect moderately sized effects with an alpha level of.05, as determined by G*POWER (Faul et al., 2007).

Materials and procedure

We used the identical materials and procedure from the numeric confidence conditions in Experiment 2, except for the following changes: first, instead of asking participants to judge guilt, we asked them to judge the likely accuracy of the eyewitness's identification. Specifically, we stated, 'Given the eyewitness's above expression of certainty, how accurate is the eyewitness's identification?' Underneath this statement we provided a six-point numeric scale, ranging from 0% to 100% with the anchors of 'Not at All Accurate' and 'Completely Accurate', respectively. All other details were identical to the numeric conditions in Experiment 2. The overall design of this experiment is 3 (Justification Condition: Confidence Only, Featural Justification, Unobservable Justification) × 2 (Statement Strength: High Confidence vs. Moderate Confidence) × 2 (Lineup Race: Black vs. White), with the first factor between-participants and the remaining two factors within-participants.

Results and discussion

We examined participants' estimates of the likely accuracy of the eyewitness's identification with a 3 (Justification Condition: Confidence Only, Featural Justification, or Unobservable Justification) × 2 (Lineup Race: Black vs. White) × 2 (Statement Strength: High Confidence vs. Moderate Confidence) ANOVA. As in the earlier experiments, there was a main effect of Statement Strength, $F(1, 237) = 324.17$, $MS_e = 177.87$, $p < .0001$, $\eta_p^2 = 0.58$, 95% CI [0.51, 0.63], with higher perceived accuracy when eyewitnesses expressed high confidence ($M = 79.08$, SD = 23.06) than moderate confidence ($M = 63.58$, SD = 19.74). Critically, the main effect of Justification Condition, $F(2, 237) = 29.28$, $MS_e = 1175.18$, $p < .0001$, $\eta_p^2 = 0.20$, 95% CI [0.12, 0.27] was qualified by the interaction between Justification Condition and Statement Strength, $F(2, 237) = 11.90$, $MS_e = 177.87$, $p < .0001$, $\eta_p^2 = 0.09$, 95% CI [0.04, 0.15].

As shown in Figure 5, participants perceived eyewitnesses as nearly 25% less accurate when a highly confident identification was accompanied by a featural justification (M = 64.19, SD = 27.16) than by either a confidence statement only ($M = 87.38$, SD = 18.03) or an unobservable justification ($M = 85.69$, SD = 14.26), $t(318) = 9.00$, $p < .0001$, Cohen's d = 1.01, 95% CI [0.77, 1.24] and $t(318) = 8.87$, $p < .0001$, Cohen's $d = 0.99$, 95% CI [0.76, 1.22], respectively. Perceived accuracy was comparable in the latter two conditions, $t(318) = 0.93$, $p = .35$. The Justification × Confidence Statement Strength interaction is

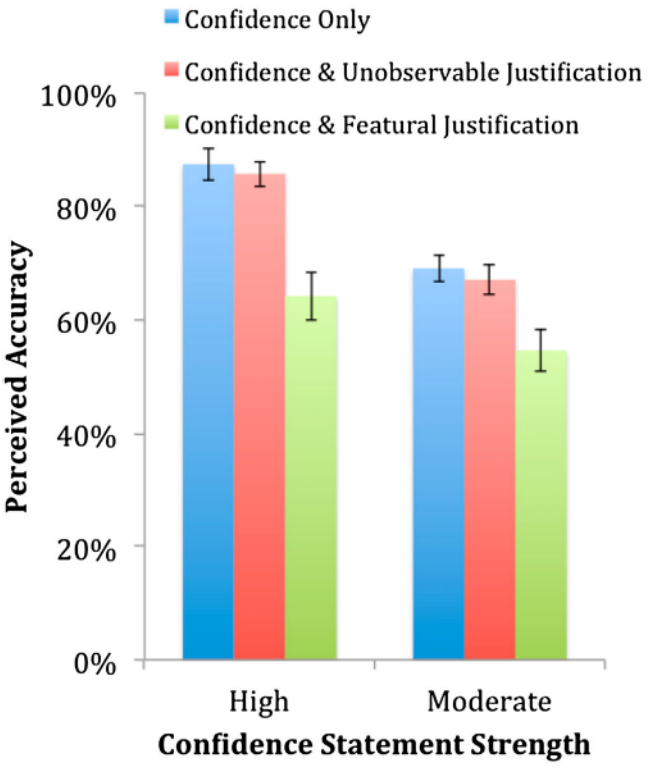

Figure 5. Perceived accuracy of the eyewitness's identification in Experiment 3 when the eyewitness provided either (a) Confidence Statement Only, (b) Confidence and Unobservable Justification, or (c) Confidence and Featural Justification. Confidence was expressed numerically only (e.g. 'I am 100% sure'). Error bars are 95% CI.

produced because the same pattern occurs in the moderate confidence condition but the effect-sizes are not as large. When eyewitnesses were moderately confident about their identification, participants perceived the eyewitness as less accurate when the identification was justified with a featural justification ($M = 54.62$, SD = 23.52) than with a confidence statement only ($M = 69.06$, SD = 14.83) or an unobservable justification ($M = 67.06$, SD = 16.66), $t(318) = 6.57$, $p < .0001$, Cohen's $d = 0.73$, 95% CI [0.51, 0.96] and $t(318) = 5.46$, $p < .0001$, Cohen's $d = 0.61$, 95% CI [0.38, 0.83], respectively. Perceived accuracy was similar in the confidence statement only and in the unobservable justification conditions, $t(318) = 1.13$, $p = .26$.

The ANOVA also produced a main effect of Lineup Race, $F(1, 237) = 25.20$, $MS_e = 111.19$, $p < .0001$, $\eta_p^2 = 0.10$, 95% CI [0.04, 0.16], which was qualified by two interactions. The interaction between Lineup Race and Statement Confidence, $F(1, 237) = 9.36$, $MS_e = 71.26$, $p < .002$, $\eta_p^2 = 0.04$, 95% CI [0.01, 0.09] shows that eyewitnesses were perceived as comparably accurate when they identified white suspects ($M = 79.96$, SD = 22.68) and black suspects ($M = 78.21$, SD = 23.46) with high confidence, $t(239) = 1.83$, $p = .07$. But, when eyewitnesses were moderately confident in their identification, they were perceived as more accurate when they identified a white suspect ($M = 66.12$, SD = 18.62) than a black suspect ($M = 61.04$, SD = 20.52), $t(239) = 6.42$, $p < .001$, Cohen's $d = 0.26$, 95% CI [0.08, 0.44], although this is a small effect.

There was also an interaction between Lineup Race and Justification Condition, $F(2, 237) = 4.03$, $MS_e = 111.19$, $p = .02$, $\eta_p^2 = 0.03$, 95% CI [0.00, 0.07]. Regardless of the race of the suspect, eyewitness identifications were perceived as similarly accurate in the Unobservable Justification condition (white $M = 77.12$, $SD = 16.84$ vs. black $M = 75.62$, $SD = 19.25$), $t(159) = 1.49$, $p = .14$, and the Featural Justification condition (white $M = 60.75$, $SD = 25.39$ vs. black $M = 58.06$, $SD = 26.24$), $t(159) = 2.17$, $p = .03$, Cohen's $d = 0.10$. However, eyewitnesses were perceived as more accurate in the Confidence Only condition when they identified a white suspect ($M = 81.25$, $SD = 16.55$) than a black suspect ($M = 75.19$, $SD = 20.53$), $t(159) = 6.35$, $p < .001$, Cohen's $d = 0.33$, 95% CI [0.10, 0.54]. We discuss the effects of race on perceived accuracy in the General Discussion.

Overall, this experiment shows that eyewitnesses are perceived as nearly 25% less accurate when they provide a featural justification for their identification than when they provide either an unobservable justification or merely a confidence statement without a justification.

Experiment 4

According to our Perceived-Diagnosticity account, the reason why featural justifications decrease an eyewitness's perceived accuracy is because observers can see the referenced feature on the face (e.g. 'I remember his nose') and note that the feature is not particularly diagnostic – it does not distinguish the suspect from the other members of the lineup. Observers then conclude that the identification is less likely to be accurate when it is based on a featural justification than either no justification or an unobservable one. Dodson and Dobolyi (2015) showed that this featural justification effect was larger when participants viewed a lineup of faces than a single face. From the perspective of our Perceived-Diagnosticity account, it makes sense that there would be a larger effect when the featural justification refers to a face within a lineup than a face by itself because the lineup encourages participants to evaluate the relative distinctiveness of the feature on the referenced face as compared to the other faces in the lineup.

However, there is an alternative explanation for the featural justification effect: regardless of the presence or absence of a face, featural justifications are always perceived as weak justifications which cause participants to perceive the eyewitness as less accurate relative to situations when the eyewitness provides either an unobservable justification or no justification (confidence statement only).

These two accounts make different predictions about the effect of a featural justification on the perceived accuracy of an eyewitness when there is either an accompanying lineup of faces, a single face or no face at all and only the justification. According to our Perceived-Diagnosticity account, the effect of featural justifications greatly depends on the presence of an accompanying face. This account predicts that participants will view an eyewitness as less accurate when a featural justification is accompanied by a lineup of faces than a single face – conceptually replicating Dodson and Dobolyi (2015). And, critically, participants should view eyewitnesses as more accurate when the featural justification is accompanied by no face at all (only the justification itself) than when there is either a lineup of faces or a single face. In sum, when it is impossible to evaluate the diagnosticity of the featural justification because of the absence of an accompanying face then the

Perceived-Diagnosticity account predicts that participants will accept the featural justification at face value and assume it justifies the level of confidence.

By contrast, the alternative account predicts that participants will view eyewitnesses as comparably inaccurate in all conditions that involve a featural justification, regardless of whether or not there is an accompanying face. Again, according to this account, featural justifications are weak pieces of evidence and the effect of this justification is uninfluenced by an accompanying face.

To test both of these accounts, we instructed participants to judge the likely accuracy of an eyewitness and we manipulated two factors. First, we manipulated whether participants saw either (a) a lineup of faces, as in the previous experiments, (b) a single face, as in Dodson and Dobolyi (2015), or critically, (c) no face at all. Thus, the no face condition serves as an important control condition for measuring how people interpret the confidence statement and justification in the absence of an accompanying face. Second, we manipulated whether participants saw either a confidence statement only or both a confidence statement and a featural justification. We did not include the Unobservable Justification condition to simplify our design and because our predictions focus on the Featural Justification condition.

Method

Participants

Five hundred and forty-six participants were randomly and in equal numbers assigned to one of six between-participants conditions that were formed by crossing the between-participant factors of Justification Condition (Confidence Only vs. Featural Justification) and Lineup Format (Lineup, Face Only, No Face). As in the previous experiments, participants were located in the United States and completed the task on either a desktop or laptop computer over the internet via Amazon's Mechanical Turk (mean age = 36.99, SD = 11.50, 51.65% female). Overall, this sample size gave us over 99% power to detect moderately sized effects with an alpha level of 0.05, as determined by G*POWER (Faul et al., 2007).

Materials and procedure

As in Experiment 3, participants judged the likely accuracy of an eyewitness's numeric expression of confidence about an identification. The materials and procedure were identical to what we used in Experiment 3, except for two differences. First, we used only the Confidence Only and the Featural Justification conditions. Second, in addition to the eyewitness's confidence statement, participants saw either (a) a lineup of faces, as in the previous experiments, (b) a single face, or (c) no face at all and only the confidence statement. All other details were identical to those in Experiment 3.

Results and discussion

We used a 2 (Justification Condition: Confidence Only, Featural Justification) × 3 (Lineup Format: Lineup, Face Only, No Face) × 2 (Statement Strength: High Confidence vs.

Moderate Confidence) × 2 (Lineup Race: Black vs. White) ANOVA to examine participants' estimates of the likely accuracy of the eyewitness's identification. Unsurprisingly, there was a main effect of confidence Statement Strength, $F(1, 540) = 901.10$, $MS_e = 166.65$, $p < .0001$, $\eta_p^2 = 0.63$, 95% CI [0.58, 0.66], with higher perceived accuracy when eyewitnesses expressed high confidence ($M = 79.66$, SD = 22.31) than moderate confidence ($M = 63.08$, SD = 18.94) in their identification.

More interestingly, there was an interaction between Justification Condition and Lineup Format, $F(2, 540) = 5.47$, $MS_e = 1094.27$, $p < .01$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.03], which qualifies the main effects of Justification Condition, $F(1, 540) = 89.80$, $MS_e = 1094.27$, $p < .0001$, $\eta_p^2 = 0.14$, 95% CI [0.09, 0.20], and Lineup Format, $F(2, 540) = 13.98$, $MS_e = 1094.27$, $p < .0001$, $\eta_p^2 = 0.05$, 95% CI [0.02, 0.09]. As shown in Figure 6, the effect of featural justifications clearly depends on the presence or absence of a face or lineup of faces. The left side of Figure 6 shows that when participants saw a confidence statement only they generally perceived the eyewitness as comparably accurate in the Lineup ($M = 76.87$, SD = 18.68), Single Face ($M = 77.28$, SD = 18.21) and No Face ($M = 80.08$, SD = 15.93) conditions. Although perceived accuracy in the No Face condition was significantly higher than perceived accuracy in the other two conditions, both t's (726) > 2.20, p's < .05, both Cohen's d's < 0.18, the size of these effects is very small, as shown by the Cohen's d's. By contrast, when participants saw the confidence statement accompanied by a featural justification, they perceived the eyewitness as 10% and 15% *less* accurate when they also saw either a single face ($M = 62.80$, SD = 23.93) or a lineup of faces ($M = 58.32$, SD = 24.93) than when

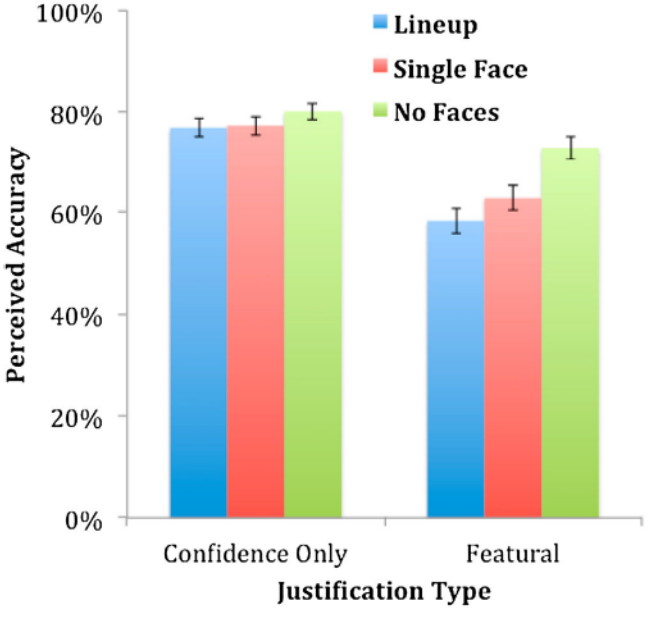

Figure 6. Perceived accuracy of an eyewitness's identification in Experiment 4 when the eyewitness provided either a Confidence Statement Only or a Confidence Statement and a Featural Justification about a suspect that was shown either (a) in a Lineup, (b) as a Single Face, or (c) was not shown – no faces were presented. Confidence was expressed numerically (e.g. 'I am 100% sure'). Error bars are 95% CI.

there were no faces present, ($M = 72.86$, SD = 21.71), both t's (726) > 5.93, p's < .001, both Cohen's d's > 0.46.

Featural justifications have a greater effect on perceived accuracy when faces are present than absent. This pattern is consistent with our Perceived-Diagnosticity account. A featural justification is evaluated in terms of how well it identifies the face (e.g. is it plausible that this is a memorable nose?) and discriminates the face from the lineup of faces (e.g. is it plausible that this chin would distinguish this face from the other faces?). These results rule out the alternative hypothesis that featural justifications are viewed as weak evidence in all situations.

As in the previous experiments, there was an interaction between Justification Condition and Confidence Statement Strength, $F(1, 540) = 18.72$, $MS_e = 166.65$, $p < .0001$, $\eta_p^2 = 0.03$, 95% CI [0.01, 0.07], which reflects a stronger (i.e. larger effect size) featural justification effect when eyewitnesses were highly than moderately confident about their identification. Participants perceived eyewitnesses as less accurate when they were highly confident and provided a featural justification ($M = 71.76$, SD = 25.13) than a confidence statement only ($M = 87.56$, SD = 15.47), $t(1090) = 12.51$, $p < .0001$, Cohen's $d = 0.76$, 95% CI [0.63, 0.88]. While the same pattern occurred when eyewitnesses were moderately confident and provided a featural justification ($M = 57.56$, SD = 21.23) than a confidence statement only ($M = 68.59$, SD = 14.37), $t(1090) = 10.05$, $p < .0001$, Cohen's $d = 0.61$, 95% CI [0.49, 0.73], the effect size for this difference was not as large as when eyewitnesses were highly confident.

As in Experiment 3, there was a main effect of Suspect Race, $F(1, 540) = 29.96$, $MS_e = 110.60$, $p < .0001$, $\eta_p^2 = 0.05$, 95% CI [0.02, 0.09], which was qualified by three interactions. The interaction between Suspect Race and Lineup Format, $F(2, 540) = 10.94$, $MS_e = 110.60$, $p < .0001$, $\eta_p^2 = 0.04$, 95% CI [0.01, 0.07], shows that there was no effect of suspect race on perceived accuracy of the eyewitness's identification when participants viewed either no faces ($M = 76.15$, SD = 19.26 vs. $M = 76.79$, SD = 19.49), $t(363) = 1.23$, $p = .22$, or a single face (White suspect: $M = 70.71$, SD = 21.62 vs. Black suspect: $M = 69.37$, SD = 23.27), $t(363) = 1.82$, $p = .07$. However, when participants viewed a lineup of faces then eyewitnesses were perceived as more accurate when they identified a white suspect ($M = 70.30$, SD = 22.57) than a black suspect ($M = 64.89$, SD = 24.88), $t(363) = 6.26$, $p < .0001$, Cohen's $d = 0.24$, 95% CI [0.14, 0.34].

There also was an interaction between Suspect Race and Confidence Statement Strength, $F(1, 540) = 20.42$, $MS_e = 75.11$, $p < .0001$, $\eta_p^2 = 0.04$, 95% CI [0.01, 0.07], which is qualified by the interaction between Suspect Race, Confidence Statement Strength, and Justification Type, $F(1, 540) = 7.51$, $MS_e = 75.11$, $p < .01$, $\eta_p^2 = 0.01$, 95% CI [0.00, 0.04]. As seen in Table 1, this triple interaction reflects the pattern that participants always perceive eyewitnesses as more accurate when they identify a white suspect than a black suspect, all t's (272) > 2.87, all p's < .004, except for the one instance when eyewitnesses are highly confident and provide a featural justification, $t(272) = 0.58$, $p = .56$.

General discussion

There are three important findings from these studies. First, we show that there are large consequences to how eyewitnesses explain their level of confidence in an identification. When highly confident eyewitnesses provide a featural justification (e.g. 'I remember his

Table 1. Perceived accuracy in Experiment 4 of an eyewitness's moderate confidence or high confidence identification when the eyewitness identified a black or white suspect and provided either a confidence statement only or a confidence statement and a featural justification.

	Suspect Race			
	Black		White	
Condition	*M*	(SD)	*M*	(SD)
Confidence Only, Moderate	66.85	(15.52)	70.33	(12.93)
Confidence Only, High	86.48	(15.84)	88.65	(15.05)
Featural, Moderate	55.16	(22.11)	59.96	(20.06)
Featural, High	72.05	(25.32)	71.46	(24.99)

hair') for their identification, Experiments 1 and 2 show that participants are much less likely to judge the accused as guilty as compared to when a highly confident eyewitness provides either no justification or an unobservable justification (e.g. 'He looks like someone I know') – a phenomenon that we call the featural justification effect.

The second finding is that, contrary to our expectations and conclusions in Dodson and Dobolyi (2015), Experiments 2, 3, and 4 show that the featural justification effect is not caused by participants misunderstanding the eyewitness's intended level of confidence in the identification. If this misunderstanding account had been correct then when eyewitness confidence was expressed numerically (e.g. 'I am 100% sure') so that there is little possibility of misunderstanding the level of confidence there should have been no (or least a greatly reduced) featural justification effect on guilt judgments. Instead, Figure 3 shows that the featural justification effect is just as large when confidence is presented numerically as it is when confidence is presented verbally.

The third significant finding is that a featural justification reduces the perceived accuracy of an eyewitness's identification. Experiment 3 shows that even when eyewitnesses express absolute certainty in their identification (e.g. 'I am 100% sure'), they are viewed as nearly 25% less accurate when they justified their level of confidence by referring to a visible feature about the suspect than when they provided a confidence statement only and no justification. Finally, Experiment 4 shows that the effect of featural justifications greatly depends on the presence of a lineup of faces (or even just a single face). Participants perceive eyewitnesses as much less accurate when an eyewitness's featural justification is accompanied by a lineup of faces than no face at all.

Altogether, the experiments support a Perceived-Diagnosticity account (Dodson & Dobolyi, 2015). Featural justifications are evaluated in terms of the perceived distinctiveness of the referenced feature on the suspect's face. When the feature is perceived as not particularly diagnostic because it does not appear to discriminate the suspect from the other lineup faces then eyewitnesses are viewed as inaccurate and not credible. We suggest that changes in perceived accuracy and credibility are the root cause of the featural justification effects that we have observed here and in Dodson and Dobolyi (2015). It is this reduction in perceived accuracy as a result of the featural justification that causes participants (a) to perceive the witness as less confident and (b) to give the witness's testimony less weight when determining a suspect's guilt, as compared to when a witness has provided either no justification (confidence statement only) or an unobservable justification.

One might question why an eyewitness would ever cite a non-distinctive feature as a justification for an identification. The answer, however, involves the difference in

experience and perspective by the eyewitness, relative to jurors, police and others. By virtue of prior experience with the suspect, the eyewitness may readily view the feature as distinctive. By contrast, when police and jurors see the suspect for the first time in the context of a fair lineup, they may view the referenced feature as not particularly memorable or distinctive. It is this difference in perspective that can cause eyewitnesses to cite a feature as justification for an identification and yet others may perceive the feature as not particularly diagnostic.

Further support for our Perceived-Diagnosticity account comes from Cash and Lane (in press) who manipulated the distinctiveness of a feature on the suspect's face. In one condition, they showed participants biased lineups in which the suspect had a facial feature, such as very bushy eyebrows, that clearly discriminated the suspect from the other lineup faces. When eyewitnesses provided a featural justification that referred to this objectively distinctive feature then participants perceived the eyewitness as similarly confident as compared to the Confidence Only condition – that is, no featural justification effect on perceived confidence. In another condition, Cash and Lane showed participants 'fair lineups' in which the suspect did not have a distinctive facial feature. When the featural justification referred to a non-distinctive feature then, replicating Dodson and Dobolyi (2015), eyewitnesses were perceived as less confident than in the Confidence Only condition. These results support our view that individuals interpret featural justifications in terms of how well the feature discriminates the referenced face from other lineup faces (i.e. the feature's diagnostic value) or in terms of the perceived memorability of the feature, as in the single face condition. One question deserving further research is whether the featural justification effect will occur when eyewitnesses provide holistic descriptors, such as 'His face reminded me of a rat', or 'He was very attractive'.

A consistent finding in all of the experiments is that the race of the suspect influences participants' judgments about the suspect's guilt and their perceptions about the accuracy of the eyewitness's identification. All four experiments showed a main effect of the suspect's race, which was qualified in Experiments 2, 3, and 4 by the eyewitness's level of confidence in the identification. When eyewitnesses were highly confident in their identification then participants perceived white and black suspects as comparably guilty (i.e. in Experiment 2) and they perceived eyewitnesses as comparably accurate in their identification (i.e. in Experiment 3). Experiment 4 suggests that it is only when highly confident eyewitnesses provide a featural justification that they are viewed as comparably accurate in their identification of a white or a black suspect. By contrast, all four experiments show that when eyewitnesses are moderately confident in their identification then white suspects are viewed as more likely to be guilty than are black suspects and eyewitnesses are perceived as more accurate in their identification of a white than a black suspect.

One explanation for these results is that participants assumed (a) that the eyewitness was white, and (b) that eyewitnesses are less accurate at identifying cross-race than same-race suspects when they are moderately confident but there is no effect of same- vs. cross-race identifications when eyewitnesses are highly confident. This explanation is consistent with results on the effects of race on identification accuracy. We (Dodson & Dobolyi, 2016) have shown identification accuracy is worse for cross-race than same-race faces. And, the magnitude of this cross-race effect on accuracy appears to diminish when participants are highly confident than not confident in their identification.

Many studies have examined how laypeople interpret eyewitness testimony to make inferences about the culpability of a defendant and the likely accuracy of an eyewitness's identification (e.g. Bradfield & Wells, 2000; Cutler et al., 1988, 1990). Cutler and colleagues showed that individuals are sensitive to an eyewitness's level of confidence and perceive an eyewitness as more likely to be accurate and the suspect as more likely to be guilty when the eyewitness expressed 100% rather than 80% confidence in the identification. Notably, however, individuals were insensitive to a variety of factors that are known to affect the accuracy of an eyewitness's identification, such as whether or not the robber wore a disguise (i.e. Cutler et al., 1988, 1990).

But, we and Cash and Lane (in press) have shown that individuals are sensitive to *how* eyewitnesses justify their level of confidence. One way to illustrate the power of this featural justification effect is to point out – as is visible in Figure 5 – that highly confident eyewitnesses who provided a featural justification (e.g. 'I am 100% sure. I remember his hair') were perceived as significantly *less* accurate than moderately confident eyewitnesses who provided no justification (e.g. 'I am 80% sure'), $t(318) = 1.99$, $p = .047$, Cohen's $d = 0.22$, 95% CI [0.00, 0.44]. So, although eyewitnesses are perceived as more accurate when they are 100% than 80% confident and provide no additional justification (i.e. our Confidence Only condition) – replicating Cutler and colleagues – we show that a featural justification can cause others to perceive a highly confident eyewitness as less accurate than an eyewitness who is only moderately confident and says nothing else about why they are moderately confident.

More broadly, our research is consistent with others who are investigating variables that influence how laypeople interpret eyewitness confidence (e.g. Brewer & Burke, 2002; Spellman & Tenney, 2010). For example, Tenney, Spellman, and colleagues (Tenney, MacCoun, Spellman, & Hastie, 2007; Tenney, Spellman, & MacCoun, 2008; see Spellman & Tenney, 2010 for review) show that people evaluate an individual's (e.g. eyewitness's) credibility based on that individual's record of the calibration of their confidence with accuracy. Given two individuals who are both highly confident about a target detail, laypeople will prefer and view as more credible the (a) individual who was of mixed confidence about another detail that turned out to be wrong as compared to (b) an individual who was highly confident and wrong about the other detail. Overall, our research and that of Tenney and Spellman and colleagues show that individuals are not blindly swayed by an eyewitness's high confidence. The layperson's apparent default belief that high confidence signals high accuracy can be (rightly or wrongly) overcome depending on either (a) what the eyewitness says to justify this confidence or (b) the eyewitness's track record of their calibration of confidence to accuracy.

Before speculating about potential practical implications of these findings it is important to note that there are clear differences between the procedure in these experiments and what jurors experience. Most notably, jurors may not see the lineup of faces that was shown to the witness. Instead, they may only see the suspect in court. Although Experiment 4 and Dodson and Dobolyi (2015, Experiment 3) show that the featural justification effect on perceived accuracy and confidence persists when individuals see a single suspect and no lineup, this effect is larger when individuals see a lineup, for reasons described earlier.

In our view, the main reason why it is premature to make policy recommendations based on the current findings is that no one knows if it is an optimal or suboptimal strategy

for individuals to discount an eyewitness's accuracy when they provide a featural justification. No one has documented the confidence–accuracy relationship when an identification is accompanied by either a featural justification, an unobservable justification or another justification type. Lineup identifications that are based on a featural justification may be either less or more accurate (or no different) than those associated with other justification types. As yet, no one knows. However, Reinitz et al. (2012) have shown that face recognition is more accurate when it is based on remembering a specific feature than when it is based on the face's familiarity. It is conceivable then that lineup identifications that are associated with featural justifications may actually be more accurate than those associated with other justification types. This means that discounting an eyewitness's perceived accuracy – and thus, the suspect's guilt – when the eyewitness has provided a featural justification may be the *opposite* of what laypeople should do.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the National Science Foundation under Grant No. 1632174.

References

Bradfield, A. L., & Wells, G. L. (2000). The perceived validity of eyewitness identification testimony: A test of the five Biggers criteria. *Law and Human Behavior, 24*, 581–594. doi:10.1023/A:1005523129437

Behrman, B. W., & Richards, R. E. (2005). Suspect/foil identification in actual crimes and in the laboratory: A reality monitoring analysis. *Law and Human Behavior, 29*, 279–301.

Brewer, N., & Burke, A. (2002). Effects of testimonial inconsistencies and eyewitness confidence on mock-juror judgments. *Law and Human Behavior, 26*, 353–364. doi:10.1023/A:1015380522722

Budescu, D. V., Broomell, S., & Por, H. H. (2009). Improving communication of uncertainty in the reports of the Intergovernmental Panel on Climate Change. *Psychological Science, 20*, 299–308.

Cash, D. K., & Lane, S. M. (in press). Context influences interpretation of eyewitness confidence statements. *Law and Human Behavior*.

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

Cutler, B. L., Penrod, S. D., & Dexter, H. R. (1990). Juror sensitivity to eyewitness identification evidence. *Law and Human Behavior, 14*, 185–191.

Cutler, B. L., Penrod, S. D., & Stuve, T. E. (1988). Juror decision making in eyewitness identification cases. *Law and Human Behavior, 12*, 41–55.

Dobolyi, D. G., & Dodson, C. S. (2013). Eyewitness confidence in simultaneous and sequential lineups: A criterion shift account for sequential mistaken identification overconfidence. *Journal of Experimental Psychology: Applied, 19*, 345–357. doi:10.1037/a0034596

Dodson, C. S., & Dobolyi, D. G. (2015). Misinterpreting eyewitness expressions of confidence: The featural justification effect. *Law and Human Behavior, 39*, 266–280. doi:10.1037/lhb0000120

Dodson, C. S., & Dobolyi, D. G. (2016). Confidence and eyewitness identifications: The cross-race effect, decision time and accuracy. *Applied Cognitive Psychology, 30*, 113–125. doi:10.1002/acp.3178

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*, 175–191.

Martire, K. A., Kemp, R. I., Sayle, M., & Newell, B. R. (2014). On the interpretation of likelihood ratios in forensic science evidence: Presentation formats and the weak evidence effect. *Forensic Science International*, *240*, 61–68. doi:10.1016/j.forsciint.2014.04.005

Reinitz, M. T., Seguin, J. A., Peria, W., & Loftus, G. R. (2012). Confidence-accuracy relations for faces and scenes: Roles of features and familiarity. *Psychonomic Bulletin and Review*, *19*, 1085–1093. doi:10.3758/s13423-012-0308-9

Semmler, C., Brewer, N., & Douglass, A. B. (2011). Jurors believe eyewitnesses. In B. L. Cutler (Ed.), *Conviction of the innocent: Lessons from psychological research* (pp. 185–209). Washington, DC: American Psychological Association. doi:10.1037/13085-009

Spellman, B. A., & Tenney, E. R. (2010). Credible testimony in and out of court. *Psychonomic Bulletin & Review*, *17*, 168–173.

Tenney, E. R., MacCoun, R. J., Spellman, B. A., & Hastie, R. (2007). Calibration trumps confidence as a basis for witness credibility. *Psychological Science*, *18*, 46–50.

Tenney, E. R., Spellman, B. A., & MacCoun, R. J. (2008). The benefits of knowing what you know (and what you don't): How calibration affects credibility. *Journal of Experimental Social Psychology*, *44*, 1368–1375.

Wells, G. L., & Penrod, S. D. (2011). Eyewitness identification research: Strengths and weaknesses of alternative methods. In B. Rosenfeld & S. D. Penrod (Eds.), *Research methods in forensic psychology* (pp. 237–256). Hoboken, NJ: John Wiley and Sons.