Hum Nat (2008) 19:347–373
DOI 10.1007/s12110-008-9051-3

Sex Differences in Detecting Sexual Infidelity

Results of a Maximum Likelihood Method for Analyzing the Sensitivity of Sex Differences to Underreporting

Paul W. Andrews · Steven W. Gangestad · Geoffrey F. Miller · Martie G. Haselton · Randy Thornhill · Michael C. Neale

Published online: 4 September 2008

Abstract Despite the importance of extrapair copulation (EPC) in human evolution, almost nothing is known about the design features of EPC detection mechanisms. We tested for sex differences in EPC inference-making mechanisms in a sample of 203 young couples. Men made more accurate inferences (φ_{men}=0.66, φ_{women}=0.46), and the ratio of positive errors to negative errors was higher for men than for women (1.22 vs. 0.18). Since some may have been reluctant to admit EPC behavior, we modeled how underreporting could have influenced these results. These analyses indicated that it would take highly sex-differentiated levels of underreporting by subjects with trusting partners for there to be no real sex difference. Further analyses indicated that men may be less willing to harbor unresolved suspicions about their partners' EPC behavior, which may explain the sex difference in accuracy. Finally, we estimated that women underreported their own EPC behavior (10%) more than men (0%).

Keywords Accuracy · Bias · Error · Evolutionary psychology · Extrapair copulation · Infidelity · Jealousy · Sex differences

P. W. Andrews (✉) · M. C. Neale
Virginia Institute for Psychiatric and Behavioral Genetics, Virginia Commonwealth University, P.O. Box 980126, Richmond, VA 23298-0126, USA
e-mail: pandrews@vcu.edu

S. W. Gangestad · G. F. Miller
Department of Psychology, University of New Mexico, Albuquerque, NM, USA

M. G. Haselton
Communication Studies, University of California, Los Angeles, Los Angeles, CA, USA

R. Thornhill
Department of Biology, University of New Mexico, Albuquerque, NM, USA

While expectations of sexual exclusivity are a pervasive feature of human romantic relationships (Buss 1994), evidence suggests that selection has favored a certain amount of sex outside those relationships (extrapair copulation or EPC). For instance, men across cultures tend to express more interest than women in sex with multiple partners (McBurney et al. 2005; Schmitt 2003), especially when there are no constraints or costs to consider (Fenigstein and Preston 2007). Women appear to have psychological mechanisms that enhance their likelihood of engaging in an EPC during the fertile time of the menstrual cycle if their pair-bonded mate fails to display cues of good genes (Gangestad et al. 2005; Haselton and Gangestad 2006). Evidence also suggests that men have adaptations that evolved in response to female EPC, including male sexual jealousy that is expressed particularly toward younger partners and partners near the fertile part of the menstrual cycle (Buss and Shackelford 1997; Gangestad et al. 2002; Haselton and Gangestad 2006), and discriminative parental investment that covaries with men's certainty of paternity (Anderson et al. 1999).

Because people often respond to a partner's EPC in ways that are detrimental to the partner's interests (e.g., withdrawing investment, violence), mechanisms for detecting EPC behavior have probably coevolved with mechanisms for concealing EPC and avoiding being perceived as sexually unfaithful. This should have favored those who produced fewer clues of their EPCs, and partners who made better inferences from the available cues.

Selection might have shaped EPC detection mechanisms in sex-differentiated ways. For instance, men's paternity uncertainty puts them at potential risk of cuckoldry (Buss 2000), and this may have put them under stronger selection to correctly detect their partners' EPC behavior. Men's paternity uncertainty also puts them at risk of making investment errors (i.e., failing to invest in their own offspring, because they erroneously infer that they are some other man's offspring), which may have put them under stronger selection to correctly detect when their partners have not had EPCs. Conversely, women are more likely to experience costs imposed on them by jealous partners, including mate guarding, stalking, violence, homicide, and loss of male investment (Anderson et al. 1999; Buss and Shackelford 1997; Shackelford et al. 2003; Wilson et al. 1995). For this reason, women may have been under stronger selection to conceal EPCs from their partners.

In this paper, we investigate possible sex differences in the performance of mechanisms that are involved in detecting and making inferences about partner EPC behavior. We explore two performance parameters in these mechanisms: the accuracy of inferences about a partner's EPC behavior, and bias in errors.

The coevolutionary arms race between the sexes makes it difficult to predict sex-differentiated performance patterns. If men have been under stronger selection to correctly detect when their partners have and have not had EPCs, then they may make fewer errors and be more accurate than women. However, greater selection on women to conceal their EPCs should reduce men's performance and, if strong enough, could make men perform worse than women.

People can make two kinds of errors about the EPC behavior of their partners: they can fail to detect an EPC (a *false negative*) and they can erroneously infer that their partner has had an EPC (a *false positive*). The costs of the two errors may be more asymmetrical for men than for women. Specifically, it should be more costly for men to make false negative errors (because they can be cuckolded and fail to reproduce) than false positive errors (which lead to investment errors, but not the failure to reproduce). If so, then error management theory (Haselton and Buss 2000) predicts that selection should have favored men who biased their inferences toward the less costly error. In other words, men may be more likely to make false positive errors than women.

In a recent study (Lenoir et al. 2006), researchers asked each person in 90 adolescent couples whether they had had extrapair sex partners, and to report whether their partner had had extrapair sex partners (answered in “yes–no–don’t know” format). Thus, it was possible to cross-reference the inference made by one person with the self-reported EPC behavior of the partner. Among those who answered “yes” or “no” to the question about the partner’s EPC behavior, boys had a slightly higher percentage of correct inferences than girls (69% vs. 63%). Also, the ratio of false positive to false negative errors was higher for boys than for girls (1.5 vs. 0.5).

There are several limitations to this study. First, the population was composed of adolescent couples presenting themselves for reproductive care at a health clinic and an STD clinic, a sample likely to be biased in important ways. Second, adolescents answered questions about their sexual history in face-to-face interviews, a setting that may inhibit disclosure of sensitive information. Third, the “yes–no–don’t know” format for responding to the question about the partner’s EPC behavior is problematic because the people who answered “don’t know” may have had a “best guess” about the EPC behavior of their partners but been uncertain about it. For our purposes, it is important to include people who have a “best guess,” even if it is uncertain.

In this paper, we extend the research in several ways. First, we tested predictions in a sample of 203 young heterosexual couples attending the University of New Mexico rather than couples going to health clinics. Second, each person in the couple answered the pertinent questions alone, on a paper questionnaire, under conditions designed to promote anonymity. Third, we obtained a measure of subjects’ “best guess” about the extrapair sexual behavior of their partners and the degree of certainty associated with that inference.

In addition, some subjects who had EPCs may have been reluctant to disclose it to us—even though the conditions were designed to ensure privacy and anonymity. Underreporting is a serious confound in this area of research. Ideally, our analyses should be tested on the *true data* (what people actually did) rather than on the *observed data* (what they reported they did). In principle, sex differences in underreporting could generate sex differences in accuracy or error bias in the observed data, even if there were no such differences in the true data. If the underreporting rates were known, the researcher could reverse the effects of underreporting and estimate the true population by adjusting the proportions of people who had EPCs. The researcher could then conduct the analyses on the estimate of the true population. In practice, it is rarely possible for the researcher to ascertain actual underreporting rates for sensitive behavior. However, it should be possible to assume a specific value for the underreporting rate, obtain response pattern proportions adjusted for this rate, and then conduct the analyses on these revised estimates. By repeating the procedure across a range of plausible underreporting values, it is possible to assess how sensitive the results are to hypothesized levels of underreporting. To test our predictions about sex differences in accuracy and error bias, we develop and describe a procedure for conducting such sensitivity analyses.

If women have been under greater selection to conceal their infidelities, then they may be more likely to underreport. Using a closely related procedure that we will describe in more detail in another paper, we test this prediction by directly estimating the overall level of underreporting of EPC behavior by both men and women in this sample using maximum likelihood.

Throughout the paper, we sometimes use phrases like "faithful partner" or "partner infidelity" to avoid more burdensome phrases (e.g., "partner that did not have an EPC"). This is a matter of linguistic expedience and is not intended to imply moral judgment. Still, we recognize that such words naturally have moral connotations and we try to avoid them as much as possible.

Methods

Participants

The data used in this paper were collected more than 10 years ago as part of a broad study on couples' sexual behavior (Gangestad and Thornhill 1997). Participants were 203 heterosexual couples (men's average age=21.1 years; women's average age=20.0 years) who were involved in a romantic relationship for at least 1 month (average length of relationship=20.6 months; SD=18.6; range=1–108). At least one member of each couple was enrolled in a psychology course at the University of New Mexico and took part in this study in exchange for research credit. To create an incentive for partners to participate, a raffle was held at the end of each of the two semesters during which data were collected, at which time one couple won $100. About half of the subjects (53%) reported themselves as Caucasian, 36% as Hispanic, 5% as Native American, 3% as African American, 1% as Asian, and 2% other. Twenty of the couples were married, nine had children together, and five men and eight women had children from prior relationships (these are non-exclusive categories).

Procedure

Couples came to the study in groups of 1–4. After signing a consent form, couples were separated and each person was taken to a different room where they could complete the questionnaire in privacy. We stressed to all participants that their answers were completely anonymous. They were not to write their names on the questionnaires; we used an arbitrary identification number instead. We also stressed that their answers would not be seen by their partners, and that they did not have to answer any question they felt uncomfortable answering.

The questionnaires dealt with various aspects of sexual behavior. The EPC self-report question asked participants whether they had had sexual intercourse with someone other than their current partner while they were involved with their current partner. We phrased the question this way to avoid moral overtones that might inhibit honest answering.

The EPC inference question asked, "To your knowledge, has your partner ever had an affair behind your back?" We used the phrase "To your knowledge" because we wanted to know if subjects knew that their partner had an affair. Thus, if subjects answered "yes" to this question, they were coded as being 100% certain that their partner had an affair. If they answered "no" to this question, they were asked, "How likely do you think it is that your partner has had an affair behind your back without ever telling you?" Subjects answered this question by circling a number on a Likert scale with markings in 10% increments.

Table 1 shows the cross-tabulation of men's EPC inference probabilities with the EPC self-reports of their partners. Table 2 shows the corresponding information for women.

The EPC self-report question and the EPC inference question are different in several ways because they were not designed to address the issues that we are attempting to address in this paper. We discuss these differences and how they may have influenced the results in the limitations section of the paper.

Coding and Formatting the Data

We wanted to code the data in a way that carved nature at its joints. For instance, both affairs and EPCs are natural binary variables—either they happened or they didn't. For the EPC self-report question, we therefore coded it as a binary variable. If people reported having an EPC, we coded it as a "1"; if they reported not having an EPC, we coded it as a "0."

The inference data is essentially on a continuous scale, and the mathematical method we introduce for dealing with underreporting requires that we cross-reference a binary inference with a binary self-report. Since EPCs and affairs are both naturally binary variables, people face the task of trying to figure out which possibility is true about their partners. The information on which they base their inference will often be imperfect, so they also should assign a level of certainty to their inference. Thus, we think it natural to assume that people make inferences about their partners' affair behavior in binary format, but they also assign a level of certainty to their inferences, which can vary continuously. The inference data

Table 1 Frequency counts of men's probability ratings that their partners had an illicit EPC (broken down by the EPC self-report of the partner)

Men's estimated chance that partner had EPC (%)	Female partner's EPC self-report		
	No EPC	EPC	Missing
0	77	4	2
10	42	5	
20	12		
30	11		
40	6		
50	4		
60	1		
70	3		
80	3		
90			
99		2	
100		25	
Missing	5	1	
N	164	37	2
Total		203	

Table 2 Frequency counts of women's probability ratings that their partners had an illicit EPC (broken down by the EPC self-report of the partner)

Women's estimated chance that partner had EPC (%)	Male partner's EPC self-report		
	No EPC	EPC	Missing
0	85	9	2
10	38	14	1
20	3	3	
30	8	5	
40		3	
50	3		
60			
70	2		
80	1	1	
90		2	
99		1	
100		20	
Missing	2		
N	142	58	3
Total		203	

provides us with subjects' certainty about which of the two possibilities they think is more likely to be true. We use this information to recover the binary inferences. Those who report <50% certainty that their partners had affairs are coded as *trusting* that their partner did not have an EPC, and those reporting ≥50% certainty are coded as being *suspicious* that their partners had an EPC. This is a natural way of treating the data because it reflects the subject's "best guess" about which possibility is more likely to be true.[1] (Neither the results nor their interpretation change substantially if only those assigning a >50% chance are coded as being suspicious.) The cross-tabulation of subjects' valence with their partners' self-reports is presented in Table 3.

Each person in a couple can be defined by whether or not they reported having had an EPC and whether or not they were suspicious about their partner (four possible states). So every couple with complete information can be defined by one of 16 (4×4) possible categories. The pattern variable listed in Table 4 defines the 16 possible couple categories. From left to right, the first column of the pattern variable represents the man's self-reported EPC behavior (0=no EPC, 1=EPC). Column 2 represents the inference that the woman makes about her partner (0=trusting; 1=suspicious). Column 3 represents the woman's self-reported EPC behavior (0=no EPC, 1=EPC). Column 4 represents the inference that the man makes about his partner (0=trusting; 1=suspicious). Thus, a pattern of 0110 (category 7) refers to a couple in which the man reported no EPC (first column=0), the woman suspected that the man had an EPC (second column=1), the woman reported having an EPC

[1] In principle, we could also treat the inference (rather than the certainty associated with it) as a continuous variable. However, this leads to unnatural coding of the correctness of inferences. Suppose, for instance, that person A thought there was an 80% chance that her partner had had an affair, whereas person B reported a 90% chance that her partner had had an affair. Assume also that the partners of both of these people admitted to having EPCs. Under this coding scheme, we would say that A is 80% correct, B is 90% correct, and B is 10% more correct than A. However, we don't believe that people think this way. A and B would probably say that they are both inferring that their partners had affairs, and that both were correct. But they would probably agree that B was more certain about her inference than A.

Table 3 Cross-tabulation of participants' EPC self-reports and the valence of their partners' EPC inferences

		Men's EPC behavior		*N*	% Correct	Women's EPC behavior		*N*	% Correct
		No EPC	EPC			No EPC	EPC		
Women's inferences	Trusting	134	34	168	79.8				
	Suspicious	6	24	30	80.0				
	N	140	58	198					
	% Correct	95.7	41.4						
Men's inferences	Trusting					148	9	157	94.3
	Suspicious					11	27	38	69.2
	N					159	36	195	
	% Correct					93.1	75.0		

Of the total inferences made in each row or column, the percentage that are correct (% Correct) are given

(third column=1), and the man trusted that the woman did not have an EPC (fourth column=0).

This format preserved information about the non-independence between variables. There were 191 couples with complete information, and we tallied the observed couple frequencies for each of the 16 couple states (Table 4).

We cross-referenced subjects' EPC inferences about their partners with their partners' EPC self-reports. The four possible outcomes are: (1) a correct inference of EPC (*correct positive*); (2) a correct inference of no EPC (*correct negative*); (3) an

Table 4 Observed frequencies for each possible couple category

Couple Category	Pattern	Man's EPC self-report	Woman's EPC inference	Woman's EPC self-report	Man's EPC inference	Observed frequency
1	0000	0	0[b]	0	0[a]	110
2	0001	0	0[b]	0	1	5
3	0010	0	0[b]	1	0	5
4	0011	0	0[b]	1	1[a]	9
5	0100	0	1	0	0[a]	5
6	0101	0	1	0	1	1
7	0110	0	1	1	0	0
8	0111	0	1	1	1[a]	0
9	1000	1	0	0	0[a]	21
10	1001	1	0	0	1	3
11	1010	1	0	1	0	4
12	1011	1	0	1	1[a]	5
13	1100	1	1[b]	0	0[a]	9
14	1101	1	1[b]	0	1	2
15	1110	1	1[b]	1	0	0
16	1111	1	1[b]	1	1[a]	12
N						191

From left to right, the columns in the pattern variable are: (1) the man's EPC self-report (0=no EPC, 1=EPC); (2) the woman's EPC inference (0=no EPC; 1=EPC); (3) the woman's EPC self-report (0=no EPC, 1=EPC); and (4) the man's EPC inference (0=no EPC; 1=EPC)

[a] The EPC inferences of men that are correct (categories 1, 4, 5, 8, 9, 12, 13 and 16)

[b] The EPC inferences of women that are correct (categories 1–4 and 13–16)

incorrect inference of EPC (*false positive*); and (4) an incorrect inference of no EPC (*false negative*). In Table 4, we have marked the couple categories in which men and women have made correct inferences.

Calculating Accuracy and Error Bias

The simplest measure of accuracy is the proportion of inferences that are correct. If we assume, for the moment, that the observed data reflect the true couple frequencies (i.e., that there was no underreporting by either sex), then the inferences of men were correct 89.7% of the time, whereas those of women were correct 79.8% of the time.

We use a correlation measure to assess whether accuracy is equal in the two sexes. For dichotomous variables, one can use the tetrachoric correlation or the phi coefficient. The tetrachoric correlation is better for dichotomized variables whose underlying distribution is continuous, whereas the phi coefficient is better for variables that are truly dichotomous. Of our two measures, the EPC self-report is truly a dichotomous variable. However, as we have already noted, the EPC inference variable also has dichotomous properties. Moreover, for both men and women, the distribution of the EPC probabilities (Tables 1 and 2) is heavily weighted toward the extremes, which makes it somewhat like a dichotomous variable. For the analyses we report below, we assess accuracy via the phi coefficient, and we use the ratio of false positive to false negative errors to measure error bias.

The Underreporting Parameters

We assume that underreporting rates may be different for men and women, primarily because we hypothesize that women have been under stronger selection to conceal their EPCs from their partners. However, the underreporting patterns are probably even more complex. A great deal of dynamics regarding suspicions of infidelity may have already taken place by the time the couples answered the questionnaire. Some people in our sample who inferred that their mates had EPCs may have done so because (a) they already confronted their mate with proof or suspicions and the mate admitted the EPC or (b) the mate had already made an unsolicited confession. People who had already admitted their EPCs to their partners were also probably more likely to answer the EPC question on the questionnaire honestly. This would make participants' underreporting vary with the trusting or suspicious inferences of their partners, with lower underreporting probabilities among people with suspicious partners. We preserve flexibility in underreporting by providing separate parameters for men with trusting and suspicious partners (p and q, respectively) and for women with trusting and suspicious partners (r and s, respectively). If people with suspicious partners are more likely to disclose their EPCs than those with trusting partners, then $p > q$ and $r > s$.

One can imagine other ways underreporting could vary, which could lead to additional underreporting parameters, the inclusion of which would make our analyses more realistic. In the limitations section of the discussion, we discuss some of these ways. However, the more parameters we include in the model, the more difficult it becomes to analyze, and the more difficult the results become to describe.

We view our four underreporting parameters as a reasonable compromise between realism and analytical tractability.

Adjustment of Predicted Cell Frequencies for Underreporting

Table 4 shows the observed cell frequencies—it reflects what subjects reported about their own EPC behavior and what they reported about the EPC behavior of their partners. But, owing to underreporting, some subjects who had had an EPC probably failed to disclose it to us. Ideally, we would conduct our analyses on the true cell frequencies—i.e., subjects' actual EPC behavior—which could be recovered if we knew the underreporting rates. Table 5 is the transition matrix describing how the estimate of the true couple frequencies in each of the categories (on the rows) are reapportioned, via the underreporting parameters, among the categories (on the columns). Mathematically, this is simply achieved by application of the matrix formula: $\mathbf{E} = \mathbf{DY}$, where $\mathbf{D}$ is a 16×1 vector containing the estimates of the true population cell frequencies, $\mathbf{Y}$ is the 16×16 transition matrix (given in Table 5), and $\mathbf{E}$ is the (16×1) vector of the estimated cell frequencies following the effects of underreporting.

The logic for deriving the transition matrix in Table 5 is simple. As an example, consider couples whose true behavior falls in the pattern category of 1011. This category corresponds to couples in which the man reports having an EPC (first column=1), the woman trusts that the man did not have an EPC (second column=0), the woman reports having an EPC (third column=1), and the man suspects that she had an EPC (fourth column=1). Since we assumed that underreporting in subjects' self-reports may depend on what their partners say about them, men in this category underreport at rate p (because the woman's report on the man is negative) and women underreport at rate s (because the man's report on the woman is positive). Thus, underreporting makes proportion $p \times s$ of the true number of couples in category 1011 go into observed category 0001, proportion $p\times(1-s)$ go into category 0011, proportion $(1-p)\times s$ go into 1001, and proportion $(1-p)\times(1-s)$ remain in 1011. In Table 5, this partitioning can be seen by noting how couples in the row corresponding to true category 1011 are proportioned into the pertinent observed category columns. The rest of the matrix can be filled out using similar logic, which we leave to the reader. Thus, the rows in Table 5 describe how couples in a specific true category are divided into separate observed categories. For this reason, the proportions in each row all sum to 1. Furthermore, the columns describe how different true categories contribute to a specific observed category.

The Sensitivity Analyses

Ideally, we would conduct our sex difference tests on the true couple data. Although we do not know the underreporting probabilities, if we assume specific values for them, we can estimate the true couple proportions under those assumed values. To do this, we used Mx (Neale et al. 2003), a statistical program that uses maximum likelihood estimation (MLE). In the absence of underreporting, the approach would estimate 16 free parameters corresponding to the proportions in each of the 16 couple response patterns. These estimates of the true proportions in the population

Table 5 The transition matrix

True categories	Observed categories															
	0000	0001	0010	0011	0100	0101	0110	0111	1000	1001	1010	1011	1100	1101	1110	1111
0000	1	–	–	–	–	–	–	–	–	—	-	–	–	–	–	–
0001	–	1	–	–	–	–	–	–	–	–	–	–	–	–	–	–
0010	r	–	$1-r$	–	–	–	–	–	–	–	–	–	–	–	–	–
0011	–	s	–	$1-s$	–	–	–	–	–	–	–	–	–	–	–	–
0100	–	–	–	–	1	–	–	–	–	–	–	-	–	–	–	–
0101	–	–	–	–	–	1	–	–	–	–	–	–	–	–	–	–
0110	–	–	–	–	r	–	$1-r$	–	–	–	–	–	–	–	–	–
0111	–	–	–	–	–	s	–	$1-s$	–	–	–	–	–	–	–	–
1000	p	–	–	–	–	–	–	–	$1-p$	–	–	–	–	–	–	–
1001	–	p	–	–	–	–	–	–	–	$1-p$	–	–	–	–	–	–
1010	pr	–	$p(1-r)$	–	–	–	–	–	$r(1-p)$	–	$(1-p)(1-r)$	–	–	–	–	–
1011	–	ps	–	$p(1-s)$	–	–	–	–	–	$s(1-p)$	–	$(1-p)(1-s)$	–	–	–	–
1100	–	–	–	–	q	–	–	–	–	–	–	–	$1-q$	–	–	–
1101	–	–	–	–	–	q	–	–	–	–	–	–	–	$1-q$	–	–
1110	–	–	–	–	qr	–	$q(1-r)$	–	–	–	–	–	$r(1-q)$	–	$(1-q)(1-r)$	–
1111	–	–	–	–	–	qs	–	$q(1-s)$	–	–	–	–	–	$s(1-q)$	–	$(1-q)(1-s)$

The underreporting parameters are as follows: (1) the rate of underreporting by men with trusting partners (p); (2) the rate of underreporting by men with suspicious partners (q); (3) the rate of underreporting by women with trusting partners (r); and (4) the rate of underreporting by women with suspicious partners (s)

change once non-zero levels of underreporting are taken into account. In practice, the estimated true proportions are modified by the assumed underreporting probabilities via the transition matrix in Table 5. Although we do not know the actual underreporting rates, we can explore the effects of hypothetical levels of underreporting by entering them as fixed parameters in the model.

Mx permits parameter estimation subject to linear or non-linear constraints among the parameters. In the present case, we constrain each individual proportion to be non-negative and the sum of the 16 proportions to equal unity. Note that non-zero underreporting levels may result in estimates of the 16 proportions that differ greatly from those observed in the data. If underreporting exists, taking it into account will increase the estimated proportion of true EPCs.

Ascertaining the Probable Search Space We want to search the underreporting parameter space to find the regions in which there is a true significant difference between the sexes in their accuracy and the kinds of errors they make and the regions in which there are no significant sex differences. We first used Mx to find the probable bounds on the underreporting parameters so that we did not have to search all of the underreporting parameter space. The fit of a model is −2 times the natural logarithm of the likelihood (−2LL), with *lower values* indicating a better fit. To find the most likely model, we systematically varied the underreporting parameters. The model with all underreporting parameters equaling zero ($p=q=r=s=0.0$) had the best fit ($-2LL=620.83$). The fit is in chi-square units, so any model with a fit of $-2LL\geq 624.67$ (i.e., at least 3.84 chi-square units greater than the best one) is a statistically improbable model. To find the probable limits on each underreporting parameter, we set the other three parameters to zero and incrementally increased the remaining parameter to find the value for which the fit was equal to 624.67. Using this procedure, we found $0\leq p\leq 0.76$, $0\leq q\leq 0.14$, $0\leq r\leq 0.94$, and $0\leq s\leq 0.37$. In our discussions below, we refer to this four-dimensional region as the *probable search space.* In searching for the significance boundary for sex differences, there is little point in exploring regions outside this space because they are improbable.

Testing for Sex Differences Although we do not know what the true couple proportions are, we can make assumptions about the underreporting probabilities, estimate the true couple proportions, and then test for sex differences using the estimated proportions. Such a test will only be valid for the data that were estimated from that particular combination of underreporting probabilities. But we can systematically vary the underreporting probabilities to determine the regions in which there are significant sex differences and those in which there are no differences. Doing so may permit inferences about whether the true couple population is likely to fall in the significant regions of the space.

With MLE, one can test whether two models, one of which is nested within the other, are statistically different from each other. A model specifies the hypothesized or expected relationships among variables (e.g., the equations relating the observed couple frequencies to the true couple frequencies). A second model is nested within a first if the only way in which they differ is that the second model has more constraints. Because the difference in the fits of two nested models ($-2\Delta LL$) is asymptotically distributed as a chi-square, it is possible to conduct significance testing, where the degrees of freedom are the differences in the number of constraints between the two models.

Exploring the Probable Search Space to Find the Significance Boundary for a Global Sex Difference For a given set of underreporting probabilities, we want to know whether men and women are significantly different from each other with respect to accuracy or bias or both. We first conduct a test in which both accuracy and error bias can contribute to a sex difference, which we refer to as a *global* sex-difference test. As an example, assume that everyone is honest such that $p=0$, $q=0$, $r=0$, and $s=0$. We first run a model in which the true couple proportions must be estimated for these underreporting probabilities subject to two constraints: (1) the sum of the couple proportions must equal 1; and (2) all the couple categories must have non-negative proportions. This is the baseline or *saturated* model, and the fit is $-2LL=620.83$. Under this model, men were more accurate than women in their EPC perceptions ($\varphi_{men}=0.66$, $\varphi_{women}=0.46$) and were more likely to make false positive errors (the ratio of false positive errors to false negative errors was 1.22 for men and 0.18 for women).

We then run a second model, which is identical to the saturated model (including the same underreporting probabilities) except that the true proportions must be estimated subject to two additional constraints: (1) men and women must be equally accurate; and (2) men and women must have equal error bias ratios. Because the second model only differs from the saturated model in that there are additional constraints, it is nested within the saturated model. We call the second model the *constrained* model because it is more constrained than the saturated model. The fit of the constrained model is $-2LL=631.75$, and the difference in fit between the constrained and the saturated models is $-2\Delta LL=10.92$, which is in chi-square units. Because there are two additional constraints in the constrained model, the change in the degrees of freedom is two, and the constrained model is significantly different from the saturated model ($p=0.004$). Technically, this means that it is highly improbable that a population of couples in which there were no sex differences in either accuracy or error bias could have generated the observed couple frequencies if everyone was honest. Put more simply, it is evidence that, if everyone was honest, there is a significant global sex difference.

If we systematically varied the underreporting probabilities, each time testing for significance in the way we just described, we would find that there were some regions of the probable search space in which men were significantly different from women, and some regions in which the differences were not significantly different. We want to identify the critical underreporting values that demark the significance boundaries. To do this, we assume values for three of the underreporting probabilities that fall within the probable search space and solve for the value of the fourth that is required for there to be no significant global sex difference. For instance, if we assume that $p=0$, $q=0$, and $s=0$, the critical value of r needed for there to be no significant global sex difference in either accuracy or error bias is 0.48.

Exploring the Probable Search Space to Find the Significance Boundary for a Sex Difference in Accuracy We conducted another set of analyses in which we explored the probable search space for regions in which there were sex differences in accuracy. In these analyses, only accuracy is allowed to contribute to a sex difference. The procedure is identical to those involved in the global analyses except in the following ways. First, the saturated model involves estimating the true couple frequencies under the assumed underreporting probabilities subject to three constraints: (1) the sum of the couple proportions must equal 1; (2) couple categories cannot have negative proportions; and (3) the sexes must have equal error bias because only accuracy is allowed to contribute to a sex difference. Second, in the constrained model, the true couple proportions are estimated subject to an additional constraint: men and women must have equal accuracy. A significant difference in fit means that it is highly improbable that a population of couples in which there was no sex difference in accuracy could have generated the observed couple proportions under the assumed set of values for the underreporting parameters.

Exploring the Probable Search Space to Find the Significance Boundary for a Sex Difference in Error Bias Finally, we conducted a set of analyses in which we explored the probable search space for regions in which there were significant sex differences in error bias. In these analyses, only error bias is allowed to contribute to a sex difference. The saturated model involves estimating the true couple frequencies under the assumed underreporting probabilities subject to three constraints: (1) the sum of the couple proportions must equal 1; (2) couple categories cannot have negative proportions; and (3) the sexes must have equal accuracy because only error bias is allowed to contribute to a sex difference. With the constrained model, the true couple frequencies are estimated subject to the additional constraint that men and women have equal error biases. A significant difference in fit means that it is highly improbable that a population of couples in which there was no sex difference in error bias could have generated the observed couple frequencies under the assumed set of values for the underreporting parameters.

Results

Descriptives

The first EPC inference question asked subjects, “To your knowledge, has your partner had an affair behind your back?” We used the phrase “to your knowledge” because we wanted to know if subjects *knew* that their partner had had an affair. Twenty five men and 20 women answered “yes” to this question, so they were coded as being 100% certain that their partner had had an affair. Everyone who answered “no” also answered the second question (which asks subjects to rate the chance that their partner had an affair) with less than a 100% probability. Moreover, three of the subjects (two men, one woman) who answered “no” to the first question answered the second question by writing in by hand 99% on the Likert scale. Altogether, these results suggest that subjects tended to interpret the "to your knowledge" question strictly, to the point of answering "no" when they only had a slight doubt about it.[2]

The subjects who answered "yes" to the "to your knowledge" question constitute an interesting subsample because they report knowing that their partners had affairs. We did not ask questions to ascertain the basis for that knowledge. It may well be that, for these individuals, the partner's affair is an open fact within the relationship. However, we do not exclude them from the analyses because these couples have a lot of relationship history behind them at the time we gave them this questionnaire. Even when an affair is an open fact within the relationship at the time of the questionnaire, most such affairs were probably concealed at some earlier time. Thus, open acknowledgment of a prior affair between a couple will often be the product of the EPC detection processes we are attempting to study. We return to this point in the discussion.

Of the 198 men who provided information about their own EPC behavior and their inferences about the affair behavior of their partners, 58 (29.3%) reported having an EPC during their current relationship, while 38 (19.5%) tended to believe that their female partners had had an affair (Table 2). Of the 195 women who provided both pieces of information, 36 (18.5%) reported having an EPC during their current relationship, while 30 (15.1%) tended to believe that their male partners had had an affair.

We assessed the phi coefficients between all binary variables: (a) men's EPC inferences with their own EPC behavior ($\varphi=0.33$, $N=194$, $p<0.001$); (b) women's EPC inferences with their own EPC behavior ($\varphi=0.23$, $N=200$, $p=0.001$); (c) men's and women's EPC inferences ($\varphi=0.34$, $N=195$, $p<0.001$); (d) men's and women's EPC behavior ($\varphi=0.30$, $N=198$, $p<0.001$); (e) men's EPC inferences with women's EPC behavior ($\varphi=0.67$, $N=195$, $p<0.001$); and (f) women's EPC inferences with men's EPC behavior ($\varphi=0.47$, $N=198$, $p<0.001$).

Underreporting Analyses

The Significance Boundary for a Global Sex Difference Figure 1 represents the results of the analyses in which both accuracy and error bias can contribute to a global difference between the sexes. The points denote positions on the significance boundary. The four-dimensional underreporting parameter space has been compressed to three dimensions. The compressed dimension is q and variation in this parameter influences the thickness of the boundary, which is represented by the point

[2] It is possible that some subjects who answered "yes" to the "to your knowledge" question were not actually 100% certain that their partners had affairs. However, it seems reasonable to assume that if there were such subjects, they would still have been at least 50% certain that their partner had an affair. Thus, this would not influence our coding of them as being suspicious that their partner had an affair or the results of our analyses about accuracy and error bias that we report. It would only influence the certainty of that inference, and the results that depend on certainty.

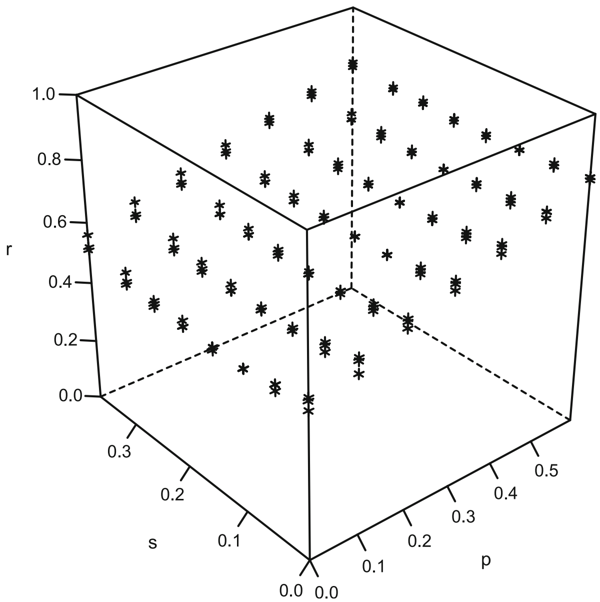

Fig. 1 The significance boundary in the plausible search space where both accuracy and error bias can contribute to a global difference between the sexes. The *x*-axis is the probability of underreporting by men with trusting partners (*p*), the *y*-axis is the probability of underreporting by women with suspicious partners (*s*), and the *z*-axis is the probability of underreporting by women with trusting partners (*r*). The influence of variation of the fourth underreporting parameter, the probability of underreporting by men with suspicious partners (*q*), is represented by the closely clustered points in the *z*-axis

clusters in the *r* axis. The possible variation in *q* (0.0–0.14) does not have a great impact on the significance boundary, so we neglect further discussion of it.

The space below the plane, which encompasses the origin, is the region where there is a significant difference between the sexes. That is, it is where men are more accurate than women, or they make more false positive errors than women, or both. The space above the plane is the region of non-significance. We do not know precisely where the couple population truly falls in the probable search space. However, if there were no underreporting (the point represented by the origin), both accuracy and error bias contribute to a significant sex difference, with $\varphi_{men}=0.66$, $\varphi_{women}=0.46$, 1.22 as the ratio of false positive to false negative errors for men, and 0.18 as the ratio of false positive to false negative errors for women. Fairly high values for the probability of underreporting by women with trusting partners must be assumed to move the population out of the region of significance. The minimum value of *r* that falls on the significance boundary is 0.46 (when $p=0.0$, $q=0.14$, and $s=0.15$). This is the smallest possible value of *r* that could potentially move the population out of the region of significance. For larger values of *p*, the minimum value of *r* needed to move the population out of the region of significance also increases. We now consider how differences in accuracy and error bias contribute separately to the global patterns.

The Significance Boundary for a Sex Difference in Accuracy Figure 2 represents the results of the analyses in which only the accuracy of EPC inferences can contribute to a sex difference. The space below the significance boundary is the region where men are significantly more accurate than women. If there were no underreporting, then the couple population would fall in the region of significance. The region of significance is smaller than it was for the global tests, which suggests that accuracy cannot fully explain the global pattern. Of interest is the fact that the significance plane is torqued toward the origin. The smallest possible value that r can take to move the population out of the significance region is 0.23 (when p=0.0, q=0.14, and s=0.0), and it climbs with both p and s.

The Significance Boundary for a Sex Difference in Error Bias Figure 3 represents the results of the analyses in which only error bias in EPC inferences can contribute to a sex difference. The region below the boundary is where the ratio of false positive to false negative errors is significantly higher for men than for women. If we assume that there was no underreporting, then the couple population would fall in the region of significance. The region of significance is smaller than it is for the

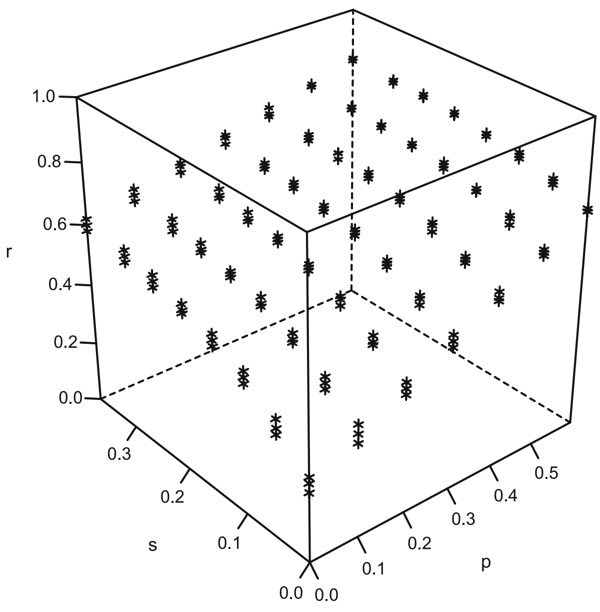

Fig. 2 The significance boundary in the plausible search space where only the accuracy of EPC inferences can contribute to a difference between the sexes. The x-axis is the probability of underreporting by men with trusting partners (p), the y-axis is the probability of underreporting by women with suspicious partners (s), and the z-axis is the probability of underreporting by women with trusting partners (r). The influence of variation of the fourth underreporting parameter, the probability of underreporting by men with suspicious partners (q), is represented by the closely clustered points in the z-axis

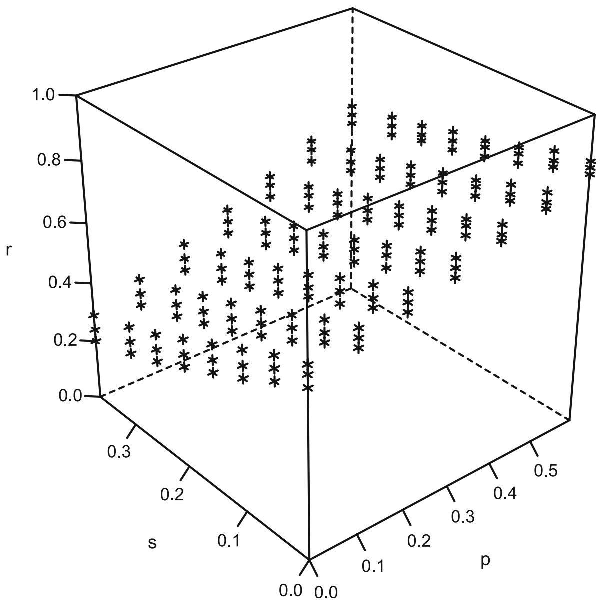

Fig. 3 The significance boundary in the plausible search space where only error bias can contribute to a difference between the sexes. The *x*-axis is the probability of underreporting by men with trusting partners (*p*), the *y*-axis is the probability of underreporting by women with suspicious partners (*s*), and the *z*-axis is the probability of underreporting by women with trusting partners (*r*). The influence of variation of the fourth underreporting parameter, the probability of underreporting by men with suspicious partners (*q*), is represented by the closely clustered points in the *z*-axis

global tests, which suggests that error bias cannot fully explain the global pattern. Since accuracy also couldn't fully explain the global pattern, it appears that accuracy and error bias both contribute to the global pattern. The minimum value that r can take for there to be no true sex difference in error bias is 0.20 (when $p=0.0$, $q=0.0$, and $s=0.37$).

Discussion

We undertook this study to explore the design of EPC detection mechanisms. If we assume that everyone was honest about their EPC self-reported behavior, men were more accurate than women in their EPC inferences ($\varphi_{men}=0.66$, $\varphi_{women}=0.46$) and were more likely to make false positive errors (the ratio of false positive errors to false negative errors was 1.22 for men and 0.18 for women).

However, the observed data are probably not completely accurate because some people probably failed to disclose their EPC behavior on the questionnaire. We first explored the probable search space to identify the regions in which accuracy and error bias both contributed to a significant global difference between the sexes and the regions in which there was no significant global sex difference (Fig. 1). While we do not know where the couple population truly falls in this space, we must assume a fairly high underreporting probability by women with trusting partners to move the population out of the region of significance ($r \geq 0.46$). If men with trusting partners are also underreporting with a non-zero probability (i.e., $p > 0$), it becomes even more difficult to move the population out of the region of significance.

We think it unlikely that the true probability of underreporting by women with trusting partners is high enough (nearly 50%) to put the couple population in the region of non-significance. When asking people to provide sensitive information about themselves, research indicates that the most important issue in reducing underreporting and socially desirable responding is how anonymous the conditions are (Schaeffer 2000; Tourangeau and Smith 1996; Turner et al. 1998). Several features of this study were designed to promote anonymity. First, participants were not interviewed in person, but were given questions in pencil and paper format. Second, participants knew that arbitrary identification numbers would be used so their names could not be associated with their data. Third, each person in a couple was put in a separate room so they could answer the questionnaire in private. Finally, subjects were reminded that they need not answer any question they felt uncomfortable answering. Only two women (Table 1) and three men (Table 2) refused to answer the EPC question, which suggests that the number of people who answered the question dishonestly may have been low as well.

In short, it seems likely to us that there is a significant global sex difference in the true couple population. Even if there is no significant global sex difference, our analyses indicate that the level of underreporting would have to be highly sex-differentiated, with a high absolute probability of underreporting by women with trusting partners, under conditions that were designed to promote anonymity. This would instead be consistent with the hypothesis that women have been under stronger selection than men to conceal their EPCs.

In any event, these analyses allowed both accuracy and error bias to contribute to the sex difference. Can we make any inference about whether accuracy or error bias contributes more to the global sex-difference pattern? We conducted separate analyses in which we identified the significance boundary where only accuracy was allowed to contribute to a sex difference (Fig. 2), and the boundary where only error bias was allowed to contribute to a sex difference (Fig. 3). These boundaries differ from the global pattern in a crucial way. Whereas the global significance boundary is relatively invariant with respect to the probability of underreporting by women with suspicious partners (see Fig. 1), the accuracy boundary is torqued toward low values of s (Fig. 2) and the error bias boundary is torqued toward high values of s (Fig. 3). Whether accuracy or error bias contributes more to the global pattern may depend largely on the probability of underreporting by women with suspicious partners. If the probability is low (i.e., s is close to 0), then sex differences in error bias are probably contributing more to the global sex-difference pattern. If the probability is high (i.e., s is closer to 0.37), then sex differences in accuracy are contributing more to the global pattern.

In these analyses, we did not make any assumptions about the values of the underreporting parameters save that we restricted them to the probable search space ($0 \leq p \leq 0.76$, $0 \leq q \leq 0.14$, $0 \leq r \leq 0.94$, and $0 \leq s \leq 0.37$). However, even within the probable search space, some parameter values might be implausible. In the graph of the global tests (Fig. 1), there is a boundary point defined by $p=0.0$, $q=0.07$, $r=$

0.59, and s=0.0. This point might be implausible for two reasons. First, it specifies that men with trusting partners have a lower probability of underreporting than have men with suspicious partners ($p<q$). But, as we discussed above, one likely reason why people in this sample might have reported suspicion about their partners is because some of their partners admitted to having an EPC. Under such circumstances, there seems less of an incentive or motive to deny the EPC on the questionnaire, and this should make underreporting less likely among those with suspicious partners (i.e., it should be the case that $p>q$ and $r>s$). Second, it assumes that men with suspicious partners had a higher probability of underreporting than women with suspicious partners (i.e., $q>s$). However, because women are more likely to be victimized for having EPCs, they have probably been under greater selection to conceal them. If so, women should be more likely to underreport than men (i.e., it should be the case that $r>p$ and $s>q$).

If we assume that these additional constraints hold, then the true couple population is probably pushed away from the origin into regions of the plausible search space where the contribution of error bias to the global sex difference is smaller and the contribution of accuracy is higher. This is because the requirement that $s > q$ forces the probability of underreporting by women with suspicious partners to take on non-zero values.

What Happens When Underreporting Rates Are Constant within Each Sex?

In conducting our analyses, we allowed underreporting to vary within each sex as a function of whether the partner was suspicious. To some, this might seem unduly complex. If underreporting is kept constant within each sex, then the probable search space is restricted such that $0 \leq (p=q) \leq 0.14$ and $0 \leq (r=s) \leq 0.37$. Under these restrictions, there is no combination of underreporting parameters that can move the true couple population out of the region of significance for either a global difference or a difference in accuracy. The reason is that the underreporting rates by people with suspicious partners and the rates by people with trusting partners have opposing effects on accuracy. (This can be verified in Table 3 by considering how underreporting by a person with a suspicious or trusting partner shifts a couple into or out of a cell representing a correct inference.) When these parameters are forced to be equal within each sex, their effects tend to cancel each other out and there is not enough leverage to move the population out of the significance region.

Why Are Men More Accurate Than Women?

We must consider the possibility that men's greater accuracy is simply an artifact of the way the questions were phrased. As noted above, there are several differences in the phrasing of the EPC self-report question and the EPC inference question, and these differences could have contributed to the results. First, the EPC self-report question asks about sexual intercourse, and the EPC inference question asks about an affair. It is possible that, for some subjects, "affair" does not necessarily connote extrapair sexual relations—some may have interpreted it as including emotional infidelity. Similarly, it is possible that some subjects interpreted "affair" as including non-copulatory sex (e.g., oral sex).

We surveyed 16 online dictionaries to ascertain the common understanding of "affair." "Affair" has multiple definitions, but the pertinent definition directly referred to a "sexual relationship" in nine of the dictionaries, and four defined it as a "romantic or sexual relationship." The remaining three defined it as a "romantic" or "amorous" relationship, or a "love affair," but these terms were then defined as having a sexual component. It was common for the dictionaries to say that the sexual relationship is usually "secret," "illicit," or "between two people who are not married to each other." It was also common for the dictionaries to say that the relationship could be of "limited" or "brief" duration. Thus, dictionaries indicate that "affair" typically connotes an illicit sexual relationship, which may be of a brief or long duration, and which may or may not involve feelings of love or attachment.

Another potential problem derives from the fact that an affair is usually viewed as illicit, which was highlighted in the affair inference questions by using the phrase "behind your back." Some subjects may have considered their partners' EPCs to be licit, thus reporting no affair inference even though they were aware of the EPC or suspected it. If there were sex differences in whether subjects viewed their partners' EPCs as illicit, then these could have contributed to the results. However, evidence suggests that this may not have been a problem. An early study found that young men and women dating in college develop expectations very early in their relationships that their partners will be sexually exclusive (Hansen 1985). A more recent study found that these expectations develop even in dating situations that are not *explicitly* exclusive, and extrapair sexual behavior that violates those expectations is viewed by both men and women as unfaithful (Yarab et al. 1999). Importantly, no sex differences were found in the degree to which extrapair sexual behavior was considered to be unfaithful. Thus, even in the early phases of a relationship where a subject's partner may also still have been dating other people, the subject is likely to view a partner's EPC as illicit. For similar reasons, the number of couples with "open" relationships in which EPCs were explicitly sanctioned was probably negligible.

This research suggests that most subjects in our sample probably interpreted the affair question as asking whether their partner had engaged in illicit extrapair sexual relations, and subjects would tend to have a broad presumption of illicitness about their partners' extrapair sexual behavior.[3]

While we cannot completely rule out the possibility that some subjects may have interpreted "affair" in different ways, we think it unlikely that the sex differences we found are an artifact of the methodology. If some people had based their affair inferences on information that their partners had only engaged in non-copulatory sex (e.g., oral sex) or an emotional infidelity, we would expect their partners to deny having an EPC. However, of the 50 men and women who were highly certain that their partners had affairs (≥90%), all their partners (100%) admitted to having EPCs (see Tables 1 and 2). Conversely, of the 69 men and women who reported intermediate levels of certainty that their partners had affairs (20–80%), only 18

[3] Of course, subjects might have had a different view of whether their own extrapair sexual behavior was illicit, but they were only asked to provide information on the affair behavior of their partners, not themselves.

(26%) of their partners admitted to having an EPC. This pattern suggests that if some subjects used information about emotional infidelity or non-copulatory sex to make affair inferences about their partners, they were most likely to result in weak inferences, not strong ones. This, in turn, would suggest that emotional infidelity is not perceived to be direct evidence of an affair, whereas an EPC is strong evidence of an affair. We therefore interpret the pattern in the following way. If some subjects did use evidence of an emotional infidelity (or non-copulatory sex) to make affair inferences, they did so because they thought it more likely that their partner had had an EPC. This would be consistent with the evidence from our survey of dictionaries that "affair" connotes sexual relations, and with evidence that when one's partner develops strong feelings of caring, attraction, or affection for another person of the opposite sex, it is considered a risk factor for sexual infidelity (Harris and Christenfeld 1996; Shackelford and Buss 1997).

If men's greater inferential accuracy is not an artifact of the methodology, then it is possible that men were better at processing information and drawing inferences from subtle clues. There is some evidence that women have better theory of mind skills than men (Geary 1998), which would seem to argue against this possibility. In any event, we have no ability to test this hypothesis in our sample.

Another possibility is that men are more accurate because, in heterosexual relationships, women freely reveal more information about their affairs and EPCs to their partners without prompting. This possibility is at odds with our hypothesis that women, more than men, have been under stronger selection to conceal their affairs and EPCs. Indeed, as we discuss in more detail below, we found evidence that the underreporting rates for women were higher than those for men. Thus, while it is possible that men are more accurate because women freely communicate more EPC and affair information to their partners, we think it unlikely. Still, this possibility, and others, could be explored more rigorously in future work.

A related possibility is that men are more accurate because they confront and rigorously question their partners when they become suspicious, and women might be more likely to reveal EPC and affair information under those circumstances. More generally, the costs that derive from paternity uncertainty may select for men who are highly motivated to resolve any uncertainties they have about the EPC behavior of their partners. Men who are suspicious that their mates have had EPCs may be more motivated to search for clues that confirm or disconfirm their suspicions and reduce their uncertainty, which may include confronting and interrogating their partners. People who harbor suspicions about their partners should be more likely to find confirming evidence when their partners have actually had EPCs, provided they are motivated to seek such information. This predicts that, of subjects with partners who report having EPCs, men will be less likely than women to harbor unresolved suspicions about their partners because suspicious men were more motivated to seek out confirming information, which they found. Conversely, of subjects with partners who report not having EPCs, men should be less certain about the EPC behavior of their partners because any suspicions they have will tend to go unresolved for lack of confirming evidence. Statistically, the certainty of people's inferences about the EPC behavior of their partners should depend on an interaction between the sex of the person making the inference and whether or not the partner has had an EPC.

From Table 1, men with unfaithful mates were either highly certain that their partners had not had an affair (≤10%) or highly certain that they had had an affair (≥99%). Conversely, women with unfaithful mates were more likely to endorse greater uncertainty about their mates' EPC behavior. This pattern is generally in line with our prediction. To formally test it, we assumed that maximum uncertainty occurred when people perceived that their partners were equally likely to be faithful as unfaithful (50% chance of EPC). We then calculated the certainty as twice the absolute value of the deviation between the rated chance and 50%.

$$\text{Inferential certainty} = 2\times|\text{chance of EPC} - 50\%|$$

This formula puts certainty on a scale from 0% to 100% and treats an inference of 0% as having the same degree of certainty as an inference of 100% (i.e., both inferences are absolutely certain, but differ in valence). For men, the average inferential certainty was 80% if their partners did not report an EPC and 98% if their partners did report an EPC. For women, the average inferential certainty was 86% if they had faithful partners and 82% if they had unfaithful partners. This pattern is suggestive of the predicted interaction, as can be seen in Fig. 4.

We used PROC GLIMMIX in SAS to test for significance, which allowed us to deal with the non-independence of men's and women's data. Certainty was reverse-coded to approximate the gamma family of distributions. Certainty was the dependent variable, and the independent variables were the subject's sex (0=male, 1=female), the partner's self-reported EPC behavior (0=faithful, 1=unfaithful), and

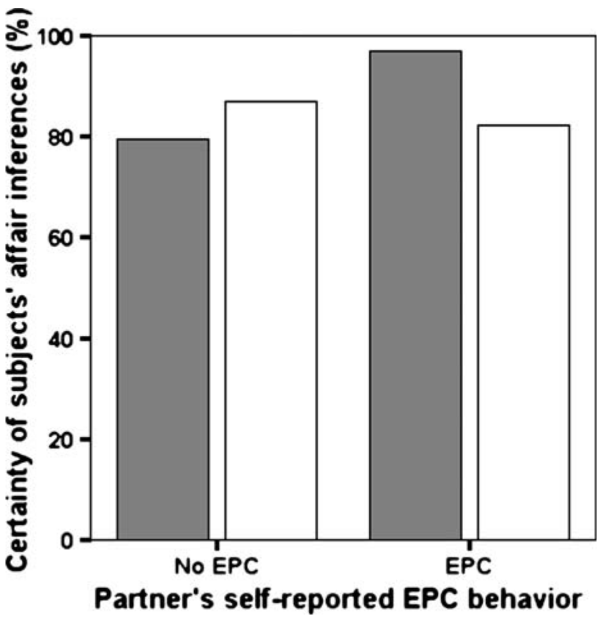

Fig. 4 The certainty of men's and women's EPC inferences as a function of their partners' self-reported EPC behavior. The gray bars represent how certain men are about the inferences they have made about their partner's extrapair behavior. The white bars represent how certain women are about the inferences they have made about their partner's extrapair behavior

the interaction between the two. As predicted, the interaction term was significant (β=+1.03, SE=0.29, p=0.001). The certainty of men's EPC inferences varied with their partners' EPC self-report, but women's certainty did not. When we dropped the interaction term, the subject's sex was not a significant predictor of certainty (β=−0.06, SE=0.10, p = n.s.), nor was the partner's self-reported EPC status (β=−0.14, SE= 0.13, p = n.s.).

In summary, among subjects with partners who have had EPCs, men (but not women) do not report intermediate probabilities that their partners have had affairs. We interpret this as evidence supporting the hypothesis that men have greater motivation to resolve uncertainties about their partners' EPC behavior. This motivation makes them more likely to find confirming evidence that resolves their uncertainty when their partners have actually had EPCs, which may contribute to their greater inferential accuracy. We also interpret this as supporting our contention that the 100% certainty that some subjects reported about their partners having affairs was the product of the EPC detection mechanisms that we are attempting to study, which was why we included them in our analyses.

The Contingencies Contributing to Men's Greater Accuracy

Table 3 shows the contingencies between men's and women's inferences and their partners' reported EPC behavior. Among those who tended to believe that their partners had not had an EPC, men were more likely to be correct than women (94% vs. 80%). Also, of participants with partners who had an EPC, men were more likely to detect the EPC (75% vs. 41%). These contingencies may reflect the degree to which sex-differentiated selection pressures influenced the design of EPC detection mechanisms. Relative to women, it may have been more important for men to be correct when they made inferences of fidelity and to detect EPC when their partners had been unfaithful. Enhanced accuracy with respect to these contingencies would tend to minimize the risk of cuckoldry, because they minimize the number of undetected EPCs.

In contrast, the other two contingencies did not contribute to men's accuracy. Of subjects with partners who reported not having an EPC, men were slightly less accurate than women (93.7% vs. 95.1%). Also, of those who thought that their partners had EPCs, men were less accurate than women (69.2% vs. 80.0%). It is noteworthy that if men had been more accurate in these two contingencies, it would not have reduced their risk of cuckoldry, because men could still have failed to detect more EPCs than women.

Do Women Underreport More Than Men?

We have hypothesized that women have been under stronger selection than men to conceal their EPC behavior. If so, then they might be more likely to underreport than men. Using a closely related goodness-of-fit procedure that we will describe in another paper, we were also able to directly estimate the overall level of underreporting by men and women in this sample. We do not have the capacity to estimate two parameters per sex (i.e., p, q, r, and s cannot all be estimated), but we can estimate underreporting if we assume that underreporting is constant within each sex. Consistent with our prediction, the best-fitting model estimated the proportions of men and women who had an EPC but failed to disclose it as 0% and 10%, respectively. However, this estimate should be treated with caution. The model had difficulty settling on a best-fitting estimate because of the zero frequency counts in some of the cells in the observed data (see Table 1). Problems with cells having low frequency counts could be solved with larger sample sizes. Also, these are merely the estimates that are most consistent with the overall pattern of observed results. The actual rates of underreporting could still vary. Because these estimates assume that underreporting is constant within each sex, they are not enough to move the population out of the region in which there is a global sex difference and a sex difference in accuracy (see above).

Limitations

We have already discussed how the EPC self-report and the EPC inference questions were different from each other. There are several other important limitations. First, couples knew that they would be participating in a study that would inquire broadly into their sexual behavior. Some people probably did not participate in the study because of its content, and it is possible that this could have influenced our results.

Second, although we assumed that subjects' underreporting may have depended on what their partners said about them, it may also have depended on what they suspected about their partners. For instance, a subject might be more likely to reveal her own EPC behavior if she thought that her partner had had an affair. To make our underreporting analyses tractable, we did not take this into account, and, depending on how it affects accuracy and error bias, it could change some of our conclusions.

Third, people may have been reluctant to disclose their true suspicions about their partners. We did not model this kind of underreporting, largely to make our analyses tractable. Even so, underreporting of EPC suspicions may have been limited. Because we asked each member of the couple both the EPC self-report question and the EPC inference question, they may have had an incentive to divulge their true suspicions so as to not look like dupes if their partners answered the EPC self-report question in the affirmative. Also, it is possible that some people might have found the opportunity to disclose concealed worries about their partners in an anonymous setting to be emotionally cathartic.

Fourth, our analyses presume that couples did not assortatively pair for underreporting of EPC behavior. In principle, a variable representing the degree of assortative pairing can be incorporated into the transition matrix, and its effects explored. Even if the attraction between men and women is highly correlated with underreporting, the effects on sex differences in accuracy or error bias are likely to be small in situations where the base rate of underreporting is low in at least one of the sexes. For instance, the best-fitting estimate of men's underreporting is less than that for women (0% vs. 10%). Under such circumstances, the net level of assortative pairing for underreporting will be small, and its effects minimal.

Fifth, in some situations, people might falsely inflate their own EPC behavior. In other words, they might overreport rather than underreport their own infidelities.

This may be most likely for men who could potentially get some reputational advantage out of trying to inflate their sexual desirability. The anonymous setting was designed to minimize incentives to do so, but it is a possible confound.

Finally, our sample was composed mostly of dating couples around 20 years of age from an Anglo-Hispanic New Mexico university population. People in other relationship stages or from different populations may exhibit different accuracy or bias patterns. Being accurate about partner EPC behavior is probably most important to men during late adolescence and young adulthood. Sexual experimentation is frequent in adolescence and young adulthood, relationships are relatively unstable (Weisfeld 1999), and sexual infidelity is not uncommon (Feldman and Cauffman 1999). Moreover, since female mate value is highly correlated with fertility (Buss 1994), female mate value should be high at this time, which may make it more important for their male partners to monopolize female sexuality and be vigilant for EPC behavior. Also, as men leave young adulthood, their mate value will rise as they acquire status and resources (Buss 1994), and women may be less likely to be sexually unfaithful to them. The patterns might be very different in older couples where relative differences in mate value have changed. Finally, couples who have been together for a while may have successfully negotiated much of their conflicting interests and transitioned to a phase where they exhibit greater trust. Under such circumstances, sex differences in EPC detection mechanisms may be minimized. We therefore predict that a male-biased sex difference in the accuracy of inferences about the EPC behavior of mates is most likely to occur in late adolescence and young adulthood and in less established couples. Further research is needed to address these phenomena in additional populations.

Conclusion

In our observed data, there were significant sex differences in accuracy and in error bias. Men made more accurate EPC inferences than did women, and they made more false positive errors (i.e., men were more likely to infer EPC when their partners said that they had been faithful). We also found evidence suggesting that men may be more motivated to seek out information that resolves suspicions about the sexual fidelity of their partners, and this may contribute to their greater accuracy. Finally, we also found evidence suggesting that women may have underreported at higher rates than men, supporting the hypothesis that women have been under stronger selection to conceal their EPC behavior. The issue of whether EPC detection and EPC concealment mechanisms have been designed by selection in sex-differentiated ways will not be conclusively resolved with a single study. Nevertheless, although there are important limitations in the current study that should be addressed in future work, this study provides preliminary evidence that such differences may exist.

Acknowledgments PWA was supported by a National Research Service Award from the National Institutes of Health, P32 MH-20030 (PI: Michael C. Neale). Rosalind Arden, Judith Easton, Todd Shackelford, Andy Thomson, Tina Wagers, and two anonymous reviews provided comments. Chuck Gardner provided statistical advice.

References

Anderson, K. G., Kaplan, H., & Lancaster, J. (1999). Paternal care by genetic fathers and stepfathers I: Reports from Albuquerque men. *Evolution and Human Behavior*, *20*, 405–431.

Buss, D. M. (1994). *The evolution of desire: Strategies of human mating*. New York: Basic Books.

Buss, D. M. (2000). *The dangerous passion: Why jealousy is as necessary as love and sex*. New York: Free Press.

Buss, D. M., & Shackelford, T. K. (1997). From vigilance to violence: Mate retention tactics in married couples. *Journal of Personality and Social Psychology*, *72*, 346–361.

Feldman, S. S., & Cauffman, E. (1999). Sexual betrayal among late adolescents: Perspectives of the perpetrator and the aggrieved. *Journal of Youth and Adolescence*, *28*, 235–258.

Fenigstein, A., & Preston, M. (2007). The desired number of sexual partners as a function of gender, sexual risks, and the meaning of "ideal". *Journal of Sex Research*, *44*, 89–95.

Gangestad, S. W., & Thornhill, R. (1997). The evolutionary psychology of extrapair sex: The role of fluctuating asymmetry. *Evolution and Human Behavior*, *18*, 69–88.

Gangestad, S. W., Thornhill, R., & Garver, C. E. (2002). Changes in women's sexual interests and their partners' mate-retention tactics across the menstrual cycle: Evidence for shifting conflicts of interest. *Proceedings of the Royal Society of London, Series B: Biological Sciences*, *269*, 975–982.

Gangestad, S. W., Thornhill, R., & Garver, C. E. (2005). Adaptations to ovulation. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 344–371). Hoboken, NJ: Wiley.

Geary, D. C. (1998). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.

Hansen, G. L. (1985). Dating jealousy among college students. *Sex Roles*, *12*, 713–721.

Harris, C. R., & Christenfeld, N. (1996). Jealousy and rational responses to infidelity across gender and culture. *Psychological Science*, *7*, 378–379.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: A new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology*, *78*, 81–91.

Haselton, M. G., & Gangestad, S. W. (2006). Conditional expression of women's desires and men's mate guarding across the ovulatory cycle. *Hormones and Behavior*, *49*, 509–518.

Lenoir, C. D., Adler, N. E., Borzekowski, D. L. G., Tschann, J. M., & Ellen, J. M. (2006). What you don't know can hurt you: Perceptions of sex-partner concurrency and partner-reported behavior. *Journal of Adolescent Health*, *38*, 179–185.

McBurney, D. H., Zapp, D. J., & Streeter, S. A. (2005). Preferred number of sexual partners: Tails of distributions and tales of mating systems. *Evolution and Human Behavior*, *26*, 271–278.

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. H. (2003). *Mx: Statistical modeling* (6th ed.). Richmond: Department of Psychiatry, Virginia Commonwealth University.

Schaeffer, N. C. (2000). Asking questions about threatening topics: A selective overview. In N. Schwarz, & S. Sudman (Eds.), *Autobiographical memory and the validity of retrospective reports* (pp. 141–160). New York: Springer.

Schmitt, D. P. (2003). Universal sex differences in the desire for sexual variety: Tests from 52 nations, 6 continents, and 13 islands. *Journal of Personality and Social Psychology*, *85*, 85–104.

Shackelford, T. K., & Buss, D. M. (1997). Cues to infidelity. *Personality and Social Psychology Bulletin*, *23*, 1034–1045.

Shackelford, T. K., Buss, D. M., & Weekes-Shackelford, V. A. (2003). Wife killings committed in the context of a lovers' triangle. *Basic and Applied Social Psychology*, *25*, 137–143.

Tourangeau, R., & Smith, T. W. (1996). Asking sensitive questions: The impact of data collection mode, question format, and question context. *Public Opinion Quarterly*, *60*, 275–304.

Turner, C. F., Forsyth, B. H., O'Reilly, J., Cooley, P. C., Smith, T. K., Rogers, S. M., et al. (1998). Automated self-interviewing and the survey measurement of sensitive behaviors. In M. P. Couper, R. P. Baker, J. Bethlehem, C. Z. F. Clark, J. Martin, W. L. Nicholls, & J. M. O'Reilly (Eds.), *Computer-assisted survey information collection* (pp. 455–473). New York: Wiley.

Weisfeld, G. E. (1999). *Evolutionary principles of human adolescence*. New York: Basic Books.

Wilson, M., Daly, M., & Daniele, A. (1995). Familicide: the killing of spouse and children. *Aggressive Behavior*, *21*, 275–291.

Yarab, P. E., Allgeier, E. R., & Sensibaugh, C. C. (1999). Looking deeper: Extradyadic behaviors, jealousy, and perceived unfaithfulness in hypothetical dating relationships. *Personal Relationships*, *6*, 305–316.

Paul W. Andrews received his PhD in biology at the University of New Mexico in 2002. He is a postdoctoral research fellow at the Virginia Institute for Psychiatric and Behavioral Genetics at Virginia Commonwealth University. His current research interests include understanding the evolution of mental health traits, especially depression and suicidal behavior, developing evidentiary criteria for distinguishing psychological adaptations for coping with stressors from mental disorders, and developing statistical methods for dealing with underreporting.

Steven W. Gangestad received his PhD in psychology at the University of Minnesota in 1985. He is Distinguished Professor of Psychology at the University of New Mexico. He is also currently President of the Human Behavior and Evolution Society. His recent research and interests have focused on adaptations and by-products involved in human mating and sexuality, individual differences in developmental instability, and evidentiary standards for distinguishing adaptation from other evolved outcomes.

Geoffrey Miller is an evolutionary psychologist at the University of New Mexico, with a PhD from Stanford University in 1993. His research concerns human mate choice, fitness indicators, evolutionary behavior genetics, intelligence, personality, psychopathology, and consumer behavior.

Martie Haselton is associate professor of communication studies and psychology at UCLA, with a PhD from the University of Texas at Austin in 2000. Her research concerns female sexuality, ovulatory cycle shifts in women's mate preferences and social behaviors, adaptive biases in social judgment, strategic conflict between the sexes, and adaptationist theory.

Randy Thornhill received his PhD in biology at the University of Michigan in 1974. He is Distinguished Professor of Biology at the University of New Mexico. His current research interests include sexual selection processes, the evolution of human sexuality, and the role of infectious diseases in cultural diversity.

Michael Neale obtained his PhD in Psychology from the University of London in 1985 and is a professor of psychiatry and human genetics at the Virginia Institute for Psychiatric and Behavioral Genetics at Virginia Commonwealth University. He specializes in the development of statistical models and software for fitting them to genetically informative datasets. Substantive areas of interest include psychopathology, substance use, cognition, and magnetic resonance imaging.