CrossMark

Positive and negative effects of gender expertise on episodic memory

Ainat Pansky[1,2] · Yaniv Oren[1] · Hadas Yaniv[1] · Ortal Landa[1] · Adi Gotlieb[1] · Eitan Hemed[1]

Abstract

In two experiments, we examined the role of differential levels of knowledge between the genders in different domains, which we term *gender expertise*, in accounting for differences in episodic memory performance. In Experiment 1, we validated the assumption of differential gender expertise among men and women and selected the categories for the subsequent experiments. In Experiment 2, participants from both genders studied exemplars from these female-oriented, male-oriented, and gender-neutral categories and were tested after 24 hours on studied items, critical lures, and unrelated lures. A gender-congruity effect was found in terms of the recognition rates of both studied items and critical lures: Participants from each gender recognized more studied items and more critical lures from gender-congruent categories than from gender-incongruent categories. A parallel pattern of results was found for subjective confidence, supporting the notion that gender congruity enhanced the phenomenological experience that an item was studied. Our findings highlight the unique role of gender expertise in accounting for gender-congruity effects in episodic memory performance, using a well-defined operationalization of gender expertise. These findings show that in addition to benefits in terms of enhancing true memory, gender expertise also has a "dark side" of increasing false memories.

Keywords Gender differences · False memory · Recognition · Confidence · Expertise

The examination of gender differences in episodic memory performance has yielded a mixed pattern of results: Whereas several large-scale studies have demonstrated superior performance of females over males (e.g., Herlitz, Nilsson, & Backman, 1997; Rehnman & Herlitz, 2007), other studies have failed to find any gender differences (e.g., Dehon, Laroi, & Van der Linden, 2011; Seamon, Guerry, Marsh, & Tracy, 2002). However, regardless of whether or not overall gender differences were found, distinctions between types of target stimuli in terms of their orientation—feminine or masculine—revealed an interesting interaction between the gender orientation of the stimulus and the gender of the rememberer. For example, compared with female participants, male participants demonstrated superior memory for car photographs but inferior memory for faces (Davies & Robertson, 1993; see also McKelvie, Standing, St. Jean, & Law, 1993). In

a similar vein, Powers, Andriks, and Loftus (1979) found superior memory for female-oriented event details among females (e.g., the female character's description) and superior memory regarding male-oriented details (e.g., a nearby automobile) among males.

Two classes of explanations were proposed to account for the interaction between the rememberer's gender and the gender orientation of the stimulus, hence the gender-congruity effect. One account attributes this effect to differential interest among the two genders in the target stimulus or its domain (e.g., Rehnman & Herlitz, 2007), which is expressed in the attention allocated to that stimulus (e.g., Powers et al., 1979). Alternatively, it was suggested that the advantage among each gender in remembering gender-oriented stimuli could be attributed to differential levels of knowledge (e.g., Davies & Robertson, 1993) or familiarity (e.g., Lewin & Herlitz, 2002) regarding the stimuli's domain. One should note that none of the reviewed studies attempted to tease apart these two potential contributions, nor do they provide evidence that distinctly supports one over the other. In fact, in many (if not most) of them, both accounts were proposed as plausible explanations for the findings (e.g., Davies & Robertson, 1993; McKelvie et al., 1993).

Our goal in the present study was to examine the role of differential levels of knowledge between the genders in

---

✉ Ainat Pansky
    pansky@research.haifa.ac.il

1   Department of Psychology, University of Haifa,
    3498848 Haifa, Israel

2   Institute of Information Processing and Decision Making, University
    of Haifa, 3498848 Haifa, Israel

different domains in accounting for differences in memory performance. Looking at this differential knowledge as *gender expertise* allowed us to gain from the rich existing literature on expertise. Previous studies in that field have demonstrated the role of expertise in boosting the memory for studied information (for a review, see Bédard & Chi, 1992). For example, compared with novices, expert chess players were better able to recall chess-game positions (Chase & Simon, 1973), baseball experts were found to exhibit superior memory performance for baseball-related information (Chiesi, Spilich, & Voss, 1979), and experienced bird-watchers were better at recognizing previously studied pictures of birds (Peeck & Zwarts, 1983).

However, several findings suggest that expertise-based knowledge organization and processing can also be a liability, alongside its advantages. For example, Arkes and Freedman (1984) found that compared with nonexperts, experts in a particular domain not only demonstrated a superior ability to distinguish between target sentences and nonrelated distracters but also demonstrated an inferior ability to distinguish between targets and distractors that were paraphrases or inferences that high-knowledge participants were likely to make. They suggested that expertise enables one to go beyond the information given, which is typically an advantage, but that expertise might turn out to be a disadvantage when it results in false memory for information that is merely inferred based on the prior knowledge. More recently, Baird (2003) found that investment experts recalled more investment-related words that were studied earlier than novices, but also exhibited more investment-related intrusions. In a similar vein, as the title of their recent paper suggests, Castel, McCabe, Roediger, and Heitman (2007) also demonstrated the "dark side of expertise," with football experts not only correctly recalling more studied names of football teams than nonexperts but also falsely recalling more unstudied distractors of the same nature. They interpreted their findings to suggest that the organizational processing that benefits experts' memory performance can also lead to the recollection of domain-relevant information that was not presented. Taking into account that experts tend to have more and stronger links among concepts and tend to base the organization of their knowledge on meaning compared with novices (Bédard & Chi, 1992), these findings of experts showing both more true and more false memory (Arkes & Friedman, 1984; Baird, 2003; Castel et al., 2007) are consistent with findings showing that, compared with more shallow processing, deep semantic processing increases both true and false memories (e.g., Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia, Neuschatz, & Goodwin, 1999).

In the present study, we applied these ideas and logic to the examination of gender differences in memory performance. In Experiment 1, we tested and confirmed the prediction that females and males develop differential gender expertise in

certain domains by comparing the number of items that come to mind among females compared with males, for a set of ostensibly feminine categories (e.g., cosmetics brands), ostensibly masculine categories (e.g., beer manufacturers), and ostensibly gender-neutral categories (e.g., European capitals). In Experiment 2 (the main experiment), female and male participants studied items from these female-oriented and male-oriented categories and were given a recognition test after 24 hours. Consistent with previous findings (e.g., Davies & Robertson, 1993; McKelvie et al., 1993), we expected a gender-congruity effect for the target items: The female participants were expected to recognize more true items from the female-oriented categories and fewer true items from the male-oriented categories than the males. As mentioned before, such a finding could be due to different interests between the genders, differential prior knowledge, or both. In real life, interest and familiarity tend to go hand in hand and are difficult to tease apart (Shepherd, 1981; Tobias, 1994). In the present study, our goal was not to rule out an interest hypothesis altogether but merely to demonstrate that gender expertise could in itself account for the gender-congruity effect. Toward this end, we also examined memory performance for nonpresented exemplars from the studied categories (i.e., critical lures), regarding which differential predictions could be derived from each theoretical account.

Interest is recognized as a critical cognitive and affective motivational variable that guides and focuses attention (see Hidi, 1995; Renninger & Hidi, 2011; Renninger & Wozniak, 1985; Schiefele & Krapp, 1996). Hence, according to the interest account of the gender-congruity effect (e.g., Powers et al., 1979), the larger interest in gender-congruent compared with gender-incongruent categories should be expected to increase focused attention to the items belonging to gender-congruent categories, which should result in enhanced item-based distinctive processing at encoding (e.g., Thomas & Sommers, 2005; see also Seamon et al., 2003). This item-based processing, in turn, should help to discriminate between studied items and critical lures (see, e.g., Gallo, 2010; Huff & Aschenbrenner, 2018; Hunt, 2003; McCabe, Presmanes, Robertson, & Smith, 2004), resulting in fewer false alarms for critical lures in the gender-congruent compared with the gender-incongruent categories.

In contrast, the gender-expertise account would make different predictions regarding the false recognition of critical lures. According to this account, given that more exemplars come to mind for high-knowledge categories than for low-knowledge categories, more categorical and associative activation can be expected for high-knowledge categories, both at encoding and at retrieval (see Castel et al., 2007). Previous studies have shown that nonstudied exemplars are more likely to be falsely recalled the more easily they tend to come to mind in norming studies when presented with the name of the category (Smith, Ward, Tindell, Sifonis, & Wilkenfeld,

2000; see also DeSoto & Roediger, 2014; Roediger & DeSoto, 2014). Given the higher gender expertise in the gender-congruent than in the gender-incongruent categories, nonstudied exemplars from these categories are expected to be more accessible, and are therefore more likely to be falsely recollected. Hence, consistent with the previous findings obtained for experts compared with nonexperts with expertise-relevant categories (e.g., Arkes & Freedman, 1984; Baird, 2003; Castel et al., 2007), the gender-expertise account would predict a higher false recognition rate of the critical lures from the gender-congruent categories (compared with the gender-incongruent categories), in parallel to the higher hit rate for the studied items from these categories, with no advantage in the ability to discriminate between these two types of items.

Experiment 1

Experiment 1 was aimed at validating the assumption that females and males develop differential gender expertise and at providing the data for selecting the target categorized lists for the memory experiment (Experiment 2).

Method

Participants Forty undergraduates from the University of Haifa, half females and half males, participated in the experiment.

Materials Twenty-four category names were used, eight of which were ostensibly female-oriented, eight were ostensibly male-oriented, and eight were neither (see Appendix Table 3 for the list of categories).

Procedure The participants received the 24 category names, one at a time, and each on a separate sheet of paper along with an example item from that category. For each category, the participants were asked to list all the items that came to mind (other than the listed example). The categories were presented in pseudorandom order, which was reversed for half of the participants within each gender group.

Results and discussion

The number of items listed for each category name served as a measure of category accessibility. First, we compared the accessibility of the ostensibly gender-neutral categories between the two genders and found similar category accessibility among the females (8.72) and among the males (9.80), $t(38)$ = 1.40, p = .170, d = 0.45.

In order to test our prediction that females and males have differential gender expertise, we conducted an ANOVA on category accessibility, with category orientation (feminine,

masculine) serving as a within-subjects factor and the gender of the participant (female, male) serving as a between-subjects factor. Category accessibility was comparable for the ostensibly feminine categories (5.54) and the ostensibly masculine categories (5.90), $F(1, 38)$ = 1.46, p = .234, η_p^2 = .04, and was also comparable overall for females (5.45) and for males (5.98), $F(1, 38)$ = 1.78, p = .190, η_p^2 = .05. However, as expected, a significant interaction was found between category orientation and gender, $F(1, 38)$ = 78.99, p < .001, η_p^2 = .68, with higher category accessibility among the females (6.59) than among the males (4.48) for the ostensibly feminine categories, $t(38)$ = 5.01, p < .001, d = 1.62, and lower category accessibility among the females (4.31) than among the males (7.48) for the ostensibly masculine categories, $t(38)$ = 5.67, p < .001, d = 1.84. In other words, we confirmed our prediction that males and females have differential gender expertise for different categories of information.

Another goal of Experiment 1 was to select a subset of the most feminine categories, the most masculine categories, and gender-neutral categories for Experiment 1. Toward this end, for each category, we calculated the difference between its mean category accessibility among the females and among the males. Based on this measure, we selected the 10 categories for Experiment 2: The three categories with the largest positive difference in category accessibility (M_{diff} = 3.97), $t(38)$ = 6.77, p < .001, d = 2.20, were selected as female-oriented categories, the three categories with the largest negative difference (M_{diff} = −3.67), $t(38)$ = 4.58, p < .001, d = 1.48, were selected as male-oriented categories, and the four categories with the smallest difference (M_{diff} = −0.86), $t(38)$ = 0.955, p = .346, d = 0.31, were selected as gender-neutral categories. The means (and standard deviations) of category accessibility by gender for each of these categories, as well as the comparisons of the gender means, are presented in Appendix Table 4.

Experiment 2

The feminine, masculine, and neutral categories selected on the basis of the data collected in Experiment 1 were used in Experiment 2 as target categories, and exemplars from these categories were presented for study and then tested after 24 hours. First, we expected to replicate previous findings of a gender-congruity effect for the studied items, with higher recognition rates for items from gender-congruent categories than for items from gender-incongruent categories (e.g., Davies & Robertson, 1993; McKelvie et al., 1993). In an attempt to differentiate a gender-expertise account of these findings from a differential-interest account, we also examined memory performance for nonpresented exemplars. As mentioned above, obtaining more false alarms for critical lures from gender-congruent compared with gender-incongruent categories

would support the gender-expertise account, whereas obtaining the opposite pattern of results would support the differential-interest account.

Method

Participants Eighty undergraduates from the University of Haifa, half females and half males, participated in the experiment. For half of the participants from each gender, the experiment was run by a female experimenter, whereas for the other half, the experiment was run by a male experimenter.

Materials As mentioned above, the 10 target categories were selected based on the data collected in Experiment 1: The three categories with the largest positive difference in category accessibility (kitchen utensils, cosmetics brands, and makeup products) were selected as female-oriented categories, the three categories with the largest negative difference (beer manufacturers, tools, and professional ball games) were selected as male-oriented categories, and the four categories with the smallest difference were selected as gender-neutral categories (European capitals, household appliances, Israeli prime ministers/presidents, and Israeli comedians). The first three gender-neutral categories served as target categories, whereas the fourth served as a practice category.

Following Smith et al. (2000), we calculated the output dominance of each item that was listed in Experiment 1—the frequency of participants who listed it. The 14 items with the highest output dominance in each categorized list, sorted in descending order, were used in Experiment 2. Two of the three items in the center of each list served as critical lures and were not presented to the participants in the learning phase. The remaining 12 items in each list served as the target stimuli.

The memory test was an old/new recognition test containing 54 items, one third of which were studied items (the two items presented in the middle of each studied list), one third were the critical lures (two for each list), and one third were unrelated lures.

Procedure In the study session, each participant was instructed to learn several word lists for a subsequent memory test. The nine target lists were presented on a computer screen, one at a time, preceded by one gender-neutral practice list. Each item was presented on the screen for 1 s, followed by a blank screen presented for 0.5 s. After the presentation of all 12 items in each list, the participant performed a nonverbal filler task for approximately 1 minute. The test session took place 24 hours after the study session. For each of the 54 test items that was presented on the screen, the participant was asked (1) whether or he or she recognizes it as having appeared in one of the lists that were studied the day before, and (2) how confident he or she was, between zero and 100, that this item was studied,

with zero representing complete confidence that the item was not studied and 100 representing complete confidence that it was. The participant's responses were provided orally and were written down by the experimenter.

Results and discussion

As the gender of the experimenter had no effect (or interaction with any of the independent variables) in any of the analyses, the data collected by the male and female experimenters were pooled. First, we compared the male and female participants' accuracy rates for the gender-neutral categories to ensure that the two groups did not differ in terms of memory performance for non-gender-specific information. Indeed, the females and the males exhibited comparable accuracy rates both for the studied items (.84 and .83, respectively), $t(78) = 0.25$, $p = .803$, $d = 0.06$, and for the critical lures (.30 and .32, respectively), $t(78) = 0.25$, $p = .804$, $d = 0.06$.

Next, we examined to what extent the recognition of true and false items from the gender-oriented categories was affected by gender congruity. The means and standard deviations of the proportion of items recognized in the various experimental conditions are presented in Table 1. Two ANOVAs were conducted, one on the proportion of hits, and one on the proportion of false alarms for critical lures, both with category orientation (feminine, masculine) as a within-subjects factor and the gender of the participant (female, male) as a between-subjects factor.

True recognition The overall hit rate was comparable for females (.75) and males (.73), $F(1, 78) = 0.364$, $p = .548$, $\eta_p^2 = .01$, and somewhat higher for the feminine categories (.77) than for the masculine categories (.71), $F(1, 78) =$

Table 1 Mean proportion recognized and confidence by item type, gender, and category orientation (Experiment 2)

Item type	Gender	Category orientation	Proportion recognized	Confidence
Studied	Female	Feminine	.85 (.16)	79 (12)
		Masculine	.65 (.20)	63 (17)
	Male	Feminine	.70 (.19)	67 (14)
		Masculine	.77 (.20)	71 (19)
Critical lure	Female	Feminine	.76 (.23)	65 (16)
		Masculine	.57 (.25)	54 (17)
	Male	Feminine	.44 (.28)	46 (20)
		Masculine	.65 (.25)	61 (20)
Mean	Female	Feminine	.80 (.20)	72 (16)
		Masculine	.61 (.23)	58 (17)
	Male	Feminine	.57 (.27)	56 (20)
		Masculine	.71 (.23)	66 (20)

Standard deviations appear in parentheses

4.53, $p = .036$, $\eta_p^2 = .06$. As expected, the interaction between the gender of the participant and the gender orientation of the category was significant, $F(1, 78) = 21.40$, $p < .001$, $\eta_p^2 = .22$. As shown in Fig. 1, the participants (correctly) recognized more studied items from the gender-congruent categories (.81) than from the gender-incongruent categories (.68), $t(79) = 4.53$, $p < .001$, $d = 0.69$.[1]

False recognition The false-alarm rate for the critical lures was higher among the females (.67) than among the males (.55), $F(1, 78) = 6.73$, $p = .011$, $\eta_p^2 = .08$, and comparable for the feminine categories (.60) and the masculine categories (.61), $F(1, 78) = 0.10$, $p = .756$, $\eta_p^2 < .001$. The interaction between the gender of the participant and the gender orientation of the category was significant here too, $F(1, 78) = 35.22$, $p < .001$, $\eta_p^2 = .31$. As shown in Fig. 1, the same pattern obtained for the hits was also found for the false alarms: The participants falsely recognized more critical lures from the gender-congruent categories (.70) than from the gender-incongruent categories (.51), $t(79) = 5.97$, $p < .001$, $d = 0.76$.

Signal detection analyses To determine whether the gender-congruity effects reported above are attributable to differences in sensitivity or in response bias, signal detection analyses were conducted. Following Koutstaal and Schacter (1997; see also Kensinger & Schacter, 1999; Schacter, Israel, & Racine, 1999), we chose to use nonparametric measures of sensitivity (A′) and response bias ($B_D″$) that are not based on any assumptions regarding the data distribution (see Grier, 1971; Hodos, 1970; Pollack & Norman, 1964). Values of A′ can range between zero and one, with higher values indicating greater sensitivity and .5 indicating chance performance, whereas values of $B_D″$ can range between −1 (extremely liberal responding) and +1 (extremely conservative responding). Because measures of A′ and $B_D″$ are undefined for hit rates and false alarm rates of zero or one, the data were first transformed, as recommended by Snodgrass and Corwin (1988), by setting P(x) = (x + .5)/(N + l) rather than P(x) = x/N. When participants showed below-chance sensitivity (A′ < .5, signifying a lower hit rate than false alarm rate), modified formulas were used, based on Aaronson and Watts (1987).

---

[1] The false recognition rate of unrelated lures was very low (0.16) and was not affected by gender, $F(78) = 1.02$, $p = .316$, $\eta_p^2{}_P = .01$, or by the interaction between gender and the gender-orientation of the category, $F(1,78) = 1.03$, $p = .313$, $\eta_p^2 = .01$. Therefore, the same ANOVA conducted for corrected recognition (i.e., hit rate corrected for response bias by subtracting the false alarm rate for unrelated lures) yielded the same pattern of results as we report for the uncorrected hits rates. In a similar vein, the ANOVA conducted for corrected false recognition (i.e., false-alarm rate for critical lures corrected for response bias by subtracting the false alarm rate for unrelated lures) yielded the same pattern of results as we report next for the uncorrected false alarm rate for critical lures.
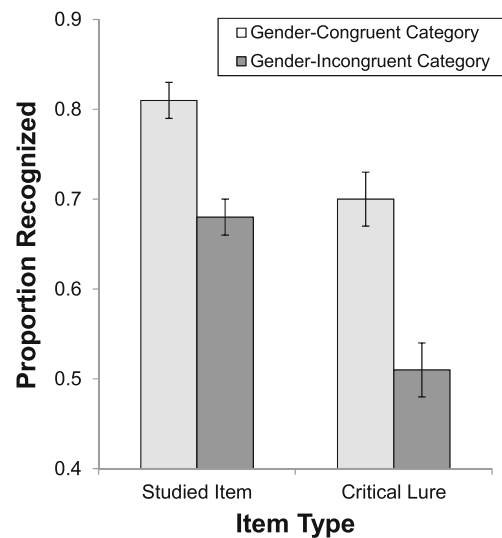

Fig. 1 Mean proportion of recognized studied items and recognized critical lures for gender-congruent and gender-incongruent categorized lists (Experiment 2). Error bars indicate 1 SEM

Following Koutstaal and Schacter (1997), three different types of signal detection analyses were performed. The first analysis assessed the sensitivity and bias associated with item-specific memory for studied items versus unrelated lures. As shown in Table 2, higher sensitivity was found for gender-congruent (A′ = .86) than for gender-incongruent categories (A′ = .80), $t(79) = 3.92$, $p < .001$, $d = .44$, indicating that gender congruity yielded a superior ability to discriminate between studied items and unrelated lures. A more liberal response criterion was also found for gender-congruent categories ($B_D″ = .12$) than for gender-incongruent categories ($B_D″ = .23$), $t(79) = 2.14$, $p = .035$, $d = .24$. The second analysis assessed the sensitivity and bias associated with item-specific memory for studied items versus critical-related lures. Here, gender congruity yielded no advantage in sensitivity, with relatively low and comparable A′s for gender-congruent (.58) and for gender-

Table 2 Signal detection measures of sensitivity (A′) and bias ($B_D″$) as a function of gender congruity, based on three different comparisons of "old" responses to (1) studied items versus unrelated lures (2) studied items versus critical lures (3) critical lures vs. unrelated lures

Comparison	Gender congruity	A′	$B_D″$
Hits vs. false alarms to unrelated lures (item-specific memory)	Gender-congruent categories	.86	.12
	Gender-incongruent categories	.80	.23
Hits vs. false alarms to critical lures (item-specific memory)	Gender-congruent categories	.58	-.14
	Gender-incongruent categories	.61	-.07
False alarms to critical lures vs. false alarms to unrelated lures (categorical knowledge)	Gender-congruent categories	.82	.18
	Gender-incongruent categories	.72	.25

incongruent categories (.61), $t(79) = 1.22$, $p = .227$, $d = .14$, but a more liberal response criterion for gender-congruent categories ($B_D'' = -.14$) than for gender-incongruent categories ($B_D'' = -.07$), $t(79) = 2.34$, $p = .022$, $d = .26$. The third and final signal detection analysis focused on the false recognition of critical-related lures versus unrelated lures, assessing the tendency of the participants to rely on categorical knowledge rather than on item-specific memory. Indeed, the tendency to rely on categorical knowledge was higher for gender-congruent categories ($A' = .82$) than for gender-incongruent categories ($A' = .72$), $t(79) = 4.82$, $p < .001$, $d = .54$, but the response criteria were not significantly different ($B_D'' = .18$ and $B_D'' = .25$ for gender-congruent and gender-incongruent categories, respectively), $t(79) = 1.51$, $p = .135$, $d = .17$.

Confidence Next, we examined the confidence judgments to gain more insight into the phenomenological experience associated with each type of item. One possibility is that although more gender-congruent items were recognized than gender-incongruent items, this difference is based on relatively low-confidence inferences that the items may have been studied. If this is the case, we would expect lower to comparable confidence judgments for items from gender-congruent categories compared with items from gender-incongruent categories. Alternatively, if the higher recognition rate of gender-congruent items is based on a relatively strong phenomenological experience that these items were indeed studied, we would expect higher confidence judgments for items from gender-congruent than from gender-incongruent categories.

In order to tease these two options apart, we repeated the main analyses conducted for the studied items and the critical lures with subjective confidence as the dependent variable (see Table 1 for the confidence means and standard deviations). Confidence that studied items were indeed studied was comparable for females (71) and males (69), $F(1, 78) = 0.822$, $p = .368$, $\eta_p^2 = .01$, and somewhat higher for the feminine categories (73) than for the masculine categories (67), $F(1, 78) = 7.76$, $p = .007$, $\eta_p^2 = .09$. The interaction between the gender of the participant and the gender orientation of the category was significant, $F(1, 78) = 22.52$, $p < .001$, $\eta_p^2 = .22$. As shown in Fig. 2, the participants were more confident in a studied item when it belonged to a gender-congruent category (75) than to a gender-incongruent category (65), $t(79) = 4.55$, $p < .001$, $d = 0.65$.

Confidence that critical lures were studied was not significantly different for females (60) and males (54), $F(1, 78) = 3.15$, $p = .08$, $\eta_p^2 = .04$, nor for the feminine categories (56) and the masculine categories (57), $F(1, 78) = 0.50$, $p = .483$, $\eta_p^2 = .01$. However, again, the interaction between the gender
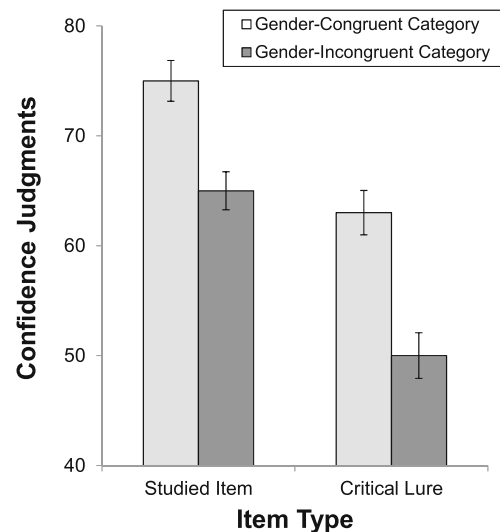

Fig. 2 Mean confidence that an item was studied for studied items and critical lures for gender-congruent and gender-incongruent categorized lists (Experiment 2). Error bars indicate 1 *SEM*

of the participant and the gender orientation of the category was significant, $F(1, 78) = 32.63$, $p < .001$, $\eta_p^2 = .30$. As for the studied items, the participants were also more confident that a critical lure was studied when it belonged to a gender-congruent category (63) than to a gender-incongruent category (50), $t(79) = 5.73$, $p < .001$, $d = 0.72$ (see Fig. 2).

To summarize, we replicated the previous findings of gender-congruity effects (e.g., Davies & Robertson, 1993; McKelvie et al., 1993; Powers et al., 1979), with higher (correct) recognition rates when one's gender was congruent with the gender orientation of a studied item. In addition, we also obtained the same pattern of results for the critical lures: Both genders falsely recognized more critical lures from gender-congruent than from gender-incongruent categories. The signal detection analyses showed that these gender-congruity effects are due to an increased reliance on categorical knowledge rather than to enhanced item-based sensitivity. This is evident in the superior sensitivity for gender-congruent than for gender-incongruent categories with regard to the discrimination of both studied items and critical-related lures from unrelated lures, but no difference in sensitivity with regard to the discrimination between that two former types of items. Rather, compared with gender-incongruent categories, gender-congruent categories were characterized by a more liberal response criterion in accepting both studied items and critical lures as having been studied. Finally, a similar pattern of gender congruity (both for studied items and for critical lures) was also found for the continuous measure of confidence, supporting the notion that gender congruity enhanced the phenomenological experience that an item was studied and not merely increased the occurrence of low-confidence inferences.

General discussion

Our main goals in the present study were (a) to replicate the previously obtained gender-congruity effect in terms of memory for the studied items, and (b) to examine the role of gender expertise in accounting for this gender-congruity effect on episodic memory.

Indeed, replicating previous findings (e.g., Davies & Robertson, 1993; McKelvie et al., 1993; Powers et al., 1979), we found higher (correct) recognition rates when one's gender was congruent with the gender orientation of a studied item than when it was incongruent, such that females correctly recognized more studied items belonging to feminine categories and males correctly recognized more studied items belonging to masculine categories. Most importantly, the same pattern of results was also obtained for the critical lures: Both females and males falsely recognized more critical lures from gender-congruent than from gender-incongruent categories, with no advantage in discriminating between studied items and critical lures. Finally, the same pattern (both for studied items and for critical lures) was also found for the continuous measure of confidence, which reflects the strength of the phenomenological conviction that an item was studied.

Because the gender-congruity effect in the recognition of studied items could be attributed to differential interests among the genders, gender expertise (i.e., differential prior knowledge), or both (e.g., Davies & Robertson, 1993; McKelvie et al., 1993), the examination of the effect of gender congruency on false memories was critical to tease apart the contribution of each factor. Obtaining fewer false alarms for critical lures for gender-congruent than for gender-incongruent categories would have supported the role of differential interests of the two genders, leading each gender to direct more focused attention to items from gender-congruent categories (see, e.g., Powers et al., 1979), resulting in enhanced item-based distinctive processing at encoding (e.g., Thomas & Sommers, 2005) and a better ability to discriminate between studied items and critical lures from these categories (e.g., Gallo, 2010; McCabe et al., 2004). The opposite pattern, which we obtained, of more false alarms to critical lures in gender-congruent categories, consistent with previous findings obtained for experts (e.g., Arkes & Freedman, 1984; Baird, 2003; Castel et al., 2007), and a comparable ability to discriminate between studied items and critical lures for gender-congruent and for gender-incongruent categories, supports the role of gender expertise in accounting for the findings. According to this account, females and males develop richer knowledge representations and stronger links among concepts in gender-relevant domains. Consequently, gender expertise may benefit performance for true memories, but may also yield more false memories

for highly accessible exemplars within the domain of expertise (see Castel et al., 2007; Smith et al., 2000). The confidence data support this interpretation, in demonstrating that both studied items and critical lures from gender-congruent categories were associated with a stronger phenomenological experience of having been earlier studied than their gender-incongruent parallels.

It is important to note that we do not claim that the differential interest of males and females cannot play a role in memory or that it did not play a role in previous studies that have demonstrated gender-congruity effects (e.g., Davies & Robertson, 1993; McKelvie et al., 1993; Powers et al., 1979). In fact, it is highly likely that differential interest and exposure, influenced perhaps by social norms, played a significant role in guiding attention to gender-oriented stimuli, and, consequently, in developing the richer gender-congruent representations in the first place (see McKelvie, 1981). However, once these richer representations exist, they can account for the gender-congruency effects in episodic memory, as the present findings show.

It is also important to stress that the classification of categories in the present study as feminine or masculine was based on sample means, such that women showed, on average, more expertise with regard to the feminine categories and men showed, on average, more expertise with regard to the masculine categories. Obviously, we are making no claims about individual females and males, each of whom can certainly be an expert in any of these categories (e.g., a female can be the leading word expert on beers, whereas a male can be the leading word expert on kitchen utensils). Furthermore, we are not making any value judgments about any of the categories, about people's expertise in them, or about people's interest in them.

With respect to generalizability, we feel that it is important to highlight the limits of the current research (see Simons, Shoda, & Lindsay, 2017). The participants in the experiments reported in this article were Israeli undergraduate students. Given that the gender-congruity effect obtained for the studied items replicates earlier findings obtained with American undergraduates (Powers et al., 1979), Canadian undergraduates (McKelvie et al., 1993), and both British children and undergraduates (e.g., Davies & Robertson, 1993) serving as participants, we believe that the new and parallel effects we found for false alarms and for confidence would also be reproducible in other countries and cultures across both student and nonstudent samples. Note, though, that the specific experimental materials used may need to be adapted for different cultural settings and age groups, and differential gender expertise for these materials should be validated in a preliminary experiment as was done in Experiment 1 of our study. We have no reason to believe that the results depend on other characteristics of the participants, materials, or context.

To conclude, our findings demonstrate the role of gender expertise in accounting for gender differences in episodic memory, using a clear operational definition of gender expertise (validated in Experiment 1). These findings show that in addition to benefits in terms of enhancing true memory, gender expertise also has a "dark side" of increasing false memories. It thus joins a host of other factors that have also been shown to increase both true and false memory, such as repeated questioning (e.g., Hyman & Pentland, 1996) and imagination (e.g., Goff & Roediger, 1998), in what has been termed the more-is-less effect (Toglia et al., 1999). In fact, the more-is-less effect of gender expertise we obtained in the present study could perhaps be seen as a particular case in which deeper semantic processing increases both true and false memories (Rhodes & Anastasi, 2000; Thapar & McDermott, 2001; Toglia et al., 1999).

Appendix A

Table 3 List of ostensibly feminine, ostensibly masculine, and ostensibly gender-neutral categories presented to the participants, along with the example item specified for each category (Experiment 1, translated from Hebrew)

Ostensibly feminine categories	Ostensibly masculine categories	Ostensibly neutral categories
Makeup products (e.g., *lipstick*)	Israeli soccer teams from the first and national leagues (e.g., *Maccabi Haifa*)	Central and south American countries (e.g., *Colombia*)
Israeli songs that competed in the Eurovision contest (e.g., *Abanibi*)	Empires throughout history (e.g., *the Roman Empire*)	Summer fruit (e.g., *grapes*)
Cosmetic brands (e.g., *Revlon*)	Israeli soccer and basketball players who played in Europe in the past and today (e.g., *Yossi Benayoun*)	European capitals (e.g., *Madrid*)
Israeli fashion retailers (e.g., *Castro*)	Beer manufacturers (e.g., *Heineken*)	Household appliances (e.g., *refrigerator*)
Israeli soap operas (e.g., *The Life Game*)	Kings who have ruled Israel throughout history (e.g., *King David*)	Israeli prime ministers and presidents (e.g., *Chaim Herzog*)
Kitchen utensils (e.g., *grater*)	Professional ball games (e.g., *baseball*)	Israeli comedians (e.g., *Shalom Asayag*)
Israeli jewelry retailers (e.g., *Magnolia*)	Israeli current ministers (e.g., *Shaul Mofaz*)	Birds (e.g., *hoopoe*)
Classic fairy tales (e.g., *Cinderella*)	Tools (e.g., *hammer*)	Trees that are not fruit trees (e.g., *pine*)

Appendix B

Table 4 Mean category accessibility by category and gender (Experiment 1)

	Category accessibility Females	Category accessibility Males	Category accessibility comparison
Feminine categories			
Kitchen utensils	8.80 (3.16)	6.70 (3.28)	$t(38) = 2.06, p = .046, d = 0.67$
Cosmetics brands	7.85 (3.91)	2.15 (1.69)	$t(38) = 5.98, p < .001, d = 1.94$
Makeup products	7.95 (1.57)	3.85 (1.90)	$t(38) = 7.44, p < .001, d = 2.41$
Masculine categories			
Beer manufacturers	5.55 (2.84)	9.45 (4.57)	$t(38) = 3.24, p = .002, d = 1.05$
Tools	5.75 (2.38)	10.10 (4.01)	$t(38) = 4.17, p < .001, d = 1.35$
Professional ball games	7.75 (2.31)	10.20 (3.07)	$t(38) = 2.85, p = .007, d = 0.92$
Gender-neutral categories			
European capitals	7.95 (3.36)	9.85 (4.97)	$t(38) = 1.42, p = .165, d = 0.46$
Household appliances	12.60 (3.48)	12.75 (2.07)	$t(38) = 0.16, p = .870, d = 0.05$
Israeli presidents and prime ministers	10.65 (2.81)	11.25 (3.11)	$t(38) = 0.64, p = .526, d = 0.21$
Israeli comedians	11.30 (5.11)	12.10 (6.46)	$t(38) = 0.43, p = .666, d = 0.14$

Standard deviations appear in parentheses

References

Aaronson, D., & Watts, B. (1987). Extensions of Grier's computational formulas for "A and B" to below-chance performance. *Psychological Bulletin, 102*, 439–442.

Arkes, H. R., & Freedman, M. R. (1984). A demonstration of the costs and benefits of expertise in recognition memory. *Memory & Cognition, 12*, 84–89.

Baird, R. R. (2003). Experts sometimes show more false recall than novices: A cost of knowing too much. *Learning and Individual Differences, 13*, 349–355.

Bédard, J., & Chi, M. T. (1992). Expertise. *Current Directions in Psychological Science, 1*, 135–139.

Castel, A. D., McCabe, D. P., Roediger, H. L., & Heitman, J. L. (2007). The dark side of expertise—Domain-specific memory errors. *Psychological Science, 18*, 3–5.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4*, 55–81.

Chiesi, H. L., Spilich, G. J., & Voss, J. F. (1979). Acquisition of domain-related information in relation to high and low domain knowledge. *Journal of Verbal Learning and Verbal Behavior, 18*, 257–273.

Davies, G., & Robertson, N. (1993). Recognition memory for automobiles: A developmental study. *Bulletin of the Psychonomic Society, 31*, 103–106.

Dehon, H., Laroi, F., & Van der Linden, M. (2011). The influence of encoding style on the production of false memories in the DRM paradigm: New insights on individual differences in false memory susceptibility? *Personality and Individual Differences, 50*, 583–587.

DeSoto, K. A., & Roediger, H. L. (2014). Positive and negative correlations between confidence and accuracy for the same events in recognition of categorized lists. *Psychological Science, 25*, 781–788.

Gallo, D. A. (2010). False memories and fantastic beliefs: 15 years of the DRM illusion. *Memory & Cognition, 38*, 833–848.

Goff, L. M., & Roediger, H. L. (1998). Imagination inflation for action events: Repeated imaginings lead to illusory recollections. *Memory & Cognition, 26*, 20–33.

Grier, J. B. (1971). Nonparametric indexes for sensitivity and bias: Computing formulas. *Psychological Bulletin, 75*, 424–429.

Herlitz, A., Nilsson, L. G., & Backman, L. (1997). Gender differences in episodic memory. *Memory & Cognition, 25*, 801–811.

Hidi, S. E. (1995). A reexamination of the role of attention in learning from text. *Educational Psychology Review, 7*, 323–350.

Hodos, W. (1970). Nonparametric index of response bias for use in detection and recognition experiments. *Psychological Bulletin, 74*, 351–354.

Huff, M. J., & Aschenbrenner, A. J. (2018). Item-specific processing reduces false recognition in older and younger adults: Separating encoding and retrieval using signal detection and the diffusion model. *Memory & Cognition.* Advance online publication. https://doi.org/10.3758/s13421-018-0837-1

Hunt, R. R. (2003). Two contributions of distinctive processing to accurate memory. *Journal of Memory and Language, 48*, 811–825.

Hyman, I. E., & Pentland, J. (1996). The role of mental imagery in the creation of false childhood memories. *Journal of Memory and Language, 35*, 101–117.

Kensinger, E. A., & Schacter, D. L. (1999). When true memories suppress false memories: Effects of ageing. *Cognitive Neuropsychology, 16*, 399–415.

Koutstaal, W., & Schacter, D. L. (1997). Gist-based false recognition of pictures in older and younger adults. *Journal of Memory and Language, 37*, 555–583.

Lewin, C., & Herlitz, A. (2002). Sex differences in face recognition: Women's faces make the difference. *Brain and Cognition, 50*, 121–128.

McCabe, D. P., Presmanes, A. G., Robertson, C. L., & Smith, A. D. (2004). Item-specific processing reduces false memories. *Psychonomic Bulletin & Review, 11*, 1074–1079.

Mckelvie, S. J., Standing, L., St. Jean, D., & Law, J. (1993). Gender differences in recognition memory for faces and cars: Evidence for the interest hypothesis. *Bulletin of the Psychonomic Society, 31*, 447–448.

Peeck, J., & Zwarts, J. (1983). Recognition memory for pictures of birds in relation to bird-watching skill. *American Journal of Psychology, 96*, 553–566.

Pollack, I., & Norman, D. A. (1964). A non-parametric analysis of recognition experiments. *Psychonomic Science, 1*, 125–126.

Powers, P. A., Andriks, J. L., & Loftus, E. F. (1979). Eyewitness accounts of females and males. *Journal of Applied Psychology, 64*, 339–347.

Rehnman, J., & Herlitz, A. (2007). Women remember more faces than men do. *Acta Psychologica, 124*, 344–355.

Renninger, K. A., & Hidi, S. (2011). Revisiting the conceptualization, measurement, and generation of interest. *Educational Psychologist, 46*, 168–184.

Renninger, K. A., & Wozniak, R. H. (1985). Effect of interest on attentional shift, recognition, and recall in young children. *Developmental Psychology, 21*, 624–632.

Rhodes, M. G., & Anastasi, J. S. (2000). The effects of a levels-of-processing manipulation on false recall. *Psychonomic Bulletin & Review, 7*, 158–162.

Roediger, H. L., & DeSoto, K. A. (2014). Confidence and memory: Assessing positive and negative correlations. *Memory, 22*, 76–91.

Schacter, D. L., Israel, L., & Racine, C. (1999). Suppressing false recognition in younger and older adults: The distinctiveness heuristic. *Journal of Memory and Language, 40*, 1–24.

Schiefele, U., & Krapp, A. (1996). Topic interest and free recall of expository text. *Learning and Individual Differences, 8*, 141–160.

Seamon, J. G., Goodkind, M. S., Dumey, A. D., Dick, E., Aufseeser, M. S., Strickland, S. E., & Fung, N. S. (2003). "If I didn't write it, why would I remember it?" Effects of encoding, attention, and practice on accurate and false memory. *Memory & Cognition, 31*, 445–457.

Seamon, J. G., Guerry, J. D., Marsh, G. P., & Tracy, M. C. (2002). Accurate and false recall in the Deese/Roediger and McDermott procedure: A methodological note on sex of participant. *Psychological Reports, 91*, 423–427.

Shepherd, J. W. (1981). Social factors in face recognition. In G. M. Davies, H. D. Ellis, & J. W. Shepherd (Eds.), Perceiving and remembering faces (pp. 55–79). New York, NY: Academic Press.

Simons, D. J., Shoda, Y., & Lindsay, D. S. (2017). Constraints on generality (COG): A proposed addition to all empirical papers. *Perspectives on Psychological Science, 12*, 1123–1128.

Smith, S. M., Ward, T. B., Tindell, D. R., Sifonis, C. M., & Wilkenfeld, M. J. (2000). Category structure and created memories. *Memory & Cognition, 28*, 386–395.

Snodgrass, J. G., & Corwin, J. (1988). Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *Journal of Experimental Psychology: General, 117*, 34–50.

Thapar, A., & McDermott, K. B. (2001). False recall and false recognition induced by presentation of associated words: Effects of retention interval and level of processing. *Memory & Cognition, 29*, 424–432.

Thomas, A. K., & Sommers, M. S. (2005). Attention to item-specific processing eliminates age effects in false memories. *Journal of Memory and Language, 52*, 71–86.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*, 37–54.

Toglia, M. P., Neuschatz, J. S., & Goodwin, K. A. (1999). Recall accuracy and illusory memories: When more is less. *Memory, 7*, 233–256.