Successful cuing of gender source memory does not improve location source memory

Jason L. Hicks¹ · Jeffrey J. Starns²

Published online: 25 January 2016
© Psychonomic Society, Inc. 2016

Abstract In three experiments we explored cross-dimensional cuing effects in a multidimensional source encoding and retrieval paradigm. We employed a bias-controlled experimental method of source cuing at retrieval (Starns & Hicks, 2013) in an attempt to improve retrieval of location information indirectly by cuing gender information. Encoded words were situated on the left or right side of a computer monitor and associated with either a male or a female face. When multiple faces were used across the set of encoded words, reinstating the correct face at retrieval alongside an incorrect, opposite-gender face cue improved male/female source decisions for test words. However, this powerful test cue did not improve memory for the encoded location of the words, suggesting that within-dimension cuing does not produce cross-dimensional cuing. This null outcome was found when gender decisions were required (Experiments 1A and 2) or not required (Experiment 1B) prior to location decisions. Nor was cross-dimension cuing found when subjects were told to expect a source test of both gender and location information at retrieval (Experiment 2). Our findings reinforce prior work demonstrating that multiple context dimensions can be bound to item information without any direct binding between the contexts.

Keywords Context effects · Source memory · Recognition · Cuing · Stochastic dependence

✉ Jason L. Hicks
jhicks@lsu.edu

¹ Department of Psychology, Louisiana State University, Baton Rouge, LA 70803-5501, USA

² University of Massachusetts Amherst, Amherst 01003, MA, USA

Source memory refers generally to one's ability to remember features of the presentation context during retrieval of an earlier learned episode. Such features might include the time or place associated with the learned event, what one was thinking or doing at the time, or associated perceptual features, such as auditory or visual information. Source monitoring refers to the processes in which such features are used as input to decisions about alternative possible sources (e.g., Is this information old or new? Are the retrieved perceptual features more consistent with a male or female friend that told me a rumor?; Johnson, Hashtroudi, & Lindsay, 1993). Source decision environments, much like associative memory decisions, provide data about theoretical mechanisms that underlie binding among events, items, or objects and their contextual features. One goal of source monitoring research is to understand the manner in which information is stored in episodic memory. Another goal is to understand how that information is accessed and used in controlled laboratory environments and in more naturalistic environments. Typical laboratory paradigms require people to learn information along a source dimension that is usually categorized into two or three sources. An example is remembering whether information was learned visually or aurally along a modality dimension (e.g., Hintzman, Block, & Inskip, 1972).

In addition to traditional one-dimensional source paradigms, multidimensional paradigms have been used increasingly to study the binding during encoding of items or objects to sources and the binding of sources to other sources (for a recent review, see Hicks & Starns, 2015). By multidimensional, we mean the learning of information from at least two separable dimensions that are manipulated orthogonally. For example, Meiser and Bröder (2002) asked people to study words on either the top or bottom of a computer monitor and either in a large or a small font size. Marsh and Hicks (2002) asked people to imagine receiving an object from one of two

sources but then also giving the object away to one of another pair of sources. Tests in these paradigms require that participants retrieve information along both source dimensions rather than on only one or the other. An important finding in these paradigms is that the probability of correctly retrieving one source attribute is often dependent on correctly retrieving the other. This phenomenon was dubbed *stochastic dependence* by Meiser and Bröder.

Theoretically, stochastic dependence has been interpreted as indicating the successful binding of item to source information but also of source to source information, sometimes referred to as context–context binding (Boywitt & Meiser, 2012), or as the successful creation of an episodic engram of highly integrated, interdependent elements (e.g., Horner & Burgess, 2013, 2014). Other studies have reinforced this interpretation. Meiser and Sattler (2007) demonstrated that the level of retrieval dependence is higher when encoding procedures immediately recapitulate the presentation of item information but is worse when encoding procedures take one's focal attention away from the source characteristics. The vast majority of research in this field shows that above-chance stochastic dependence is found only in the context of a feeling of recollection rather than in the context of familiarity (e.g., Meiser & Bröder, 2002; Meiser & Sattler, 2007). Meiser and colleagues have also developed multinomial processing tree models that provide parameter estimates capturing the degree to which stochastic dependence is present (Boywitt & Meiser, 2012; Meiser, 2014). The most recent model offered by Boywitt and Meiser isolates a parameter dedicated to joint retrieval of source information, which they argue is a measure of context–context binding. We note that empirical evidence of an above-chance joint retrieval parameter in multinomial modeling is evidence only of stochastic dependence, not the theoretical cause of such dependence. The interpretation of such dependence as reflecting encoding processes is certainly not controversial, although whether it represents direct context–context binding (cf. Boywitt & Meiser, 2012) or indirect context–item–context binding (cf. Starns & Hicks, 2005, 2008) is an open question.

However, one puzzling finding in the literature is that externally cuing one source attribute at retrieval has not proved capable of improving source memory for the second attribute. So despite the fact that stochastic dependence suggests binding of source attributes to one another, these attributes appear incapable of assisting retrieval of one another via cue reinstatement (Starns & Hicks, 2005; Vogt & Bröder, 2007). This suggests that stochastic dependence only appears to be related to direct context–context binding, but in reality it is related only to item–context binding. That is, some items are better learned than others, and these items tend to have accurate memory for both source dimensions, creating a dependency between the two. One major limitation in this prior external cuing work is that the cues used during retrieval have also

been limited in demonstrating cuing effects on the cued dimension itself, let alone in cuing the other source. For example, Starns and Hicks attempted to cue at test the font size of the item as large or small. The obvious prediction is that correctly cuing an item's font size should help one retrieve the font size information at a higher probability, but this did not happen. Vogt and Bröder (2007) reasoned appropriately that one should not expect much cross-dimension cuing when within-dimension cuing is absent. In their work, in at least one experimental condition they were able to establish within-dimension cuing but nonetheless could not find cross-dimension cuing (see also Starns & Hicks, 2008, for similar results in an associative recognition paradigm).

Study rationale

The present study is a further attempt to explore cross-dimension cuing with a new paradigm developed by Starns and Hicks (2013) that successfully produced within-dimension cuing while controlling for response bias due to cuing. As discussed more fully in Starns and Hicks (2013), paradigms in which external reinstatement is used to cue source memory retrieval are often unsuccessful for one or both of two reasons. First, the cue can sometimes bias source responding, which confounds the ability to separate bias from a memory benefit. Imagine a case in which a studied word was accompanied by a female voice. At test, that item might be presented again in the female voice or in a male voice for a gender source judgment. Performance increases with the female test cue, but this could be due to better memory or due to a higher likelihood of responding with the gender of the test cue (e.g., Dodson & Shimamura, 2000). Deciding to go along with a gender cue when it is correct will be countered by decisions to go along with the gender cue when it is incorrect. Second, theoretically Starns and Hicks reasoned that source memory decisions already require internal reinstatement of source details to create an appropriate retrieval cue and that providing people with an external cue is often redundant (Bjork & Richardson-Klavehn, 1989; Murnane & Phelps, 1994; Smith, 1979).

Starns and Hicks (2013) used a source cuing paradigm that emphasized the degree to which an external cue provides non-redundant information and also controlled for response bias. People studied words accompanied visually by male or female faces and made gender source memory decisions at retrieval. In one condition, half of the words were studied with a single female face and the other half with a single male face. This is similar to most traditional source memory paradigms with two sources along a single dimension. At retrieval, half of the test items were presented without a face cue, and others were presented with face cues. This represents a traditional external cuing manipulation that is often unsuccessful in source

paradigms (e.g., Starns & Hicks, 2005). Unique to the procedure was a second encoding condition that provided more variability among faces. Half of the words were studied with a female face, but there were eight different female faces used across the set of words. Similarly, eight different male faces were used across the other half of the words. This added variability among faces makes it difficult for people to internally reinstate information specific to each of 16 different possible faces at test. In this multiple-face condition, and as predicted, externally cuing people with the correct, specific face improved male–female gender decisions as compared with no-cue trials, but this was not true with the simpler single-gender face condition where internal reinstatement was sufficient to make the decision. Furthermore, the cuing technique was unique in that two faces were always presented at retrieval on cued trials—a female face and a male face. One of these faces was always correct and the other face was a familiar one used to study some of the other items. Presenting both a male and a female face cue experimentally controls for any potential response bias to go along with any one gender represented in the test cue. Increased performance can therefore be attributed solely to a memory cuing benefit rather than to a decision bias. This also obviates the need for complicated statistical modeling, such as that represented by the multidimensional multinomial applied in the Vogt and Bröder (2007) study. The experimental procedure, rather than statistical procedures, controls for the decision bias.

We capitalized on this new, powerful face-cuing procedure to examine whether reinstating a face cue at retrieval not only helps improve gender discrimination but also provides a cross-dimension benefit to a location source judgment. We modified the Starns and Hicks (2013) single-face and multiple-face encoding conditions by presenting half of the word-face pairings on the left side of the computer monitor and half on the right. This was crossed factorially with the gender of the face presented with the words. The key question of interest is whether providing male–female face cues at test improves a location decision for each word. If it is true that stochastic dependence relies on context–context binding and prior studies have simply used ineffective cuing procedures, then a powerful within-dimension cuing procedure may finally prove capable of producing cross-dimension cuing. In contrast, it may be the case that stochastic dependence only appears to reflect context–context binding, when in fact it is produced by variability in how well each context dimension is independently bound to the item information.

Experiments 1A, 1B, and 2

All of these experiments followed the same general methodology, which was very similar to that used in the first experiment of Starns and Hicks (2013). A set of words was learned

at encoding, and each word was accompanied by a male or female face. In the single-face condition, one male and one female face were used across the set of words, which meant that each face was associated with half of the words. In the multiple-face condition, eight different male and eight different female faces were associated with the words, making the face-to-word associative fan much lower. At retrieval, half of the test words were not cued by any faces, whereas the other half were accompanied by a combination male–female face cue. One of the faces was the correct one paired with that word at encoding, and the other face was from the opposite gender studied with other words. Both Experiment 1A and Experiment 2 used this methodology, with one major difference being that in Experiment 2 subjects were instructed to pay attention fully to both the face information and location information for each encoded item. A second change was that all items were presented twice in Experiment 2. These changes were made to increase source decision performance overall and to increase the likelihood that location and gender information could be integrated at encoding. The consequences should be better stochastic dependence among the sources and, theoretically, an increased chance at finding cross-dimension cuing among better integrated location–item–gender associations. Experiment 1B differed from Experiment 1A in that only location judgments were made, and only the single face actually studied with the word was presented on each cued trial. We conducted this experiment to see if providing the face cues, per se, rather than also having people make decisions about gender, had a cross-dimension influence on location discrimination. It is also possible that making gender decisions could have an interfering influence on subsequent location decisions (e.g., the presentation of two faces for gender decisions on each side of the screen or the use of similar response keys disrupts location judgments). Removing the gender decision eliminates this potential problem.

Method

Participants

Data were collected from Louisiana State University undergraduates who received course extra credit for volunteering. Participants were randomly assigned to the single-face ($n = 34$) and multiple-face ($n = 32$) conditions in Experiment 1A, with 29 and 28 participants in these respective conditions of Experiment 1B. For Experiment 2, there were 26 assigned to the single-face condition and 25 assigned to the multiple-face condition. Power analyses were used to estimate these sample sizes, given the within-subjects cuing effect size (Cohen's d ; Cohen, 1988) from the similar multiple-face conditions of Experiments 1 and 2 from Starns and Hicks (2013). In Experiment 1 of Starns and Hicks, Cohen's d was 0.53 for

the cuing effect; for Experiment 2 Cohen's d was 0.88. The present sample sizes were estimated using the software G*Power (Faul, Erdfelder, Lang, & Buchner, 2007). Using the parameters two-tailed $\alpha = .05$ and $(1 - \beta) = .80$, 30 participants were needed to achieve a power of .80 with an effect size of .53 and 10 participants were needed for similar power with an effect size of .88. Based on these estimates, we targeted sample sizes between 20 and 30 for each between-subjects condition of our experiments.

Materials

Ninety-six words were selected from the MRC psycholinguistic database (http://websites.psychology.uwa.edu.au/school/MRCDatabase/uwa_mrc.htm). All words were between five and eight letters in length, had familiarity ratings between 500 and 700, and had concreteness ratings between 300 and 500. Eight male and eight female faces were selected from Maner et al. (2003). The encoding list contained 96 words, with 48 words seen with a female face and 48 seen with a male face. The assignment of gender to individual words was counterbalanced across participants. For the single-face group, one of the male and one of the female faces was selected at random for each participant, and these faces appeared throughout the study phase. In the multiple-face group, each of the eight male faces was randomly paired with six words, and each of the eight female faces was randomly paired with six words.

Design and procedure

Each experiment formed a 2 (face condition) $\times$ 2 (cuing condition) design. Face condition was manipulated between subjects with "multiple-face" and "single-face" as the levels. Cuing condition was manipulated within subjects with the levels of this variable being "cued" and "uncued." Dependent measures included proportion correct source decisions and signal detection discrimination indexes for gender source memory and for location source memory.

Experiments 1A and 1B Prior to the study phase, participants were told that they would see words paired with faces, that the word-face pairings would appear randomly on different sides of the computer screen, and that their memory would be tested later in the experiment. The nature of the test was not specified. They were asked to imagine the person that they saw on each trial "demonstrating" the word they saw. For example, participants were told that if the word was a verb, they could imagine the person in the picture acting out the word. Conversely, if the word was a noun, they might imagine the person in the picture holding or pointing to the word's referent. After the instructions, the word-face pairs on the study list appeared in random order, and each remained on the screen

for 5 s with 200 ms of blank screen between presentations. Half of the face-word pairs were presented on the left side of the computer screen and half on the right (i.e., 24 of each gender on each side). Each face was shown equally often on the left and right side of the screen across the 96 study trials. Next, participants performed a distractor task in which they rearranged the segments of a scrambled picture on the computer screen until each segment was in its appropriate position. None of the pictures in the distractor task depicted human faces. Participants completed as many puzzles as they could solve in 3 minutes, and then they read the test instructions.

The test list contained all 96 studied words, with half of the words studied in each side-by-gender condition appearing with no face cues at test and half appearing with face cues. On cued trials, the test word appeared near the top-center of the computer screen. The female face appeared on the left side of the screen and the male face on the right. One face was always the individual face paired with the test item from the study phase and the mismatching face was of the opposite gender. The mismatching test face was the only alternative face in the single-face condition or was an individual face selected at random from the 8 of that opposite gender in the multiple-face condition. Faces were selected under the constraint that all faces appeared an equal number of times on the test. On uncued trials, gray rectangles occupied the positions where faces otherwise appeared on cued trials.

At this point the procedures differed between Experiment 1A and 1B. For Experiment 1A, test instructions informed participants that they would see words that appeared in the study phase and that they must decide whether each word appeared with a male or a female face. Participants were told that they would see a male and a female face on some of the test trials, and that one of the faces would always be the face studied with the test word. They were also informed that on other trials no faces would be shown. Participants entered a gender decision using either the "d" (female) or "j" (male) keys on the keyboard. On cued trials, immediately after making the gender decision the correct face cue remained on the screen but was shown centered underneath the test word. The mismatching face was not shown on this screen. Participants made a left/right side decision for the cued probe using the "d" key for "left" and the "j" key for "right." Thus, for every cued test item a gender decision was made first and the side decision was made next. For uncued trials, this same procedure was used except that instead of a face cue being shown in the screen center, a light-gray rectangle appeared.

Experiment 1B differed from this procedure in that no gender decisions were made; only location decisions were made for each word. On cued trials, only the correct face cue appeared centered underneath the centered test word for the left/right decision. On uncued trials, a gray rectangle appeared centered underneath the test word. As mentioned earlier, these methodological changes eliminate the potential for

interference from the gender-decision phase in Experiment 1A, interference that might be produced by the presentation of two face cues prior to gender decisions or produced by using the same response keys aligned on the left and right side of the keyboard for gender decisions and also for side decisions. All aspects of the encoding phase were identical to Experiment 1A.

Experiment 2 There were two important modifications to the encoding procedures listed above. First, during the initial instructions, people were told that their memory would be tested for the faces paired with the words and for the side on which the words were presented. Thus, encoding was intentional as to the nature of the eventual test. Second, the encoding phase was shortened slightly to 4 s per face–word pair, but the entire study sequence was repeated in a different random order after all 96 face–word pairs were presented once. This was done in order to generally increase memory for the stimuli and to enhance potential binding among the sources and words. The test phase was identical to that described for Experiment 1A. Gender and location decisions were made in sequence for each test word, with some trials being cued and other trials uncued.

Results and discussion

For ANOVA models, partial-eta squared (η_p^2) is reported as an estimate of effect size. For pairwise comparisons, the effect size Cohen's d is offered. An equal-variance signal detection measure of discriminability (d') was used to define source memory performance (e.g., Slotnick & Dodson, 2005; Starns & Hicks, 2013).¹ This measure is defined by the formula $d' = [z(\text{HR}) - z(\text{FAR})]$. Hit rates (HR) for gender memory were defined as the proportion of “male” responses for test words studied with male faces, and false alarm rates (FAR) were the proportion of “male” responses for words studied with female faces. A similar measure was calculated for the location memory decisions. “Right” decisions for words shown on the right at study served as the hits, and “right” decisions for words shown on the left were the false alarms. We used the Snodgrass and Corwin (1988) correction to avoid undefined d' values resulting from hit rates equal to one or false alarm rates equal to zero. That is, a constant of .5 was added to the numerator frequency of “male” and “right” responses before calculating the hit and false-alarm rates, with a constant of 1 added to the denominator frequency.

¹ We also assessed a two-high threshold model of discriminability by subtracting the false alarm rates directly from the hit rates. Analyses of this measure showed results similar to the analyses of d' . These similarities are consistent with the same general outcome in the Starns and Hicks (2013) report.

Experiments 1A and 1B The top portion of Table 1 displays hit rates, false alarm rates, and discriminability for gender source memory in Experiment 1A. These data are very similar to the multiple-faces gender cuing effect reported by Starns and Hicks (2013). A 2 (face condition) $\times$ 2 (cuing) mixed factorial ANOVA on gender discriminability produced a significant main effect of cuing, $F(1, 64) = 15.92, p < .01, MSE = .19, \eta_p^2 = .20$, which was qualified by a cuing $\times$ face condition interaction, $F(1, 64) = 11.12, p < .01, \eta_p^2 = .15$. Gender decisions were cued successfully in the multiple-face condition, $t(31) = 4.11, p < .01$, Cohen's $d = 0.73$ ($r = .54$ between cued and uncued trials). This was not true for the single-face condition, $t(33) = 0.67$, Cohen's $d = .11$ ($r = .27$ between cued and uncued trials). Face condition had no significant influence overall, $F(1, 64) < 0.01, MSE = 1.08, p > .95$.

In contrast, providing gender cues did not aid location source decisions. Descriptive statistics for the location judgments are shown in the top portion of Table 2. A 2 (face condition) $\times$ 2 (cuing) mixed factorial ANOVA on location discriminability produced no significant effects, all F s(1, 64) < 1.0 , all η_p^2 estimates $< .01$. Cohen's d estimates for the cuing effect were .00 for the single-face condition and .10 for the multiple-face condition. A similar null outcome occurred for Experiment 1B in which gender decisions were not made. These data are shown in the middle portion of Table 2. A 2 (face condition) $\times$ 2 (cuing) mixed factorial ANOVA on location discriminability produced no significant effects, including the main effect of cuing, $F(1, 55) = 0.97, p > .30, MSE = .21, \eta_p^2 = .017$, and the cuing $\times$ face condition interaction, $F(1, 55) = 0.11, p > .70, \eta_p^2 = .002$. Although performance nominally favored the single-face condition overall, this main effect was not significant, $F(1, 55) = 2.85, p = .09, MSE = .58, \eta_p^2 = .05$.

A complementary analysis was undertaken to analyze location discriminability conditionalized on correct gender decisions for Experiment 1A. The prior analysis of cross-dimension cuing was pooled over successful and unsuccessful gender decisions, the latter of which may have undercut our likelihood of finding a cross-dimension cuing effect. We conducted a 2 (face condition) $\times$ 2 (cuing) mixed factorial ANOVA on these conditionalized measures and again found no effect of cuing, $F(1, 64) = 0.61, p > .40, MSE = .20, \eta_p^2 = .009$, and no cuing $\times$ face condition interaction, $F(1, 64) = 0.001, p > .90, \eta_p^2 < .001$. The main effect of face condition was also nonsignificant, $F(1, 64) = 0.54, p > .45, MSE = .60, \eta_p^2 = .008$.

Experiment 2 The data for gender discriminations can be found in the bottom portion of Table 1 and for location decisions in the bottom portion of Table 2. The entire pattern of results mirrors that of Experiment 1A, albeit with slightly improved performance overall, presumably owing to the repeated encoding sequence and/or to the intentional encoding instructions. For gender discrimination in terms of d'

Table 1 Gender Source Memory Performance in Experiments 1A and 2

Experiment & Cuing Condition	Face Condition and Source Measure					
	Multiple Face			Single Face		
	“M” M	“M” F	d^a	“M” M	“M” F	d^a
Experiment 1A						
Uncued	.65 (.02)	.39 (.03)	0.74 (.14)	.70 (.02)	.36 (.02)	0.98 (.14)
Cued	.73 (.03)	.29 (.02)	1.28 (.14)	.70 (.02)	.35 (.03)	1.03 (.14)
Experiment 2						
Uncued	.72 (.03)	.27 (.03)	1.38 (.19)	.73 (.03)	.34 (.03)	1.13 (.18)
Cued	.82 (.03)	.23 (.03)	1.94 (.22)	.69 (.03)	.33 (.03)	1.07 (.19)

Note. “M”|M is the proportion of male items called “male”; “M”|F is the proportion of female items called “male.” Standard errors are in parentheses

estimates, there was a significant main effect of cuing, $F(1, 49) = 6.27, p < .05, MSE = .26, \eta_p^2 = .11$, which was qualified by a cuing $\times$ face condition interaction, $F(1, 49) = 9.68, p < .01, \eta_p^2 = .16$. Gender decisions were cued successfully in the multiple-face condition, $t(24) = 3.64, p < .01$, Cohen’s $d = 0.73$ ($r = .73$ between cued and uncued trials). This was not true for the single-face condition and in fact the effect size measure nominally favored the uncued trials, $t(25) = -0.47$, Cohen’s $d = .09$ ($r = .76$ between cued and uncued trials). Face condition had a significant effect favoring the multiple-face group, $F(1, 49) = 4.63, MSE = 1.72, p < .05, \eta_p^2 = .09$, but of course was qualified by the interaction just described.²

For location discrimination, again no significant cross-dimension cuing emerged. These data are shown in the bottom portion of Table 2. Neither the significant effect of cuing, $F(1, 49) = 1.11, p > .20, MSE = .16, \eta_p^2 = .02$, nor the cuing $\times$ face condition interaction, $F(1, 49) = 0.26, p > .60, \eta_p^2 = .005$, was significant. There was no main effect of face condition, $F(1, 49) = 0.34, p > .50, MSE = .73, \eta_p^2 = .007$. As with Experiment 1A, a complementary analysis was performed on location discriminability conditionalized on correct gender decisions. The 2 (face condition) $\times$ 2 (cuing) mixed factorial ANOVA

produced no effect of cuing, $F(1, 49) < .001, p > .99, MSE = .32, \eta_p^2 < .001$, and no cuing $\times$ face condition interaction, $F(1, 49) = 0.91, p > .75, \eta_p^2 = .002$. The main effect of face condition was not significant, $F(1, 49) = 0.15, p > .70, MSE = .90, \eta_p^2 = .003$.

Cross-dimension cuing across experiments Because the collection of evidence thus far points to a null effect of gender cuing on location judgments, it is appropriate to pool the data in order to increase the power to detect any such effect that might exist. Of course, nonsignificant results only provide meaningful evidence for the null hypothesis when experimental power is high, and the effect size for within-dimension cuing that we used to estimate our sample sizes may underestimate the statistical power required for a cross-dimension effect size. Therefore, we combined the data across experiments to increase the power to detect what could be a relatively small effect. We concentrated on the multiple-face cuing conditions for these analyses ($N = 85$ across all three experiments) because they were the only ones shown empirically to produce a within-cuing effect. For the pooled NHST analysis, we conducted a 2 (cuing: gender cued vs. uncued) $\times$ 3 (Experiment 1A, 1B, and 2) mixed ANOVA on location discriminability. The main effect of Experiment was significant, $F(2, 82) = 5.26, MSE = .384, p < .01, \eta_p^2 = .114$. Post hoc Sidak tests demonstrated a difference between Experiment 1A ($M = .40$) and 2 ($M = .745, p = .012$) and between Experiment 1B ($M = .414$) and 2 ($p = .022$). This result confirms improved location decisions overall after the changes in encoding procedures in Experiment 2. Of particular importance, the main effect of cuing was nonsignificant, $F(1, 82) = .005, MSE = .157, p > .90, \eta_p^2 < .001$, and this interpretation obviously holds whether one considered this a one-tailed or two-tailed test. The interaction of cuing with Experiment was not significant, $F(2, 82) = .915, p > .40, \eta_p^2 = .022$.

We also explored a Bayesian approach to provide more information about the probable effect sizes created by our cross-dimension and within-dimension cuing manipulations.

² Only in Experiment 2 does a main effect favor the multiple-face condition over the single-face condition in gender decision performance. In the Starns and Hicks (2013) experiments, the single-face condition was superior if there was any advantage. In Experiment 2, all of the encoded trials were repeated and the nature of the test was provided, which may have given the multiple-face encoding condition a disproportionate added quality to the encoding as opposed to the single-face condition (i.e., the multiple-face condition has more room to benefit from this help). More generally, we acknowledge that presenting multiple faces can have other effects besides just making internal reinstatement more difficult, and some of these effects could benefit performance. Regardless of these potential other effects, the effect on internal reinstatement should make external cuing manipulations more effective in the multiple-face than the single-face condition. These speculations may hint at factors to pursue in further work on the basic within-dimension cuing effect. None of this gainsays the essential finding that providing face cues improves gender decisions only in the multiple face conditions, which we found in two experiments here and is consistent with the Starns and Hicks findings.

Table 2 Location Source Memory Performance in Experiments 1A, 1B, and 2

Experiment & Cuing Condition	Face Condition and Source Measure					
	Multiple Face			Single Face		
	“R” R	“R” L	d'	“R” R	“R” L	d'
Experiment 1A						
Uncued	.59 (.02)	.45 (.02)	0.37 (.07)	.62 (.02)	.45 (.03)	0.48 (.11)
Cued	.62 (.02)	.46 (.02)	0.43 (.07)	.61 (.02)	.44 (.02)	0.47 (.10)
Experiment 1B						
Uncued	.58 (.02)	.43 (.03)	0.39 (.11)	.59 (.03)	.38 (.02)	0.60 (.12)
Cued	.60 (.02)	.43 (.02)	0.44 (.07)	.64 (.03)	.39 (.03)	0.71 (.15)
Experiment 2						
Uncued	.65 (.03)	.37 (.03)	0.81 (.14)	.62 (.03)	.38 (.03)	0.67 (.13)
Cued	.64 (.03)	.39 (.02)	0.68 (.12)	.63 (.03)	.40 (.03)	0.62 (.14)

Note. “R”|R is the proportion of “right” decisions given items studied on the right; “R”|L is the proportion of “right” decisions given items studied on the left. Standard errors are in parentheses. Gender decisions preceded location decisions in Experiments 1A and 2, but location judgments alone were made in Experiment 1B

Figure 1 shows posterior distributions on standardized effect size in our multiple-face conditions. The solid line shows cross-dimensional cuing (location judgment discriminability with vs. without face cues), and the dashed line shows the single-dimension cuing (gender judgment discriminability with vs. without face cues). The distributions were constructed using a unit normal as the prior distribution (e.g., Rouder, Speckman, Sun, Morey, & Iverson, 2009) and updating this prior with the combined data across all of our experiments. Our data support the claim that single-dimension cuing produces a relatively large cuing effect. The mode of the single-dimension distribution is near a standardized effect size of 0.7, with very little density in the region of medium and small

Fig. 1 Posterior distributions of effect size for the within-dimension face cuing of gender discriminability (dashed line) and cross-dimension face cuing of location discriminability (solid line). Data were pooled over Experiments 1A and 2 for within-dimension cuing and over Experiments 1A, 1B, and 2 for cross-dimension cuing

effect sizes. In contrast, the distribution for cross-dimension cuing is tightly peaked around an effect size of zero, indicating a high probability that this manipulation either produces no effect or a very small effect. Cohen’s recommendation for a small standardized effect size is 0.2 (Cohen, 1988) and the probability that our cross-dimension cuing manipulation produces a positive cuing effect of 0.2 or larger is quite low (.035). When we include data from both the single- and multiple-face condition (not shown), the posterior distribution for cross-dimension cuing is even more tightly concentrated around zero, and the probability of an effect size of 0.2 or larger is only .009. Therefore, our results provide strong evidence against anything other than a null effect or a very small effect of cross-dimension cuing.

Evidence for stochastic dependence All of the prior results concerning retrieval-based cuing, or lack thereof, should be interpreted within in the context of stochastic dependence or binding more generally. Consistent with much of the prior work on multidimensional source memory (Hicks & Starns, 2015), location decision performance was correlated with

Table 3 Proportion of Correct Location Decisions in Face-Cued and Uncued Conditions in Experiments 1A and 2, Contingent on Correct vs. Incorrect Gender Source Decisions

	Cued Trials		Uncued Trials	
	Gender Correct	Gender Incorrect	Gender Correct	Gender Incorrect
Experiment 1A	.61 (.01)	.53 (.02)	.60 (.01)	.53 (.02)
Experiment 2	.65 (.02)	.46 (.03)	.65 (.02)	.56 (.02)

Note. Standard errors are in parentheses

gender decision performance. Table 3 presents these data. Note that the patterns in Table 3 did not change materially when performance was calculated for gender decisions contingent on correct versus incorrect location decisions, even though gender decisions preceded location decisions.

A 2 (gender correct vs. incorrect) $\times 2$ (cuing) $\times 2$ (face condition) $\times 2$ (Experiment) mixed ANOVA produced a significant three-way interaction among gender decision performance, cuing, and Experiment, $F(1, 115) = 4.78$, $MSE = .015$, $p < .05$, $\eta_p^2 = .04$. We therefore conducted separate 2 (gender correct) $\times 2$ (cuing) $\times 2$ (face condition) mixed model ANOVAs for each experiment. In Experiment 1A, the only significant difference was that location decisions were more often correct following correct gender decisions ($M = .60$) as opposed to following incorrect gender decisions ($M = .53$), $F(1, 65) = 19.71$, $MSE = .017$, $p < .001$, $\eta_p^2 = .24$. For Experiment 2, the same main effect of gender correct vs. incorrect was produced with an even larger effect size, $F(1, 50) = 35.04$, $MSE = .028$, $p < .001$, $\eta_p^2 = .42$, but qualified by a two-way interaction between cuing and gender decision accuracy, $F(1, 50) = 9.13$, $MSE = .017$, $p < .01$, $\eta_p^2 = .16$. This interaction was produced because the difference in location performance between gender correct versus incorrect for the face-cued conditions, $t(50) = 5.78$, $p < .001$, Cohen's $d = 0.81$, was almost twice as large as the analogous difference for the uncued conditions, $t(50) = 3.14$, $p < .01$, Cohen's $d = 0.44$. In fact, these patterns seem due to variation primarily in the location decisions following incorrect gender judgments in the cued condition ($M = .46$) and for location decisions following incorrect gender judgments in the uncued condition ($M = .56$). The former mean was not significantly different from chance (.50), $t(50) = 1.61$, $p > .10$, but the latter mean was, $t(50) = 2.83$, $p < .05$. Location judgments contingent on correct gender decisions were similar between cued and uncued conditions (both M s = .65).

As we have argued elsewhere (Hicks & Starns, 2015; Starns & Hicks, 2005), these patterns of stochastic dependence in Experiments 1A and 2 reflect the likelihood that on some encoding trials, the face *and* location information were bound successfully with the stimulus word without necessarily being bound to one another. The fact that such dependence essentially doubled from Experiment 1A (about a 7 % difference between gender-correct vs. gender-incorrect) to Experiment 2 (about a 14 % difference) reinforces the notion that encoding factors are primarily responsible for these differences. Yet our attempts to cue location information with correct face cues at retrieval was not successful for either experiment, replicating the patterns shown previously by Starns and Hicks (2005, 2008) and Vogt and Bröder (2007).

Discussion

Across three experiments, reinstating associated face information at test did not improve location source memory. This

result is similar to other experimental outcomes (Starns & Hicks, 2005; Vogt & Bröder, 2007) but is unique in that gender source memory was successfully and powerfully cued when gender memory was tested. This bias-controlled method of source reinstatement developed by Starns and Hicks (2013) answers the call by Vogt and Bröder for paradigms examining cross-dimensional source memory to employ procedures that demonstrate within-dimension cuing. This null cross-dimensional cuing occurred even after an encoding procedure in which people were fully informed that their memory for both gender and source memory would be tested—a change that improved memory discrimination for both sources. This outcome argues against the possibility that people might not be paying attention enough to both source dimensions at encoding to allow for possible context–context binding.

Our findings also represent internally the paradox seen across prior studies of multidimensional source retrieval. On the one hand, we demonstrated stochastic dependence, suggesting that for many items both gender and location information were successfully bound to item information. That is, correct decisions on one dimension were more likely when the other dimension was correctly retrieved also. On the other hand, for the same set of items, reinstating face information at test improved gender discrimination only. If gender and location information were bound directly, one would expect reinstatement of face information to improve retrieval of location information as well (see Starns & Hicks, 2008). To the extent that a word alone at test cues location memory to a certain level of performance, adding the correct face cue to that test environment should improve location memory performance in a measurable way. Our null outcome therefore runs counter to what most memory models would predict in terms of how item-context cuing should influence activation of information in memory along a decision axis (e.g., Dennis & Humphreys, 2001; Hintzman, 1988).

Theoretically, the present findings reinforce the view that context or source dimensions are bound directly to item information but do not seem to be bound directly to one another naturally (i.e., context–context binding; Starns & Hicks, 2008). Stochastic dependence in our experiments was likely demonstrated for two reasons. First, the encoded word was always present in the retrieval cue and therefore capable of cuing a baseline level of both gender and location memory. The provision of face cues, if this information were directly bound to location information, should have improved the location decisions. Second, despite source dimensions being associated with learned items, it may also be possible that our encoding procedures did not favor a highly integrated memory trace that includes direct binding of sources. Even though in Experiment 2 we warned people that they would be tested on this information, this may not have prompted a thorough encoding approach that encourages ensemble encoding (e.g., Murnane & Phelps, 1993, 1994). To

underscore this point, recent work by Horner and Burgess (2013, 2014; see also Horner, Bisby, Bush, Lin, & Burgess, 2015) demonstrated stochastic dependence among triplets of encoded information that were all mutually and focally encoded. In this work, people learned triplets such as Obama–Pencil–Supermarket and were asked to associate them with a considerable encoding time. Because all three components were well integrated, each served as a good retrieval cue for the others.

On a related note, one concern about our methodology is that we did not assess old–new recognition prior to the sequence of source decisions for each stimulus, nor did we ask people for Remember–Know judgments associated with recognition. These methodological elements are common in other studies of stochastic dependence and of cross-dimension cuing (e.g., Meiser & Bröder, 2002; Starns & Hicks, 2005; Vogt & Bröder, 2007). Not requesting these decisions may allow unrecognized items that received correct guesses to contaminate our measures of gender and location discriminability and attenuate potential cuing effects. However, we believe that the lack of these decisions does not limit our findings. First, Starns and Hicks (2013) found that the within-dimension cuing effect on gender decisions was found whether people were asked for old–new recognition or were not asked. Second, the effect sizes for within-dimension cuing in our study were of a medium-to-large effect size and therefore promoted the conditions theoretically necessary to produce cross-dimension cuing if it were to be found. Third, from Experiment 1A to Experiment 2 we were able to increase performance on both source dimensions, which should theoretically create more items at retrieval that would have been associated with correct “old” responses and “Remember” responses had people been queried. The near-zero estimate of effect size for cross-dimension cuing (see Fig. 1) suggests little in terms of between-stimulus variability in the likelihood of generating cross-dimension cuing. Fourth, and finally, we produced estimates of stochastic dependence that were in line with prior work that also assessed old–new recognition and Remember–Know responses. Of course, whether our results actually generalize to conditions in which old–new and/or Remember–Know testing is required is an empirical question. Whether our results are limited to the face and location information used to define our sources is also an open question, but we note that other studies have found stochastic dependence among sources (as we did) that are not intrinsically integrated along some perceptual or conceptual classification or modality (e.g., Meiser, Sattler, & Weißer, 2008).

Key for future investigations will be the advent of experimental procedures in which evidence of context–context binding is both possible and demonstrable independent of item retrieval. Starns and Hicks (2008) attempted such a retrieval procedure by giving people aspects of the context in which objects were presented without showing the object. In such

test environments, cuing one context dimension—object color—was not capable of improving retrieval of the other—object location. But variations on such a procedure, or different procedures altogether, may indeed be successful. Perhaps modified encoding procedures in our paradigm that emphasize the association between faces and locations would be sufficient to create context–context binding, and cross-dimensional cuing would be successful. This prediction awaits further investigation. Regardless, the present results add to the evidence that direct context–context binding is not a necessary prerequisite of stochastic dependence (e.g., Boywitt & Meiser, 2012). There is no doubt that cross-dimension cuing is elusive as an empirical phenomenon, whereas stochastic dependence is almost always observed.

Author note Jason L. Hicks, Department of Psychology, Louisiana State University. Jeffrey J. Starns, Department of Psychology, University of Massachusetts Amherst.

We wish to thank Beth Turansky for her dedicated help in collecting the data.

References

- Bjork, R. A., & Richardson-Klavehn, A. (1989). On the puzzling relationship between environmental context and human memory. In C. Izawa (Ed.), *Current issues in cognitive processes: The Tulane Flowerree symposium on cognition* (pp. 313–344). Hillsdale: Erlbaum.
- Boywitt, C. D., & Meiser, T. (2012). Bound context features are integrated at encoding. *The Quarterly Journal of Experimental Psychology*, 65, 1484–1501.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale: Erlbaum.
- Dennis, S., & Humphreys, M. S. (2001). A context noise model of episodic word recognition. *Psychological Review*, 108, 452–478.
- Dodson, C. S., & Shimamura, A. P. (2000). Differential effects of cue dependency on item and source memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 26, 1023–1044.
- 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.
- Hicks, J. L., & Starns, J. J. (2015). Using multidimensional encoding and retrieval contexts to enhance our understanding of stochastic dependence in source memory. In B. H. Ross (Ed.), *The psychology of learning and motivation* (Vol. 62, pp. 101–140). Waltham: Academic Press.
- Hintzman, D. L. (1988). Judgments of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, 95, 528–551.
- Hintzman, D. L., Block, R. A., & Inskip, N. R. (1972). Memory for mode of input. *Journal of Verbal Learning and Verbal Behavior*, 11, 741–749.
- Horner, A. J., Bisby, J. A., Bush, D., Lin, W. J., & Burgess, N. (2015). Evidence for holistic episodic recollection via hippocampal pattern completion. *Nature Communications*, 6, 7462.
- Horner, A. J., & Burgess, N. (2013). The associative structure of memory for multi-element events. *Journal of Experimental Psychology: General*, 142, 1370–1383.
- Horner, A. J., & Burgess, N. (2014). Pattern completion in multielement event engrams. *Current Biology*, 24, 988–992.

- Johnson, M. K., Hashtroudi, S., & Lindsay, D. S. (1993). Source monitoring. *Psychological Bulletin*, 114, 3–28.
- Maner, J. K., Kenrick, D. T., Becker, D. V., Delton, A. W., Hofer, B., Wilbur, C. J., & Neuberg, S. L. (2003). Sexually selective cognition: Beauty captures the mind of the beholder. *Journal of Personality and Social Psychology*, 85, 1107–1120.
- Marsh, R. L., & Hicks, J. L. (2002). Comparisons of target output monitoring and source input monitoring. *Applied Cognitive Psychology*, 16, 845–862.
- Meiser, T. (2014). Analyzing stochastic dependence of cognitive processes in multidimensional source recognition. *Experimental Psychology*, 61, 402–415.
- Meiser, T., & Bröder, A. (2002). Memory for multidimensional source information. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28, 116–137.
- Meiser, T., & Sattler, C. (2007). Boundaries of the relation between conscious recollection and source memory for perceptual details. *Consciousness and Cognition*, 16, 189–210.
- Meiser, T., Sattler, C., & Weißer, K. (2008). Binding of multidimensional context information as a distinctive characteristic of remember judgments. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34, 32–49.
- Murnane, K., & Phelps, M. P. (1993). A global activation approach to the effect of changes in environmental context on recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 19, 882–894.
- Murnane, K., & Phelps, M. P. (1994). When does a different environmental context make a difference in recognition? A global activation model. *Memory & Cognition*, 22, 584–590.
- Rouder, J. N., Speckman, P. L., Sun, D., Morey, R. D., & Iverson, G. (2009). Bayesian *t* tests for accepting and rejecting the null hypothesis. *Psychonomic Bulletin & Review*, 16, 225–237.
- Slotnick, S. D., & Dodson, C. S. (2005). Support for a continuous (single-process) model of recognition memory and source memory. *Memory & Cognition*, 33, 151–170.
- Smith, S. M. (1979). Remembering in and out of context. *Journal of Experimental Psychology: Human Learning and Memory*, 5, 460–471.
- Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General*, 117, 34–50.
- Starns, J. J., & Hicks, J. L. (2005). Source dimensions are retrieved independently in multidimensional source-monitoring tasks. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31, 1213–1220.
- Starns, J. J., & Hicks, J. L. (2008). Context attributes in memory are bound to item information, but not to one another. *Psychonomic Bulletin & Review*, 15, 309–314.
- Starns, J. L., & Hicks, J. L. (2013). Internal reinstatement hides cuing effects in source memory tasks. *Memory & Cognition*, 41, 953–966.
- Vogt, V., & Bröder, A. (2007). Independent retrieval of source dimensions: An extension of results by Starns and Hicks (2005) and a comment on the ACSIM measure. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33, 443–450.