

## Revisiting absolute and relative judgments in the WITNESS model

Dustin Fife · Colton Perry · Scott D. Gronlund

Published online: 14 August 2013  
© Psychonomic Society, Inc. 2013

**Abstract** The WITNESS model (Clark in *Applied Cognitive Psychology* 17:629–654, 2003) provides a theoretical framework with which to investigate the factors that contribute to eyewitness identification decisions. One key factor involves the contributions of absolute versus relative judgments. An absolute contribution is determined by the degree of match between an individual lineup member and memory for the perpetrator; a relative contribution involves the degree to which the best-matching lineup member is a better match to memory than the remaining lineup members. In WITNESS, the proportional contributions of relative versus absolute judgments are governed by the values of the decision weight parameters. We conducted an exploration of the WITNESS model's parameter space to determine the identifiability of these relative/absolute decision weight parameters, and compared the results to a restricted version of the model that does not vary the decision weight parameters. This exploration revealed that the decision weights in WITNESS are difficult to identify: Data often can be fit equally well by setting the decision weights to nearly any value and compensating with a criterion adjustment. Clark, Erickson, and Breneman (*Law and Human Behavior* 35:364–380, 2011) claimed to demonstrate a theoretical basis for the superiority of lineup decisions that are based on absolute contributions, but the relationship between the decision weights and the criterion weakens this claim. These findings necessitate reconsidering the role of the relative/absolute judgment distinction in eyewitness decision making.

**Keywords** Eyewitness identification · Relative and absolute judgments · Computational modeling · WITNESS model

One of the most unfortunate consequences of a legal system governed by human judgment is the alarming occurrence of false eyewitness identifications; too many individuals have been incarcerated for crimes that they did not commit. Many factors contribute to false convictions, but the principal one is faulty eyewitness identification, which contributed to the convictions in 75% of the DNA exonerations (see [www.innocenceproject.org/](http://www.innocenceproject.org/)).

A major contributor to faulty eyewitness identification is thought to be an overreliance on relative judgments (Wells, 1984). A relative judgment can involve making a selection of the lineup member who most resembles the perpetrator relative to the other lineup members, which can arise when comparisons are made across lineup members. Obviously, this is problematic if the police have an innocent suspect who is compared to foils that poorly resemble the perpetrator. On the other hand, an absolute judgment involves choosing the best-matching lineup member if, and only if, the degree of match to memory is above a criterion value. It is believed that a reliance on absolute judgments would enhance the accuracy of eyewitness identifications. This is the rationale for the recommendation that lineups be conducted sequentially (with lineup members being presented one at a time; e.g., Wells et al., 1998; for a review, see Gronlund, Andersen, & Perry, 2013).

To evaluate this distinction, Wells (1993) developed the removal-without-replacement paradigm. This paradigm requires the creation of two lineups. In one lineup (the target-present lineup, which includes the perpetrator/guilty suspect), the guilty suspect is presented along with five foils. In the other lineup (the target-removed lineup), the guilty suspect is removed and NOT replaced, creating a five-person lineup. Suppose that 50% of a randomly assigned sample of participants correctly selected the guilty suspect from a target-present lineup and 20% rejected the lineup. Wells argued that if participants rely on absolute judgments, approximately 70% of the participants in the target-removed lineup should reject the lineup. This would consist of the 50% of participants who would reject because they did not see the perpetrator that they

D. Fife (✉) · C. Perry · S. D. Gronlund  
Department of Psychology, University of Oklahoma, 455 W. Lindsey Street, Dale Hall Tower, Room 705, Norman, OK 73019-2007, USA  
e-mail: dfife@ou.edu

otherwise could identify, plus the 20% who rejected the lineup anyway. Conversely, to the extent that the rejection rate is less than 70%, it would signal that participants are following a relative decision rule. Wells (1993; see also Clark & Davey, 2005) reported that a large proportion of participants viewing the target-removed lineup chose a foil instead of rejecting the lineup, signaling a reliance on relative judgments. Participants' subjective reports also supported the relative–absolute conceptualization (e.g., Dunning & Stern, 1994; Kneller, Memon, & Stevenage, 2001).

Although absolute judgments are purported to yield better performance in lineup identifications, theoretical support for this proposal has been lacking. However, Clark, Erickson, and Breneman (2011) recently provided theoretical support by showing that a computational model (the WITNESS model, Clark, 2003) predicted that absolute judgments produced better performance than relative judgments in some circumstances.

We will begin by briefly introducing the WITNESS model. Following that, we will describe a restricted version of WITNESS (which, for pedagogical reasons, we refer to as WITNESS<sub>R</sub>) that we will compare to the original, unrestricted WITNESS model. We will show that the values taken by the relative/absolute decision weight parameters in WITNESS covary with the values taken by the decision criterion, making the decision weight parameters difficult to identify. That is, in most circumstances, WITNESS<sub>R</sub> is able to fit data as well with any value of the decision weights simply by adjusting the value of the criterion. This raises questions about the theoretical rationale for the superiority of absolute judgments, and the role of the relative/absolute judgment distinction in eyewitness decision-making.

### The WITNESS model

The WITNESS model (Clark, 2003) is a direct-access matching model (for an overview of this type of model, see Clark & Gronlund, 1996) that has been adapted for eyewitness situations. WITNESS uses numerical representations of features as items in the matching process. The details of the WITNESS model are beyond the scope of this article. Instead, we highlight the aspects of the model necessary for our analysis. WITNESS is implemented in R and available as a package to download from <http://cran.us.r-project.org/>.

To begin, WITNESS generates a perpetrator vector (PERP), which serves as the basis for all of the subsequent lineup members. The model next “encodes” the features in the PERP vector into a memory vector. The parameter  $a$  governs the degree to which the memory (MEM) vector matches the PERP vector;  $a$  is the probability that each individual feature in PERP will be successfully encoded to MEM. WITNESS next creates the lineup members for comparison to memory. For target-absent lineups (i.e., an innocent suspect replaces the

guilty suspect), a new vector is created to represent the innocent suspect (SUSP). This vector is governed by the parameter  $SSP$ , or the *similarity of the suspect to the perpetrator*.  $SSP$  is the probability that each feature of SUSP will match the corresponding feature of PERP. Next, WITNESS uses one of two more parameters to create the remaining lineup members (i.e., the foils). These two parameters simulate a different method of foil selection. Foils can be selected either because they match the description of the perpetrator (description-matched foils) or because they match the appearance of the suspect (suspect-matched foils). The parameter  $SFP$ , or *similarity of the foils to the perpetrator*, simulates a description-matched lineup.  $SFP$  is the probability that a given feature of a given foil will match a corresponding feature in PERP (similar to  $SSP$ , where 0 yields *no shared features* and 1 replicates *PERP*). Note that this results in the same foils being used in target-present and target-absent lineups.  $SFS$ , the *similarity of the foils to the suspect*, is a similar parameter; however, it uses the suspect in each given lineup (target present = PERP, target absent = SUSP) to generate the foils. Note that this results in different foils being used in target-present and target-absent lineups. Of note, Clark (2003) and Clark et al. (2011) used a different parameter notation, but the parameters that we detailed are mathematically equivalent to those used previously: In Clark's notation,  $SSP$  is  $S(I, P)$ ,  $SFP$  is  $S(F, P)$ , and  $SFS$  is  $S(F, S)$ . Following the construction of the lineup, the model must simulate the actual lineup procedure. WITNESS accomplishes this by comparing each lineup vector (i.e., PERP, SUSP, and FOILS) to MEM in order to create match values, or assessments of the degree to which two vectors overlap. These match values are the dot products of each lineup vector to MEM, divided by the total number of features. Larger dot products indicate a closer match to memory. After computing these match values, the values are used to execute the decision aspect of the lineup.

In order to model relative versus absolute contributions, WITNESS employs two parameters:  $w_a$ , the decision weight for the absolute contribution, and  $w_r$ , the decision weight for the relative contribution. These parameters are proportionally complementary, in that they are constrained to sum to 1 (i.e.,  $w_a + w_r = 1$ ). WITNESS uses these weights to determine the contributions of the two lineup members with the largest match values:  $w_a$  governs the contribution of the best match to MEM (BEST), whereas the contribution of the second best match (NEXT) is governed by  $w_r$ . When making its decision, WITNESS will choose BEST if the evidence (EV) [ $EV = w_a * BEST + w_r * (BEST - NEXT)$ ] exceeds  $c$ , the decision criterion. Thus, if  $w_a = 1$  (and so,  $w_r = 0$ ), the decision would be made entirely on the basis of BEST's match to memory (absolute contribution only); if  $w_r = 1$ , the decision would be made on the basis only of the magnitude of the difference between BEST and NEXT (relative contribution only). Both relative and absolute judgments can contribute to identification

decisions (meaning that  $w_a$  can take any value within the interval of 0 to 1). If the resulting EV value does not exceed  $c$ , the lineup is rejected, meaning that no individual is selected from the lineup.

Goodsell, Gronlund, and Carlson (2010) fitted WITNESS to a variety of data sets in an effort to examine the effects of  $w_a$  and  $w_r$  in simultaneous and sequential lineups (e.g., Kneller et al., 2001; Lindsay & Wells, 1985; MacLin, Zimmerman, & Malpass, 2005). In their exploration of WITNESS, they concluded that the decision weights produced little impact on the model's ability to fit data sets using description-matched lineups. Regardless of the values of the decision weights, they were able to achieve ostensibly identical fits to data simply by adjusting the value of  $c$  (see Fig. 1 in Goodsell et al., 2010). That is, they set  $w_a$  to 1, .5, or 0, held the other parameters constant, and achieved the same fit simply by adjusting  $c$ : The resulting receiver-operating characteristic (ROC) curves were coincident.

Clark et al. (2011) conducted a broader exploration of WITNESS's parameters to investigate the effects of the decision weights on identification accuracy in description-matched and suspect-matched designs. In this exploration, they compared three primary decision strategies: BEST ABOVE ( $w_a = 1$ ), BEST-NEXT ( $w_a = 0$ ), and BEST-REST (BEST minus the average of the remaining match values) across a wide range of parameter combinations. For each parameter combination, they generated ROC curves plotting correct identifications of the guilty suspect versus false identifications of the innocent suspect. Some of their parameter combinations (e.g., panel C in Fig. 2 of Clark et al., 2011) showed little difference among the three decision strategies, which is what Goodsell et al. (2010) observed. But others showed that in some cases the three decision strategies generated different ROC curves, indicating that the relative-versus-absolute distinction sometimes does impact identification accuracy. Moreover, under some parameter combinations, particularly in suspect-matched lineups, the ROC traced out by the absolute decision rule was superior. Although we agree with this finding, we disagree with the interpretation that it supports the superiority of absolute judgments. This is because the decision weights and the criterion covary even in the situations that appear to favor absolute judgments, making the decision weights difficult to identify and problematic to interpret.

To facilitate our analysis, we explored a restricted version of WITNESS, which we refer to as *WITNESS-Restricted* (WITNESS<sub>R</sub>). This version sets  $w_a$  equal to 1. In other words, WITNESS<sub>R</sub> has one less free parameter. Note that we would have reached nearly the same conclusions about how WITNESS implements the relative and absolute contributions if we had instead set  $w_a$  equal to 0, or any other value of  $w_a$ . Clark et al. (2011) examined the “Best Above Criteria” model, which is a pure absolute model and equivalent to setting  $w_a = 1$ ,

and the “Best-Next” model, which is a pure relative model and equivalent to setting  $w_a = 0$ . We did not replicate the “Best-Rest” model, substituting  $w_a = .5$  in its place. These are not identical models, but our goal was not to replicate Clark et al. but to examine the impact of the decision weights. Also, the Best-Rest model is not essential to the claim made by Clark et al. (2011, p. 364) that “the WITNESS model showed a consistent advantage for absolute judgments over relative judgments for suspect-matched lineups.” Throughout our analysis, we will focus on suspect-matched designs, in which the foils are created on the basis of either the perpetrator or an innocent suspect in (respectively) target-present and target-absent lineups, because Clark et al. concluded that the  $w_a$ - $w_r$  distinction made the largest difference for this design.

The evaluation of WITNESS versus WITNESS<sub>R</sub> proceeded as follows: First, we examined the identifiability of the  $w_a$  parameter. A model is identifiable if different values of a parameter generate different predicted values (see Bamber & van Santen, 2000). We then conducted a parameter recovery simulation. Particular parameter values were used to generate response proportions, and then the resulting response proportions were fit by freely estimating the remaining parameters in order to determine the extent to which the initial generating parameter values could be recovered (see Lewandowsky & Farrell, 2011).

We will show that the WITNESS model is only partially identifiable, because the decision weight parameters cannot be uniquely specified. We will also show that the WITNESS model has a high degree of parameter variability involving  $w_a$  and  $c$ , meaning that the parameter values recovered by a fitting algorithm vary greatly across iterations. Both of these analyses reveal that the values of the decision weight parameters are poorly identified and not well constrained by existing data. This makes any theoretical interpretation of the values or rank order of these parameters problematic.

### Bivariate identifiability

We investigated the behavior of the decision weights at the bivariate level using contour plots. Contour plots allow one to plot one of the variables as a color, thus showing the relationship between three variables without having to resort to three-dimensional plots. For example, suppose that we want to investigate how  $a$  and  $SFS$  interact to create a particular  $RMSE$  value.<sup>1</sup> Using a contour plot, we can plot  $a$  on the  $x$ -axis and  $SFS$  on the  $y$ -axis, with  $RMSE$  ranging from light to dark throughout the plot. Darker areas of the plot show the

<sup>1</sup>  $RMSE$  stands for root-mean squared error. It is defined as the square root of the average squared difference between the model-generated response proportions and the data.

combinations of  $a$  and  $SFS$  values that produce the best fit (i.e., the closest approximation between model and data). If two parameters are poorly identified at the bivariate level, then we should see large areas of the figure where nearly equal fits are obtained (i.e., large dark areas). Conversely, if two parameters are identifiable at the bivariate level, then we should see only a small dark area in the plot, signifying a tight area of best fit centered on the parameter values that generated the predicted response proportions.

We chose a set of parameter values that produced response proportions for WITNESS<sub>R</sub> and WITNESS (see Table 1) that were similar to one another and typical of actual data. We varied each parameter between 0 and 1 (except for  $c$ , which never exceeded .2) in increments of .02, resulting in 50 different values for each parameter. We then crossed each of these values with 50 different values of  $w_a$  (ranging from 0 to 1). For each of these parameter combinations, we iterated the WITNESS and WITNESS<sub>R</sub> models 10,000 times in order to estimate response proportions with stability, and then calculated  $RMSE$ . The result was four different  $50 \times 50$  matrices of  $RMSE$  values for every combination of parameter values. These results were plotted as described above in a contour plot. Figure 1 depicts WITNESS, with  $w_a$  on the  $x$ -axis and the other four parameters, one at a time, on the  $y$ -axis. Figure 2 depicts WITNESS<sub>R</sub> (with  $w_a$  set to 1.0); the different plots show interactions between all of the pairs of parameters. The crosshairs in the plots indicate the parameter combinations used to generate the response proportions in Table 1.

The first thing to notice is that identifiability varies as a function of which pair of parameters is being considered. For example, in WITNESS,  $w_a$  is identifiable when crossed with the  $SSP$  parameter (see panel C in Fig. 1); there is only one small area in the plot where  $RMSE$  is at a minimum ( $w_a \approx .5$ ,  $SFS \approx .6$ ). Panels A ( $a$  with  $w_a$ ) and B ( $SFS$  with  $w_a$ ) also show good identifiability. In contrast,  $w_a$  is poorly identified when crossed with  $c$  in panel D:  $RMSE$  can be at a minimum with any value of  $w_a$  between 0 and 1, as long as  $c$  follows a particular pattern—as  $c$  increases,  $w_a$  also can increase through its entire range, and still fit equally well. For example,  $RMSE$  is less than .05 when  $w_a$  is 0 and  $c$  is .025, but  $RMSE$  is also less than .05 when  $w_a$  is 1.0 and  $c$  is approximately .08.

**Table 1** Response proportions from WITNESS<sub>R</sub> and WITNESS used as targets in Figs. 1 and 2

	WITNESS <sub>R</sub>			WITNESS		
	Suspect	Foil	Reject	Suspect	Foil	Reject
Target Present	.78	.13	.10	.78	.11	.11
Target Absent	.33	.42	.25	.32	.39	.28

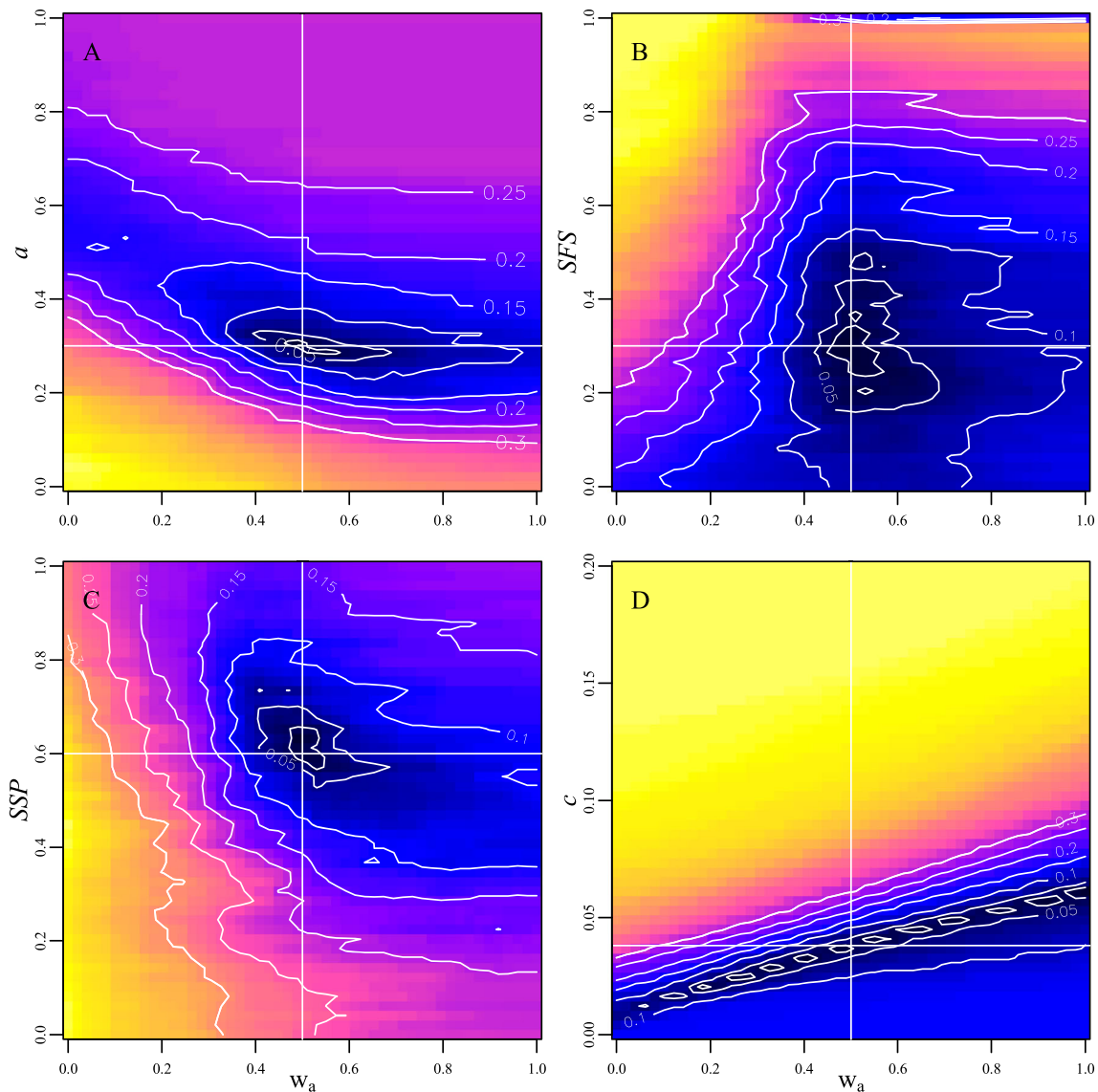
Parameter values: WITNESS,  $a = .3$ ,  $SSP = .6$ ,  $SFS = .3$ ,  $c = .038$ ,  $w_a = .5$ ; WITNESS<sub>R</sub>,  $a = .3$ ,  $SSP = .6$ ,  $SFS = .3$ ,  $c = .06$ ,  $w_a = 1.0$

This is the same problem that Goodsell et al. (2010) identified; WITNESS cannot ascertain the relative versus absolute contributions, even for this suspect-matched foil design. The results are different for WITNESS<sub>R</sub> (see Fig. 2). All parameter combinations showed a high degree of identifiability; each plot contains only one relatively small area where the fit is at a minimum. Not surprisingly, these areas were located around the parameter values that generated the response proportions in Table 1.

We next examined bivariate identifiability using a different set of parameter values:  $a = .3$ ,  $SSP = .4$ , and  $SFS = .8$ . This parameter configuration showed a large absolute judgment advantage in Clark et al. (2011, Fig. 6, panel B). We followed the same procedure described above to produce the contour plots for  $w_a$  versus  $c$  (Fig. 3). Notice that even with the parameter settings that showed the greatest advantage for absolute judgments at the univariate level (as assessed by ROC curves),  $w_a$  still covaries with  $c$  at the bivariate level. However, note that the covariation between  $w_a$  and  $c$  does not extend across the entire range of  $w_a$ . In the lower left-hand corner of Fig. 3, when  $w_a$  is close to 0, the model is unable to approximate the generated response proportions. What does this mean, and what are the implications for the absolute judgment advantage revealed by the ROC curves?

Although we agree with Clark et al. (2011, Fig. 6, panel B) that the ROC curve for the absolute judgment model is the highest, Fig. 3 reveals that it would be misleading to conclude that this is evidence of an advantage for absolute judgments. The covariation illustrated in Fig. 3 reveals that the value of  $w_a$  is interchangeable with  $c$  over most, although not the entire, range of  $w_a$ . Instead, what these results reveal is a relative judgment ( $w_a = 0$ ) disadvantage, but of a restricted sort. That is, over most of the range of  $w_a$  (in this case, once  $w_a > .25$ ), it will be very difficult to distinguish between a mixture model (e.g., one with a predominantly relative rule:  $w_r = .75$ ,  $w_a = .25$ ) and a pure absolute model ( $w_a = 1.0$ ).

To illustrate, we chose a point on the absolute ROC in Clark et al. (2011, Fig. 6, panel B) where the absolute advantage was the greatest. When the coordinates on the absolute ROC ( $w_a = 1$ ) were .37 and .07 (correct and false identifications, respectively), the relative ( $w_a = 0$ ) coordinates were much lower (.22 and .07). However, we can refit WITNESS to the absolute data for a range of different values of  $w_a$ . As is shown in Table 2,  $w_a$  can take any value between .25 and 1.0 and closely approximate correct and false ID rates of .37 and .07. Figure 4 shows that this holds true over the entire range of the ROC curve, which we swept out by varying  $c$  over its entire range. The ROCs that we traced out for  $w_a$  equals .25 and  $w_a$  equals 1.0 were nearly identical. Even if one argues that  $w_a$  equals 1.0 is still greater, this is hardly strong evidence for the superiority of absolute judgments, given that a model with  $w_a$  equals .25 (i.e., one that is predominantly relative) would be very difficult to distinguish from it. However, it does demonstrate that, despite



**Fig. 1** Bivariate identifiability of the original WITNESS model, conditional on  $w_a$ . Darker areas indicate where the fit is most optimal. Contour labels indicating the value of *RMSE* have been added for clarity. The

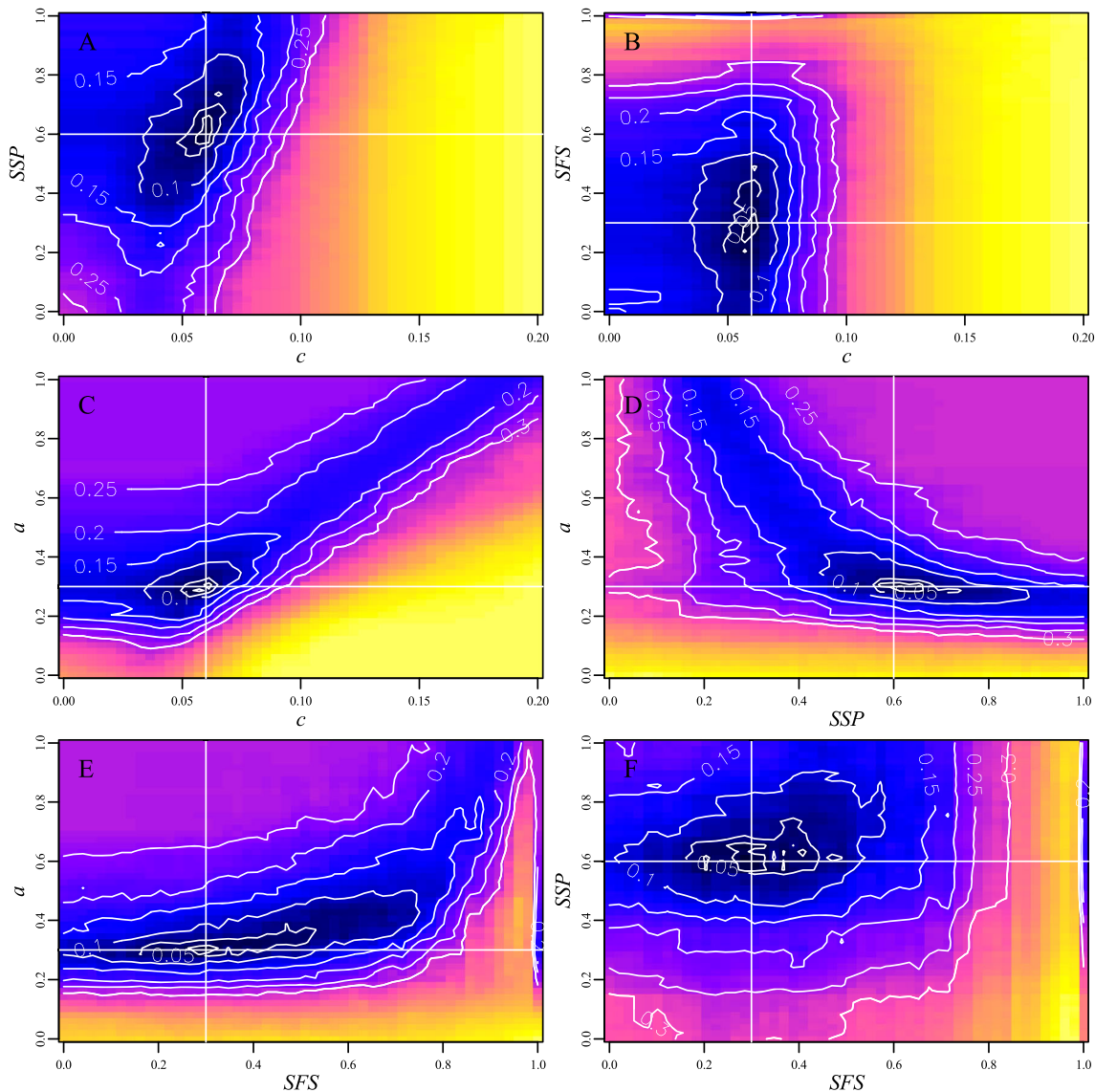
crosshairs indicate the parameter combination used to generate the data ( $a = .3$ ,  $SSP = .6$ ,  $SFS = .3$ ,  $c = .038$ ,  $w_a = .5$ )

the poor identifiability of the decision weights, it may be possible to distinguish pure relative ( $w_a = 0$ ) from pure absolute ( $w_a = 1.0$ ) rules in some circumstances.

### Parameter recovery

In the previous section, we showed that  $c$  and  $w_a$  are poorly identified, when holding all other parameters constant. But, in reality, parameters are rarely held constant. Instead, a fitting algorithm (e.g., genetic algorithm, steepest descent, or simulated annealing) is used to search the parameter space for the optimal parameter combinations. To determine whether WITNESS's partial identifiability is problematic in applied settings, we did the following:

1. Using the parameter values listed in Table 3 (which were very similar to those in Table 1), we used WITNESS and WITNESS<sub>R</sub> to generate response proportions.
2. We fit the response proportions from Step 1 using a steepest-descent algorithm, to see how well the models could recover the generating parameter values. If the fit was better than  $RMSE = .05$ , the parameter values were retained for later consideration. Otherwise, the parameter values were ignored, because the fitting algorithm failed to produce acceptable fits. Note that our intent was not to find the best parameter values. If this had been our intent, we would have used another algorithm less sensitive to local minima, such as a genetic algorithm or a simulated annealing procedure. Instead, our intent was to understand the degree of variability in parameter estimation



**Fig. 2** Bivariate identifiability of the WITNESS<sub>R</sub> model ( $w_a = 1.0$ ), conditional on each pairwise comparison of parameters ( $\alpha$ ,  $SFS$ ,  $SSP$ , and  $c$ ). Darker areas indicate where the fit is most optimal. Contour labels indicating

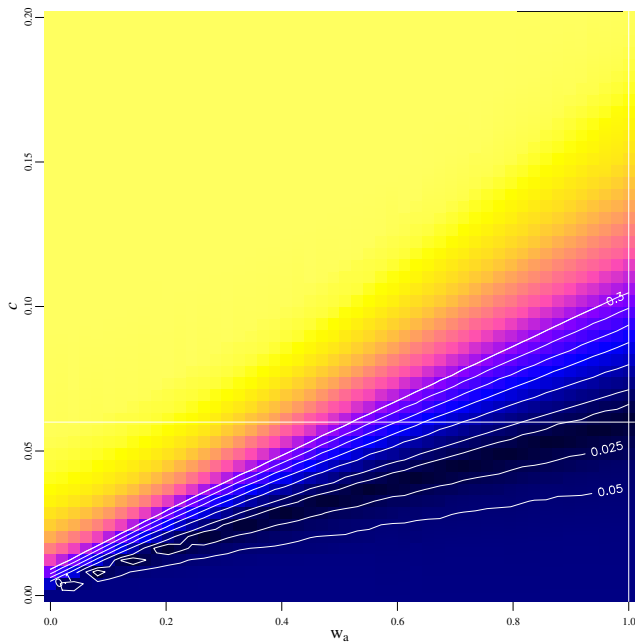
the value of  $RMSE$  have been added for clarity. The crosshairs indicate the parameter combination used to generate the data ( $\alpha = .3$ ,  $SSP = .6$ ,  $SFS = .3$ ,  $c = .06$ ,  $w_a = 1.0$ )

under conditions that produced what we have observed to be typical fits.

- Both WITNESS and WITNESS<sub>R</sub> were repeatedly fit to the response proportions until we achieved 1,000 fits with  $RMSE < .05$ . This took 3,488 attempts for WITNESS and 2,543 attempts for WITNESS<sub>R</sub>.

Figure 5 shows three histograms. The two on the top correspond to the original WITNESS model; the left histogram shows the distribution of the recovered  $c$  parameters, and the right histogram shows the distribution of the recovered  $w_a$  parameter. The bottom histogram corresponds to WITNESS<sub>R</sub> and shows the values of the  $c$  parameter.

Several things are worth noting about the histograms. First, the  $w_a$  histogram is fairly uniform. This shows that in applied situations, estimates of the  $w_a$  value will be variable, ranging from totally relative to totally absolute. Second, the  $c$  parameter also is highly variable, because it trades off with the value of  $w_a$  ( $r = .735$ ). Finally, note that the modes for the  $c$  and  $w_a$  parameters for the original WITNESS model are not located at the parameter values that generated the response proportions, in contrast to the mode for  $c$  for the WITNESS<sub>R</sub>. The  $c$  parameter is still variably estimated using WITNESS<sub>R</sub>, but less so than WITNESS. The reason for this variability can be attributable to the lax .05 fit criterion. When a more stringent fit criterion was used (i.e., only retain models with an  $RMSE$



**Fig. 3** Bivariate identifiability of the original WITNESS model with parameter settings  $a = .3$ ,  $SSP = .4$ ,  $SFS = .8$ ,  $c = .06$ , and  $w_a = .5$ . The parameter values showed a large absolute judgment advantage in Clark et al. (2011, Fig. 6, panel B). Here, the parameter  $w_a$  has been crossed with  $c$

less than .01), the variability reduced substantially for the distribution of  $c$  in WITNESS<sub>R</sub>, with over 90% of the  $c$  values falling between .04 and .08.

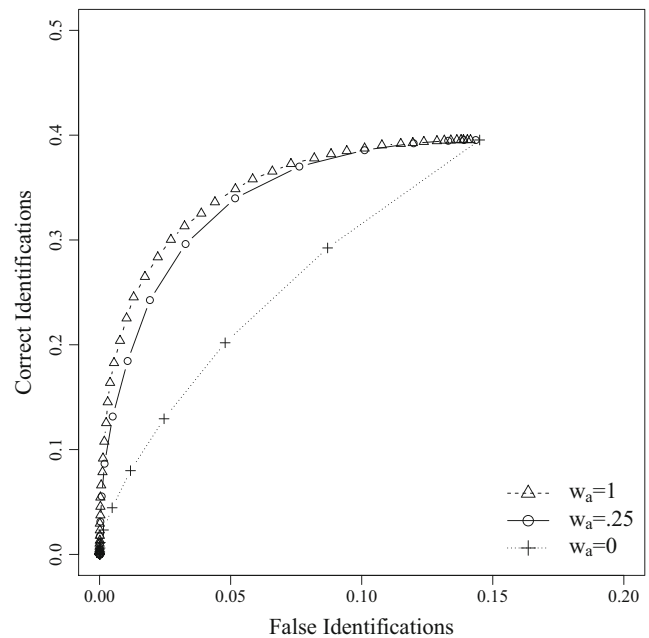
**Implications for relative/absolute judgment processes**

In WITNESS, the absolute match value is necessarily higher than any relative match value, because the relative match value is the absolute match value minus the next-best match value. Consequently, if WITNESS puts more weight on the higher evidence (the absolute match), it necessarily makes the total evidence have a higher value. But we have shown that an

**Table 2**  $w_a$  and  $c$  parameters in WITNESS tradeoff to produce the same correct and false identification rates over a wide range

$w_a$	$c$	Correct IDs	False IDs
1	0.07	.37	.07
.75	0.055	.37	.06
.5	0.0391	.37	.07
.25	0.0225	.37	.07
.1	0.0155	.37	.09
0	0.00157	.37	.13

Parameter values: WITNESS,  $a = .3$ ,  $SSP = .4$ ,  $SFS = .38$ ; number of simulations = 10,000



**Fig. 4** The receiver-operating characteristic (ROC) curves traced out for  $w_a$  equals .25 and  $w_a$  equals 1 are very similar, and both are superior to the ROC traced out by a pure relative rule ( $w_a = 0$ ). The remaining parameters ( $a = .3$ ,  $SSP = .4$ ,  $SFS = .8$ ) are taken from Clark et al. (2011, Fig. 6, panel B)

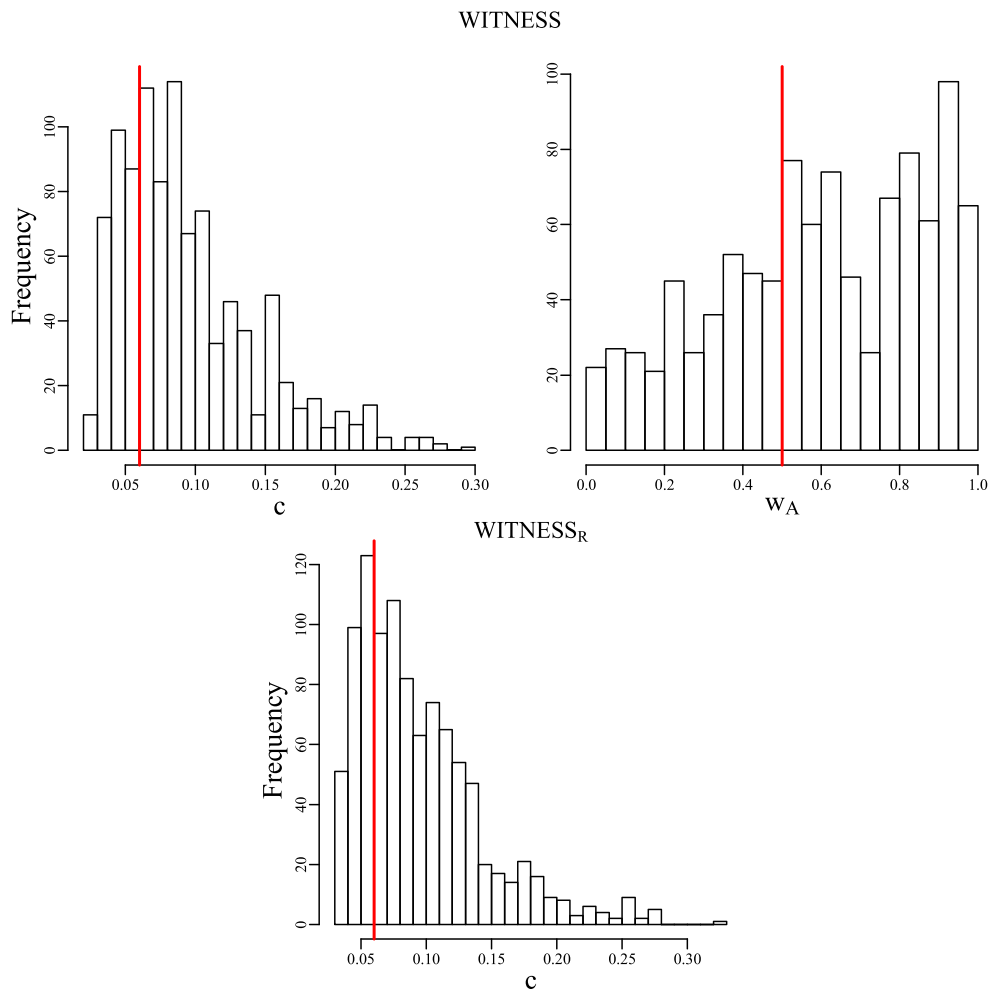
appropriate increase in the criterion generally can maintain the same predicted response proportions. As a consequence, the  $w_a/w_r$  parameters are difficult to identify within the current implementation of the WITNESS model. We believe that this undermines the theoretical support offered by Clark et al. (2011) for the superiority of absolute judgments in eyewitness identification decision-making. Moreover, the theoretical rationale for the relative/absolute conceptualization of eyewitness decision-making should be reassessed in light of these findings.

One alluring aspect of the relative-versus-absolute judgment concept is that it is introspectively intuitive. It is easy to conceptualize the experience of a relative decision process: We evaluate multiple options and seemingly make comparisons amongst them, and it seems logical for these comparisons to be incorporated into the decision process. As we previously mentioned, several studies have found relationships between

**Table 3** Response proportions from WITNESS<sub>R</sub> and WITNESS used in parameter recovery procedure

	WITNESS <sub>R</sub>			WITNESS		
	Suspect	Foil	Reject	Suspect	Foil	Reject
Target present	.78	.13	.10	.56	.04	.40
Target absent	.33	.42	.25	.15	.10	.75

Parameter values: WITNESS,  $a = .3$ ,  $SSP = .6$ ,  $SFS = .3$ ,  $c = .06$ ,  $w_a = .5$ ; WITNESS<sub>R</sub>,  $a = .3$ ,  $SSP = .6$ ,  $SFS = .3$ ,  $c = .06$ ,  $w_a = 1.0$



**Fig. 5** Distribution of parameter recovery values across 1,000 iterations of a steepest-descent fitting algorithm. The bottom plot is the distribution of the  $c$  parameter for WITNESS<sub>R</sub>, and the top plots correspond to the  $c$  and  $w_a$  parameters for the original WITNESS model

the experiential self-reports of absolute or relative decisions and witness accuracy (Dunning & Stern, 1994; Kneller et al., 2001; Lindsay et al., 1991). However, experiential reports, intuition, or introspection do not validate psychological processes, and formal modeling allows researchers to investigate phenomena in a manner constrained so as to control natural human biases and errors of intuition (see Hintzman, 1991). For example, Hintzman (1986) demonstrated that a single-store exemplar model of memory accounted for perceived abstractions from episodic memory, despite conventional intuition that a separate abstraction process and memory store were necessary to explain this phenomenon.

Clark et al. (2011) stated at several points in their article that the relative–absolute judgment contributions would be difficult to discriminate empirically, even when they found evidence for an absolute judgment advantage. We agree, and our analyses have demonstrated why this is the case. This might be a structural limitation of the WITNESS model, suggesting that alternative implementations for the relative/absolute contributions should be explored (Clark et al., 2011,

reviewed several possibilities). However, making a memory decision on the basis of an absolute comparison to memory versus a relative comparison to competitors seems like it should have a large impact on performance. But the problem lies not with WITNESS; at least, Clark (2003) specified an implementation for these decision processes. Before we start changing the way that WITNESS implements these different decision components, we need better operationalization of, and empirical evidence for, these decision contributions to make sure that they exist, and if so, how they impact performance.

The identifiability problem in WITNESS also could be a function of the paucity of eyewitness data. One of the difficulties of modeling eyewitness memory is that the data are limited, consisting of only six response proportions (and only four degrees of freedom). The conclusion that we have reached regarding how WITNESS implements relative and absolute judgments may not hold, once a richer set of data constrain the model (e.g., including confidence and latency data, in addition to response accuracy). Although the model will need to be extended to account for these additional data,



the consideration of richer data may reduce or eliminate the covariation between the decision weights and the criterion. Likewise, more definitive, objective, empirical evidence for the contributions of relative and absolute judgments to eyewitness decisions may require the consideration of data beyond response accuracy.

Sauer, Brewer, and Weber (2008) developed an alternative response format for lineups, in which witnesses provide confidence judgments for each lineup member rather than a binary decision for the entire lineup. This procedure then used an algorithm to transform these confidence judgments into a lineup decision that was typically more diagnostic than a binary identification decision, extracting more information in order to make the decision than a traditional lineup offers. A similar design may allow researchers to detect differences between absolute and relative decision-making, perhaps by generating rating profiles that are indicative of different processes, or by manipulating lineup members such that different decision strategies should result in different confidence ratings. This direction of design may also provide additional complexity that will allow researchers to overcome the relative–absolute identifiability problem that we have examined.

## References

- Bamber, D., & van Santen, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, *44*, 20–40.
- Clark, S. E. (2003). A memory and decision model for eyewitness identification. *Applied Cognitive Psychology*, *17*, 629–654.
- Clark, S. E., & Davey, S. L. (2005). The target-to-foils shift in simultaneous and sequential lineups. *Law and Human Behavior*, *29*, 151–172.
- Clark, S. E., Erickson, M. A., & Breneman, J. (2011). Probative value of absolute and relative judgments in eyewitness identification. *Law and Human Behavior*, *35*, 364–380.
- Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, *3*, 37–60. doi:10.3758/BF03210740
- Dunning, D., & Stern, L. B. (1994). Distinguishing accurate from inaccurate eyewitness identifications via inquiries about decision processes. *Journal of Personality and Social Psychology*, *67*, 818–835.
- Goodsell, C. A., Gronlund, S. D., & Carlson, C. A. (2010). Exploring the sequential lineup advantage using WITNESS. *Law and Human Behavior*, *34*, 445–459.
- Gronlund, S. D., Andersen, S. M., & Perry, C. (2013). Presentation methods. In B. Cutler (Ed.), *Reform of eyewitness identification procedures* (pp. 113–138). Washington, DC: American Psychological Association.
- Hintzman, D. L. (1986). “Schema abstraction” in a multiple-trace memory model. *Psychological Review*, *93*, 411–428. doi:10.1037/0033-295X.93.4.411
- Hintzman, D. L. (1991). Why are formal models useful in psychology? In W. E. Hockley & S. Lewandowsky (Eds.), *Relating theory and data: Essays on human memory in honor of Bennet B. Murdock* (pp. 39–56). Hillsdale: Erlbaum.
- Kneller, W., Memon, A., & Stevenage, S. (2001). Simultaneous and sequential lineups: Decision processes of accurate and inaccurate eyewitnesses. *Applied Cognitive Psychology*, *15*, 659–671.
- Lewandowsky, S., & Farrell, S. (2011). *Computational modeling in cognition: Principles and practice*. Los Angeles, CA: Sage.
- Lindsay, R. C. L., Lea, J. A., Nosworthy, G. J., Fulford, J. A., Hector, J., LeVan, V., & Seabrook, C. (1991). Biased lineups: Sequential presentation reduces the problem. *Journal of Applied Psychology*, *76*, 796–802.
- Lindsay, R. C. L., & Wells, G. L. (1985). Improving eyewitness identifications from lineups: Simultaneous versus sequential lineup presentation. *Journal of Applied Psychology*, *70*, 556–564.
- MacLin, O. H., Zimmerman, L. A., & Malpass, R. S. (2005). PC\_Eyewitness and the sequential superiority effect: Computer-based lineup administration. *Law and Human Behavior*, *29*, 303–321.
- Sauer, J. D., Brewer, N., & Weber, N. (2008). Multiple confidence estimates as indices of eyewitness memory. *Journal of Experimental Psychology: General*, *137*, 528–547. doi:10.1037/a0012712
- Wells, G. L. (1984). The psychology of lineup identifications. *Journal of Applied Social Psychology*, *14*, 89–103.
- Wells, G. L. (1993). What do we know about eyewitness identification? *American Psychologist*, *48*, 553–571.
- Wells, G. L., Small, M., Penrod, S. D., Malpass, R. S., Fulero, S. M., & Brimacombe, C. A. E. (1998). Eyewitness identification procedures: Recommendations for lineups and photospreads. *Law and Human Behavior*, *22*, 1–39.