
Sex differences in jealousy: Misinterpretation of nonsignificant results as refuting the theory

JOHN E. EDLUND AND BRAD J. SAGARIN
Northern Illinois University

Abstract

Studies examining sex differences in jealousy using continuous measures have produced inconsistent findings. To explain these inconsistencies, this study critically evaluates the criterion used to test the sex difference in jealousy, demonstrating that the Participant Sex \times Infidelity Type interaction is the only relevant effect. Then, using a U.S.-based sample, the sex difference is investigated using 14 continuous-measure response formats, revealing a highly significant overall sex difference ($p < .001$, $g = .300$), despite the sex difference being nonsignificant for 9 of 14 individual formats. Results highlight the danger of falling prey to the belief in the law of small numbers, manifest in this debate as the erroneous interpretation of individual nonsignificant results as refuting the theory.

According to the theory of evolved sex differences in jealousy (Buss, Larsen, Westen, & Semmelroth, 1992; Daly, Wilson, & Weghorst, 1982; Symons, 1979), ancestral women's challenge of ensuring paternal investment exerted selective pressures that increased women's jealousy in response to emotional infidelity, whereas ancestral men's challenge of paternal uncertainty exerted selective pressures that increased men's jealousy in response to sexual infidelity. The majority of studies examining sex differences in jealousy have used the forced-choice method (Buss et al., 1992) in which women and men are asked to choose which type of infidelity (sexual or emotional) would cause

greater distress. Researchers have favored this methodology for two reasons: (a) both men and women tend to have strong jealous reactions to sexual and emotional infidelity because both forms of infidelity were costly and they were likely to have been highly correlated with one another throughout evolutionary history, and (b) the forced-choice method allows for a cleaner separation of men and women's differential reactions to both types of infidelity. Continuous measures, in contrast, allow participants to respond identically to both types of infidelity, potentially making the sex difference more difficult to detect. Results of a meta-analysis demonstrate that the sex difference in jealousy when assessed via the forced-choice method is reliable and robust (Harris, 2003).

Some critics have argued that the sex difference in jealousy is an artifact of the forced-choice method and that the results do not replicate with continuous measures (DeSteno, Bartlett, Braverman, & Salovey, 2002; Harris, 2003). Indeed, studies that have assessed women and men's jealousy in response to both sexual and emotional infidelity using continuous measures have produced conflicting results. Nonetheless, as we argue in this

John E. Edlund and Brad J. Sagarin, Department of Psychology, Northern Illinois University.

John E. Edlund is now at the Department of Psychology, Hamilton College.

Some of these findings were initially presented at the 2007 Midwestern Psychological Association's annual conference. We are gratefully indebted to George Neuman and our anonymous reviewers for feedback on earlier drafts of the manuscript. Finally, we wish to thank our research assistants Nikki Accurso, Megan Cannon, and Karen Raff.

Correspondence should be addressed to John E. Edlund, Department of Psychology, Hamilton College, Clinton, NY 13323, e-mail: jedlund@hamilton.edu.

article, these conflicts may well be illusory, stemming from (a) misunderstandings regarding the empirical implications of the theory and (b) misinterpretations of the distributions of significant and nonsignificant effects. In the next section, we consider the typical continuous measure study in which women and men respond to sexual and emotional infidelity, and we provide a theoretical argument for the theory relevance of the Participant Sex \times Infidelity Type interaction and the theory irrelevance of the often-interpreted main effects and simple effects. Thereafter, we provide empirical evidence of a distribution of interaction effects that strongly implies a reliable sex difference in the population despite a preponderance of nonsignificant interactions.

Which effect should be interpreted?

Much confusion has arisen regarding which effects are relevant for testing the theory of evolved sex differences in jealousy. Most studies examining sex differences in jealousy using continuous measures employ a 2 (participant sex) \times 2 (infidelity type) mixed model factorial design with participant sex as the between-subjects factor and infidelity type as the within-subject factor (e.g., DeSteno et al., 2002; Edlund, Heider, Scherer, Farc, & Sagarin, 2006; see, however, Michalski, Shackelford, & Salmon, 2007; Shackelford, Michalski, & Schmitt, 2004; Sheets & Wolfe, 2001, for a different approach in which the forced-choice format is adapted to yield continuous-measure data). Each participant provides two measures: jealousy (or some other emotion) in response to sexual infidelity and jealousy (or some other emotion) in response to emotional infidelity. A 2 \times 2 analysis of variance (ANOVA) performed on such data produces a main effect of participant sex, a main effect of infidelity type, and a Participant Sex \times Infidelity Type interaction, with a significant interaction often followed by four simple effects: the simple effects of participant sex in response to sexual infidelity and in response to emotional infidelity, and the simple effects of infidelity type for women and for men. In this section, we argue that the only relevant test is the Participant Sex \times Infidelity

Type interaction. At first glance, the simple effects might appear relevant, but as we demonstrate later, the relevance of the simple effects rests on the unwarranted assumption that no other factors impact participants' responses to the jealousy measures.

The theory of evolved sex differences in jealousy posits that selection pressures selected for men who felt more jealousy in response to sexual infidelity over men who felt less jealousy in response to sexual infidelity because of the decreased likelihood that such men would be cuckolded compared to their less jealous intrasexual competitors. Likewise, selection pressures selected for women who felt more jealousy in response to emotional infidelity over women who felt less jealousy in response to emotional infidelity because of the decreased likelihood that such women would lose paternal investment in their children compared to their less jealous intrasexual competitors. Put another way, selection pressures gave men (but not women) a boost in their jealous response to sexual infidelity. Likewise, selection pressures gave women (but not men) a boost in their jealous response to emotional infidelity.

To illustrate the empirical implications of this theory, consider a study in which women and men indicate how jealous they would feel on a 1 to 10 scale in response to a sexual infidelity and in response to an emotional infidelity. Figure 1 illustrates the means in each condition assuming (a) a baseline response to 5 on the scale and (b) the evolved sex difference manifests as a 1-point increase in men's jealousy in response to sexual infidelity and women's jealousy in response to emotional infidelity. (The specific values are arbitrary, of course; alternative values would produce an analogous pattern.) This pattern of results shows an interaction in the theory-supportive direction—men, relative to women, report greater jealousy in response to sexual infidelity than in response to emotional infidelity. In addition, all four simple effects show patterns seemingly supportive of the theory: (a) men report greater jealousy in response to sexual infidelity than in response to emotional infidelity, (b) women report greater jealousy in response to emotional infidelity than in

	Men	Women	
Sexual infidelity	6 (5 baseline +1 evolved sex difference)	5 (5 baseline)	5.5
Emotional infidelity	5 (5 baseline)	6 (5 baseline +1 evolved sex difference)	5.5
	5.5	5.5	

Figure 1. Pattern of jealousy responses influenced only by evolved sex differences.

response to sexual infidelity, (c) men report greater jealousy than women report in response to sexual infidelity, and (d) women report greater jealousy than men report in response to emotional infidelity.

This example illustrates that if sexually dimorphic selection pressures were the only factor impacting responses, the simple effects would indeed be relevant for testing the theory. Past research, however, has identified a variety of factors likely to impact the overall levels of jealousy women versus men report or the overall levels of jealousy reported in response to sexual versus emotional infidelity. With respect to the overall levels of jealousy women versus men report, for example, Feldman Barrett, Robin, Pietromonaco, and Eysell (1998) found that women report more intense emotions than men on global, retrospective measures, and Sagarin and Guadagno (2004) found that women report higher levels of jealousy than men on one commonly used response scale because of a differential interpretation of the upper anchor of the scale. Alternatively, if researchers operationalize jealousy as a willingness to commit overt acts of aggression toward the rival, one might expect men to exhibit higher levels of jealousy than women.

With respect to the overall level of jealousy reported in response to sexual versus emotional infidelity, the operationalization of sexual versus emotional infidelity within a particular study is likely to have an impact. Buss and colleagues (1992), for example, used two operationalizations of each type of infidelity. They operationalized sexual infidelity as either “enjoying passionate sexual intercourse” or “trying different sexual positions” and emotional infidelity as either “forming a deep emotional attachment” or “falling in

love.” The results suggest that both women and men found passionate sexual intercourse more jealousy provoking than different sexual positions, and falling in love more jealousy provoking than a deep emotional attachment.

To see the implications of these additional factors, consider a study in which “passionate sexual intercourse” represents sexual infidelity, “a deep emotional attachment” represents emotional infidelity, and the response scale uses *extremely jealous* as the upper anchor. Consistent with Buss and colleagues (1992), the operationalizations of sexual versus emotional infidelity might cause an increase in women and men’s response to the sexual infidelity compared to the emotional infidelity. Consistent with Sagarin and Guadagno (2004), the choice of the upper anchor might cause an increase in women’s responses to both types of infidelities compared to men. Figure 2 illustrates the means in each condition assuming (a) a baseline response to 5 on the scale, (b) the evolved sex difference manifests as a 1-point increase in men’s jealousy in response to sexual infidelity and women’s jealousy in response to emotional infidelity, (c) the operationalization of the infidelities manifests as a 2-point increase in women and men’s jealousy in response to sexual infidelity, and (d) the upper anchor manifests as a 2-point increase in women’s jealousy in response to both types of infidelity.¹

1. It is important to note that an increase in jealousy (to both types of infidelity) as a result of a modified upper anchor would not affect the magnitude or meaning of the interaction because the effect of the upper anchor would be partialled out as part of the main effect of participant sex. Consistent with this, Sagarin and Guadagno (2004) found a significant Participant Sex × Infidelity Type interaction but no three-way Participant Sex × Infidelity Type × Upper Anchor interaction.

	Men	Women	
Sexual infidelity	8 (5 baseline +2 infidelity operationalization +1 evolved sex difference)	9 (5 baseline +2 infidelity operationalization +2 upper anchor)	8.5
Emotional infidelity	5 (5 baseline)	8 (5 baseline +2 upper anchor +1 evolved sex difference)	6.5
	6.5	8.5	

Figure 2. Pattern of jealousy responses influenced by evolved sex differences and other factors.

The pattern of results shows an interaction in the theory-supportive direction—men, relative to women, report greater jealousy in response to sexual infidelity than in response to emotional infidelity. But two of the simple effects might seem at first glance to refute the theory—women report greater jealousy than men in response to sexual infidelity, and women report greater jealousy in response to sexual infidelity than in response to emotional infidelity. Both of these simple effects, however, stem from other factors impacting the scores, factors that are in no way incompatible with the theory of evolved sex differences in jealousy. In particular, because reproductive competitions were intrasexual (i.e., women competed with other women whereas men competed with other men for reproductive success), the theory does not imply that selection pressures boosted men's jealous response to sexual infidelity above women's jealous response to sexual infidelity, or that selection pressures boosted women's jealous response to emotional infidelity above men's jealous response to emotional infidelity. Thus, the theory is compatible with a variety of factors that might impact the overall levels of jealousy women versus men report or that might impact the overall levels of jealousy reported in response to sexual versus emotional infidelity.²

2. We should note that a significant interaction in the theory-supportive direction represents evidence that there is a sex difference, but it leaves ambiguous whether the sex difference stems from a boost in men's jealousy in response to sexual infidelity, a boost in women's jealousy in response to emotional infidelity, or both. In addition, for the reasons stated previously, the simple effects do not represent separate tests of each of these because factors impacting the main effect of participant sex and the main effect of infidelity type also influence the simple effects. Instead, separate tests of each of these requires a different type of comparison, such as Sagarin, Becker, Guadagno, Nicastle, and Millevoi's (2003) examination of same-sex and opposite-sex infidelities.

The conclusions of one recent study that examined sex differences in jealousy using continuous measures (Sabini & Silver, 2005, Study 2) highlight the risk of focusing on theory-irrelevant effects. Based on the simple effects of infidelity type for women and men, the authors conclude, "There is little support for the evolutionary position in these data. With regard to all three dependent variables, the two sex conditions had a greater impact than did the emotion conditions for both sexes" (p. 723). Nonetheless, as Sabini and Silver (2005) report, both measured emotions showed theory-supportive interactions, one marginal ($p = .051$ for anger) and one significant ($p = .01$ for hurt).

The logic applied in this analysis of continuous measures also applies to a misunderstanding of the forced-choice method. For example, Buller (2005) characterized the theory as implying that "males care more about sexual infidelity *than they do about emotional infidelity*" (p. 280, italics in original). This is analogous, however, to the erroneous interpretation of the simple effect of infidelity type described earlier. Applying our argument to the analysis for forced-choice data, the relevant test is whether more men than women choose sexual infidelity as more jealousy provoking than emotional infidelity. The test of whether the proportion of men or women differs from 50% is irrelevant (and possibly misleading).

Finally, in response to the concern that evolutionary psychological theories are unfalsifiable, it is worth noting that the interaction test can, in fact, produce a pattern of results that would refute the theory of evolved sex differences in jealousy. Specifically, a significant interaction in which women, relative to men, reported greater jealousy in response to sexual

infidelity than in response to emotional infidelity would provide strong evidence against the theory. (For more information on the falsifiability of evolutionary psychological theories, see Conway & Schaller, 2002; Ketelaar & Ellis, 2000.)

The present study

Although a focus on the theory-relevant versus theory-irrelevant effects helps resolve some of the discrepancies in the published literature (e.g., Sabini & Silver, 2005), substantial inconsistencies remain, with some studies finding that the sex difference in jealousy does not replicate with continuous measures (e.g., DeSteno et al., 2002; DeSteno & Salovey, 1996; Green & Sabini, 2006; Harris, 2002; Sabini & Green, 2004) but other studies finding that the sex difference in jealousy replicates when using continuous measures (e.g., Edlund et al., 2006; Sagarin et al., 2003; Wiederman & Allgeier, 1993).³ Additionally, other researchers (Fenigstein & Peltz, 2002; Michalski et al., 2007; Shackelford et al., 2004; Sheets & Wolfe, 2001) have used hybrid techniques with results that support the evolutionary psychologically derived predictions.

3. The question of whether the sex difference replicates using continuous measures is not, of course, the only challenge to the theory of evolved sex differences in jealousy. Harris (2002) criticized the use of hypothetical infidelity scenarios and questioned whether the sex difference would appear in response to actual infidelity experiences. In response, Edlund and colleagues (2006) demonstrated that the sex difference in jealousy replicated when an adult sample retrospectively reported their responses to a past experience as a victim of infidelity. DeSteno and colleagues (2002) offered another challenge, suggesting that cognitive load eliminated the sex difference in jealousy. Sagarin (2005), however, reanalyzed DeSteno and colleagues' data and found that a significant sex difference remained even when participants were under cognitive load. Furthermore, Barrett, Frederick, Haselton, and Kurzban (2006) and Sagarin question the usefulness of cognitive load in investigating the sex difference in jealousy (but see DeSteno, Bartlett, & Salovey, 2006, for a response to Barrett et al., 2006). Finally, Harris (2000) challenged the theory-supportive interpretation of Buss and colleagues' (1992) physiological data. Harris's (2000) challenge remains unaddressed (but see Takahashi et al., 2006, for an intriguing demonstration of different patterns of brain activation when women and men imagine sexual and emotional infidelity).

In an attempt to resolve this conflicting set of results, we examined each study using continuous scales discussed in the preceding paragraph to try to identify reasons why some found a sex difference and others did not. One set of potential moderators that emerged represented characteristics of the response scale provided to participants. In particular, three characteristics tended to differentiate studies that failed to find a sex difference from studies that found a sex difference: number of points labeled on the scale, scale length, and the upper anchor used. Researchers who have not shown support for the sex difference in jealousy typically have used compact scales (e.g., 5-point; Harris, 2002) whereas researchers who have found support for the sex difference in jealousy have often used extended scales (e.g., 10-point; Sagarin et al., 2003). Research (Weems, 2004) has suggested that using longer scales increases the reliability of the scales and decreases the use of the midpoint. This additional reliability could contribute to the ability (or lack thereof) to detect the sex difference in jealousy.

Similarly, researchers who have not shown support for the sex difference typically have labeled more than 2 points on the scale, whereas researchers who have found support for the sex difference have often labeled only endpoints. For example, Sabini and Green (2004) labeled both endpoints and a midpoint and failed to find the sex difference in jealousy, whereas Edlund and colleagues (2006) labeled only the endpoints and found the sex difference. Many psychometricians have argued that a labeled midpoint can adversely affect the scale by drawing participants to the midpoint of the scale (Alreck & Settle, 1985) as well as decreasing the reliability of the measure (Alwin & Krosnick, 1991). Additionally, other researchers (Swearingen, 1999) have found significant mean differences between scales with a midpoint and scales without. This suggests the inclusion of a midpoint can impact responses to a scale.

Finally, researchers who have not shown support for the sex difference have often used *very* as the upper anchor label (e.g., Green & Sabini, 2006), whereas researchers who have shown support for the sex difference have often

used *extremely* as the upper anchor label (e.g., Wiederman & Allgeier, 1993). Assuming that *extremely* represents an amount of jealousy greater than *very*, the use of *very* as the upper anchor might compress the usable portion of the scale more so than the use of *extremely*.

It is notable that two of these characteristics seem to go against psychometricians' typical recommendations. For example, Krosnick (1999) suggests using 5 response points to maximize the reliability of unipolar scales and labeling all points on the scale. In the context of extreme emotions, however, some of the usual guidelines for scale construction might not apply. In the case of emotional response to an infidelity, most participants would probably feel a fairly high degree of jealousy. As a result, participants might not consider the points on the lower end of the scale valid responses, and they might need longer scales to allow for sufficient variability in responses (Newstead, Pollard, & Riezebos, 1987). Similarly, if the scale midpoint includes a label denoting moderate jealousy, this might restrict the usable range of the scale to the upper end (Weems & Onwuegbuzie, 2001).

To test the effects of these potential moderators, we conducted an experiment that assessed jealousy in response to sexual and emotional infidelity while manipulating the characteristics of the response scale provided to participants. We then analyzed the resulting data using ANOVA and meta-analytic techniques (Biostat, 2006). The meta-analysis also provided an opportunity to examine the distribution of effects and the overall effect that emerged across the studies.

Method

Description of the participants

One thousand eighty-two students from a mid-sized Midwestern U.S. university participated in this experiment (480 men, 601 women, and 1 participant who did not specify a sex; $M_{\text{age}} = 18.95$ years, $SD = 1.75$) for partial fulfillment of a class requirement. There were 697 Caucasians, 183 African Americans, 84 Hispanics, 80 Asian Americans, and 38 self-identified others, which are representative of

the student population. We excluded 2 students who did not answer all of the relevant questions from the analyses. Our rationale for using a convenience sample is based on our need for an unusually large sample, as we had 14 conditions and obtaining a random sample of that size would have been prohibitively expensive.

Procedure

Participants received a survey that asked them the following:

Please think of a serious committed romantic relationship that you have had in the past, that you currently have, or that you would like to have. Imagine that you discover that the person with whom you've been seriously involved became interested in someone else. (scenario adapted from Buss et al., 1992)⁴

Participants then answered two questions: "How jealous would you feel about your partner forming a deep emotional attachment to this other person?" and "How jealous would you feel about your partner enjoying passionate sexual intercourse with this other person?"⁵ using a randomly assigned response format.

We created 14 variations of the response format. Thirteen of these forms were unipolar measures in which we varied the total number of points labeled on the scale (endpoints only, endpoints and midpoint, and all points labeled), the length of the scale (5, 7, 9, and 10 points), and the upper anchor label (*very jealous*,

4. This adaptation preserved Buss and colleagues' (1992) basic scenario with the response format changed from a single forced choice to two continuous measures (Edlund et al., 2006; Green & Sabini, 2006; Harris, 2002; Sabini & Silver, 2005).

5. Some might question the use of the word *jealous*, as past studies have used numerous other terms, including *upset/distressed*. Initially, researchers might have used *upset/distress* because of concerns that participants would be unwilling to report feelings of jealousy. Later research (e.g., Becker, Sagarin, Guadagno, Millevoi & Nicastle, 2004), however, demonstrated that participants are willing to report high levels of jealousy. As jealousy is the primary emotion of concern in this theory and past studies have used the term *jealous* successfully, we decided to use that term here.

extremely jealous).⁶ The 14th form used a feeling thermometer, which presents the response scale in a graphical format, similar to an old-fashioned thermometer. Although these 14 forms represent a number of permutations of the manipulated factors, they do not constitute a fully crossed design. Creating a fully crossed design would have included some invalid forms (e.g., a 10-point scale with a labeled midpoint) as well as some forms that proved difficult to construct (e.g., a fully labeled 10-point scale with 10 monotonically increasing gradations of jealousy of approximately equal intervals). As such, we proceeded with a nonfully crossed design constructed to match closely the published research on this topic.

Results

Individual forms

The first analysis examined the sex difference in jealousy in each response form individually (as if we had conducted 14 separate experiments). For brevity, we present the detailed ANOVA results for all forms in a summary table (see Table 1). Each form had a minimum of 76 participants, with a minimum of 25 men.⁷

6. We selected the options for each of these factors to represent the options that typically appeared in the published literature. With respect to the number of points labeled on the scale, our review of the literature revealed no studies that used a different labeling system than the three included in the present study. With respect to the length of the scale, our review revealed one study that used a different scale length (Takahashi et al., 2006, used a 6-point scale). With respect to the upper anchor label, most used *very* or *extremely*, a few used *incredibly* or *completely*, and one used *as jealous as you could feel in a romantic relationship*. According to Sagarin and Guadagno (2004), this last option has the benefit of eliminating a main effect of participant sex that appears with other upper anchor labels (i.e., women report more overall jealousy compared to men when *extremely* is the upper anchor, but this sex difference disappears when *as jealous as you could feel in a romantic relationship* is the upper anchor). Nevertheless, we opted to use *very* and *extremely* in the present study because (a) their more common use in the published literature and (b) the irrelevance of the main effect of participant sex in testing the theory-relevant interaction.

7. Per a reviewer's suggestion, we performed additional analyses with equal numbers of men and women by dropping random women, and the conclusions did not change. As such, we report only the results from the full data set.

Following the logic laid out in earlier in the manuscript, the appropriate analysis to examine the sex difference in jealousy is the interaction between participant sex and infidelity type. A statistically reliable interaction, in which men, relative to women, report greater jealousy in response to sexual infidelity compared to emotional infidelity would provide support for the theory.

Of the 14 response formats, 5 had statistically significant interactions in the theory-supportive direction, 1 had a marginal interaction in the theory-supportive direction, and 8 had statistically nonsignificant interactions, 5 with theory-supportive patterns of means and 3 with a theory-refutational pattern of means (although all of the theory-refutational patterns yielded interactions with $ps > .66$). For example, in Form 2 (5-point scale, endpoints only, and *extremely* as the upper anchor), both men and women reported greater jealousy in response to the sexual aspects (men: $M = 4.85$, $SD = .43$; women: $M = 4.56$, $SD = .77$) than to the emotional aspects (men: $M = 4.18$, $SD = .71$, women: $M = 4.28$, $SD = .82$) of the infidelity, but the difference was significantly greater for men, $F_{\text{interaction}}(1, 74) = 6.51$, $p = .013$, which represents a medium effect size⁸ ($g = .547$).

Combined analyses of overall effect

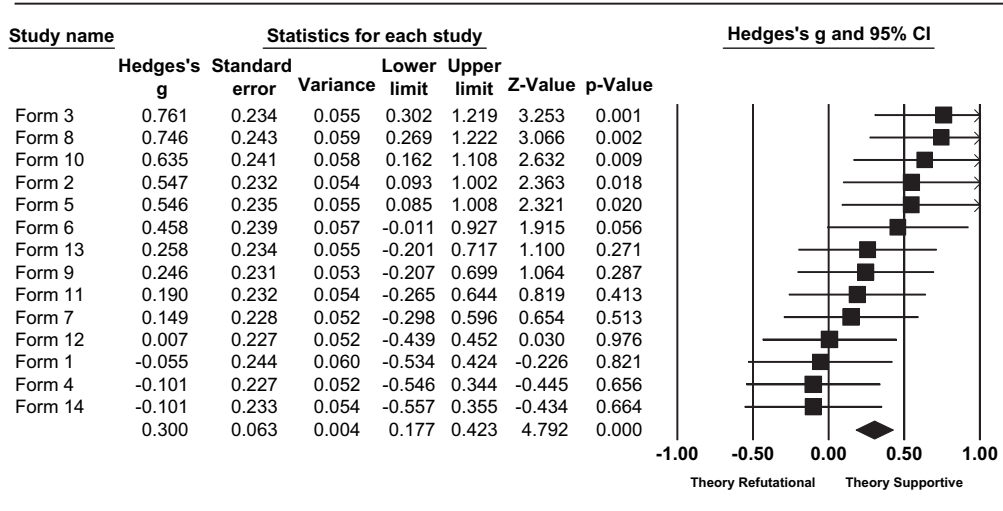
The next sets of analyses examined whether there was an overall sex difference in jealousy across all forms. To investigate this, we took two approaches. In the first approach, we used meta-analytic techniques and treated each form as a separate study. In the second approach, we converted the individual responses to z scores (standardized within forms) and then conducted an ANOVA on the standardized scores across the different response forms.

Using meta-analytic techniques, we converted each set of means, standard deviations, sample size, and correlations to a Hedges's g (see Figure 3). Positive Hedges's g s represent effects in which the means aligned in a theory-supportive pattern, whereas negative Hedges's g s represent effects in which the means aligned

8. Researchers can interpret Hedges's g using Cohen's rules of thumb for effect sizes (*small* = .2, *moderate* = .5, *large* = .8).

Table 1. Summary table of means, standard deviations, and ANOVA interactions for individual forms

Form	No. of points, point labeling, upper anchor	Men emotional infidelity		Men sexual infidelity		Women emotional infidelity		Women sexual infidelity		Interaction
		<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	<i>M</i> (<i>SD</i>)	
1	5, all points, <i>extremely</i>	3.71 (1.16)	4.33 (0.96)	4.10 (0.93)	4.77 (0.55)	$F(1, 74) = 0.05, p = .83$				
2	5, ends only, <i>extremely</i>	4.18 (0.71)	4.85 (0.43)	4.28 (0.82)	4.56 (0.77)	$F(1, 74) = 6.51, p = .01$				
3	5, middle & ends, <i>extremely</i>	4.08 (0.88)	4.89 (0.39)	4.41 (0.68)	4.67 (0.62)	$F(1, 75) = 14.50, p < .01$				
4	7, all points, <i>extremely</i>	5.64 (0.99)	6.42 (1.15)	5.89 (1.15)	6.78 (0.42)	$F(1, 76) = 0.18, p = .67$				
5	7, ends only, <i>extremely</i>	5.73 (0.98)	6.60 (0.62)	6.09 (0.80)	6.43 (1.12)	$F(1, 75) = 6.13, p = .02$				
6	7, ends only, <i>very</i>	5.39 (1.23)	6.50 (0.69)	6.06 (0.94)	6.72 (0.71)	$F(1, 73) = 3.60, p = .06$				
7	7, middle & ends, <i>extremely</i>	5.41 (1.61)	6.02 (1.85)	5.91 (0.85)	6.37 (0.97)	$F(1, 74) = 0.44, p = .51$				
8	7, middle & ends, <i>very</i>	6.00 (0.85)	6.79 (0.41)	6.46 (0.90)	6.57 (1.19)	$F(1, 69) = 14.70, p < .01$				
9	9, ends only, <i>extremely</i>	7.39 (1.73)	8.28 (1.63)	7.58 (1.57)	8.18 (1.66)	$F(1, 72) = 1.11, p = .30$				
10	9, ends only, <i>very</i>	7.31 (1.23)	8.62 (0.86)	7.96 (1.21)	8.53 (1.06)	$F(1, 72) = 7.47, p = .01$				
11	9, middle & ends, <i>extremely</i>	7.30 (1.26)	8.45 (1.20)	7.71 (1.33)	8.59 (1.20)	$F(1, 72) = 0.64, p = .43$				
12	9, middle & ends, <i>very</i>	7.19 (1.70)	8.08 (1.99)	7.33 (1.66)	8.21 (1.72)	$F(1, 74) = 0.00, p = .95$				
13	10, ends only, <i>extremely</i>	6.80 (1.83)	7.83 (2.41)	7.80 (1.66)	8.38 (1.50)	$F(1, 73) = 1.77, p = .19$				
14	Feeling thermometer	79.09 (19.55)	90.00 (20.59)	79.83 (21.89)	92.88 (15.64)	$F(1, 71) = 0.19, p = .67$				



Meta Analysis

Figure 3. Graphical depiction of all effect sizes.

in a theory-refutational pattern. We then meta-analyzed the *g* scores. The resulting mean Hedges's *g* using a fixed effect model was .300, $p < .001$. This represents a statistically significant effect in the theory-supportive direction with a small-moderate effect size. The 95% confidence interval for this effect is .177 to .423. A significant analysis of heterogeneity supported the hypothesis of heterogeneity, $Q(13) = 22.46, p = .049$, suggesting the presence of possible moderators to the effect sizes. The *z*-score data, collapsing across all response formats, also showed a statistically significant overall effect in the theory-supportive direction, $F_{interaction}(1, 1078) = 21.07, p < .001$.

Moderation analyses

For a number of reasons, the moderation analyses exclude the feeling thermometer. Past sex difference in jealousy studies have never used this response format, and it does not fit easily into the categories tested (e.g., number of points). Furthermore, the variance associated with this form was substantially greater than the variances in the other forms. As in the previous section, we took two approaches (meta-analytic, *z*-score ANOVA) in investigating whether scaling response format decisions affect the magnitude of the sex difference.

When looking at the moderator of point labeling, using the ANOVA model, type of labeling (endpoints, endpoints and midpoints, all points) significantly moderated the sex difference in jealousy, $F_{interaction}(2, 998) = 6.51, p = .002$. The fixed-effects meta-analysis model also showed a significant effect of this moderator, $Q(2) = 6.08, p = .048$. Scales that labeled only endpoints ($g = .366$) and scales that labeled endpoints and a midpoint ($g = .358$) showed larger effects than scales that labeled all points ($g = -.080$).

When looking at moderator of scale length, using the ANOVA model, there was no evidence of moderation, $F_{interaction}(1, 1000) = .33, p = .565$. The fixed-effects meta-analysis model also showed no effect of this moderator, $Q(1) = 1.51, p = .471$.

When looking at moderator of upper anchor (*very* vs. *extremely*), using the ANOVA model, there was no evidence of moderation, $F_{interaction}(1, 1000) = .29, p = .591$. The fixed-effects meta-analysis model also showed no effect of this moderator, $Q(1) = .900, p = .343$.

Discussion

We designed the present study to examine a question at the heart of the debate on evolved sex differences in jealousy—whether the sex

difference is an artifact of the forced-choice method. The present article offers three contributions that we believe answer that question and may help close this aspect of the debate. First, we provide a comprehensive explanation of the empirical implications of the theory of evolved sex differences in jealousy. Citing past research (Sabini & Silver, 2005), we show how misunderstandings regarding these implications have led researchers to erroneously interpret statistically significant theory-supportive effects as not supporting the theory. Second, across 14 independent replications, we demonstrate a highly significant aggregate sex difference of small to moderate size but one that does not come out statistically significant in even a majority of the individual replications. Third, we identify a significant moderator of the sex difference: the type of labeling of the response scale.

With respect to the 14 replications, our data illustrate that individual studies with modest sample sizes (less than 80 people) can often fail to achieve statistical significance even when there is a true effect present. In the present experiment, using traditional significance levels, we failed to detect significance in more than half of the individual forms. Nevertheless, when collapsing across all studies, we easily detected a small to moderate effect, $g = .300$, $p < .001$. Of course, this phenomenon is likely to occur in any research domain examining small effects with limited statistical power.

In addition, our data identified a significant moderator of the ability to detect the sex difference in jealousy: labeling of scale points. We found that labeling all points eliminated the sex difference in jealousy. This supports our concern that scale labeling can influence the ability to detect the sex difference in jealousy. Our other hypothesized moderators of scale length and the upper anchor term did not moderate the sex difference.

Researchers on both sides of the debate regarding sex differences in jealousy have often pointed to single studies as providing sufficient support for their position. Opponents of the theory have sometimes considered a single failure to replicate the sex difference as sufficient evidence to refute the theory. Like-

wise, proponents of the theory have often considered a single significant sex difference as firmly establishing the theory. We suggest that both sides should be wary of placing too much weight on the results of a single study—a tendency Tversky and Kahneman (1971) labeled *the belief in the law of small numbers*. According to Tversky and Kahneman, the belief in the law of small numbers is the common misconception that a small sample will be highly representative of the overall population. Describing a researcher falling prey to this belief, Tversky and Kahneman explain, “He has unreasonably high expectations about the replicability of significant results. He underestimates the breadth of confidence intervals” (p. 109).

Perhaps most revealing is the question of what the expected distribution of effects would look like if there were no sex difference in jealousy. If there were no sex difference in jealousy, one would expect to find in our data a normal distribution of effect sizes centered around 0, but that is not what we found. Instead, we found a distribution of effect sizes centered around .300 with five significant sex differences in the theory supportive direction and none in the opposite direction. If there were truly no sex difference in jealousy—if the population effect size were 0 and all the significant theory-supportive results published in the literature represented Type I errors—there should be at least some publications with significant effects in the opposite direction (women, relative to men, reporting more jealousy in response to sexual than emotional infidelity). To our knowledge, however, there has never been a published reversal of the sex difference in jealousy.

An important limitation to the results presented in this article needs to be discussed: our use of a student sample. Some readers might question whether the results reported in this manuscript would replicate with a nonstudent sample. As is the case with all convenience samples (especially student convenience samples), we cannot generalize our results beyond this sample. To obtain generalizability, researchers would ideally use a probability sample. Given the difficulties in obtaining a probability-based sample, however, the next

best option would be to replicate the results of this study in both other student and nonstudent samples. Previously, Harris's (2003) meta-analysis found that the sex difference in jealousy replicated in both student and nonstudent samples, although the difference was smaller in the nonstudent samples. Additionally, Edlund and colleagues (2006) found that the sex difference in jealousy replicated in an adult sample when using continuous measures. Thus, although the research in this article would be bolstered by the addition of a additional samples, we believe that the evidence suggests that the sex difference occurs in both student and nonstudent samples.

As an anonymous reviewer suggested, one promising option for future research would be to attempt to engineer infidelity scenarios that would produce responses in the theory-refutational direction (Greenwald, Pratkanis, Leippe, & Baumgardner, 1986). If researchers successfully design such scenarios, data from these scenarios could refute or identify important boundary conditions on the theory of evolved sex differences in jealousy. If researchers could not design such scenarios, it would further support the theory.

There are other avenues for potential work to understand the moderators of jealousy. One approach could examine whether the perceived locus of responsibility moderated the jealous response (e.g., how would jealousy change if the partner expressed interest in the rival vs. the rival expressed interest in the partner). Another approach could examine the individual acts of "cheating" to determine whether the severity of the action moderates the jealous response (e.g., would oral sex provoke more of a jealous response than passionate sexual intercourse). Finally, as with many areas of research, researchers studying sex differences in jealousy often adapt prior methods when designing their studies. This has the benefit of increasing comparability between studies but carries the risk that the results stem from the particular stimuli and measurement instruments used (and reused). Varying these elements would reduce the likelihood that inadequate stimulus sampling (Wells & Windschitl, 1999) could provide a plausible alternative explanation for the results.

Ultimately, infidelity in relationships is a very damaging event. This work demonstrates that there are measurable differences in how men and women respond to an imagined infidelity. This suggests that counselors working with couples who have experienced an infidelity should be sensitive to the different emotional reactions men and women have in response to these events.

In conclusion, our study suggests that the sex difference in jealousy is a real effect of small to moderate size but that like all effects of this magnitude, it will not appear at a statistically significant level every time researchers estimate it. We believe this should serve as a reminder to all social science researchers about the importance of not falling victim to the belief in the law of small numbers.

References

- Alreck, P. L., & Settle, R. B. (1985). *The survey research handbook*. Homewood, IL: Irwin.
- Alwin, D. F., & Krosnick, J. A. (1991). The reliability of survey attitude measurement: The influence of question and respondent attributes. *Sociological Methods and Research, 20*, 139–181.
- Barrett, H. C., Frederick, D. A., Haselton, M. G., & Kurzban, R. (2006). Can manipulations of cognitive load be used to test evolutionary hypotheses? *Journal of Personality and Social Psychology, 91*, 513–518.
- Becker, D. V., Sagarin, B. J., Guadagno, R. E., Millevoi, A., & Nicastle, L. D. (2004). When the sexes need not differ: Emotional responses to the sexual and emotional aspects of infidelity. *Personal Relationships, 11*, 529–538.
- Biostat. (2006). *Comprehensive Meta-Analysis Version 2.0* [Computer software]. Englewood, NJ: Author.
- Buller, D. J. (2005). Evolutionary psychology: The emperor's new paradigm. *Trends in Cognitive Science, 9*, 277–283.
- Buss, D. M., Larsen, R. J., Westen, D., & Semmelroth, J. (1992). Sex differences in jealousy: Evolution, physiology, and psychology. *Psychological Science, 3*, 251–255.
- Conway, L. G., III, & Schaller, M. (2002). On the verifiability of evolutionary psychological theories: An analysis of the psychology of scientific persuasion. *Personality and Social Psychology Review, 6*, 152–166.
- Daly, M., Wilson, M., & Weghorst, S. J. (1982). Male sexual jealousy. *Ethology and Sociobiology, 3*, 11–27.
- DeSteno, D. A., Bartlett, M. Y., Braverman, J., & Salovey, P. (2002). Sex differences in jealousy: Evolutionary mechanism or artifact of measurement? *Journal of Personality and Social Psychology, 83*, 1103–1116.
- DeSteno, D., Bartlett, M. Y., & Salovey, P. (2006). Constraining accommodative homunculi in evolutionary explorations of jealousy: A reply to Barrett et al. (2006). *Journal of Personality and Social Psychology, 91*, 519–523.

- DeSteno, D. A., & Salovey, P. (1996). Evolutionary origins of sex differences in jealousy? Questioning the "fitness" of the model. *Psychological Science, 7*, 367–372.
- Edlund, J. E., Heider, J. D., Scherer, C. R., Farc, M. M., & Sagarin, B. J. (2006). Sex differences in jealousy in response to actual infidelity experiences. *Evolutionary Psychology, 4*, 462–470.
- Feldman Barrett, L., Robin, L., Pietromonaco, P. R., & Eysell, K. M. (1998). Are women the "more emotional" sex? Evidence from emotional experiences in social context. *Cognition and Emotion, 12*, 555–578.
- Fenigstein, A., & Peltz, R. (2002). Distress over the infidelity of a child's spouse: A crucial test of evolutionary and socialization hypotheses. *Personal Relationships, 9*, 301–312.
- Green, M. C., & Sabini, J. (2006). Gender, socioeconomic status, age, and jealousy: Emotional responses to infidelity in a national sample. *Emotion, 6*, 330–334.
- Greenwald, A. G., Pratkanis, A. R., Leippe, M. R., & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review, 93*, 216–229.
- Harris, C. R. (2000). Psychophysiological responses to imagined infidelity: The specific innate modular view of jealousy reconsidered. *Journal of Personality and Social Psychology, 78*, 1082–1091.
- Harris, C. R. (2002). Sexual and romantic jealousy in heterosexual and homosexual adults. *Psychological Science, 13*, 7–12.
- Harris, C. R. (2003). A review of sex differences in sexual jealousy, including self-report data, psychophysiological responses, interpersonal violence, and morbid jealousy. *Personality and Social Psychology Review, 7*, 102–128.
- Ketelaar, T., & Ellis, B. J. (2000). Are evolutionary explanations unfalsifiable? Evolutionary psychology and the Lakatosian philosophy of science. *Psychological Inquiry, 11*, 1–21.
- Krosnick, J. A. (1999). Maximizing questionnaire quality. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of political attitudes* (pp. 37–57). San Diego, CA: Academic Press.
- Michalski, R. L., Shackelford, T. K., & Salmon, C. A. (2007). Upset in response to a sibling's partner's infidelities. *Human Nature, 18*, 74–84.
- Newstead, S. E., Pollard, P., & Riezebos, D. (1987). The effect of set size on the interpretation of quantifiers used in rating scales. *Applied Ergonomics, 18*, 178–182.
- Sabini, J., & Green, M. C. (2004). Emotional responses to sexual and emotional infidelity: Constants and differences across genders, samples, and methods. *Personality and Social Psychology Bulletin, 30*, 1375–1388.
- Sabini, J., & Silver, M. (2005). Gender and jealousy: Stories of infidelity. *Cognition and Emotion, 19*, 713–727.
- Sagarin, B. J. (2005). Reconsidering evolved sex differences in jealousy: Comment on Harris (2003). *Personality and Social Psychology Review, 9*, 62–75.
- Sagarin, B. J., Becker, D. V., Guadagno, R. E., Nicastle, L. D., & Millevoi, A. (2003). Sex differences (and similarities) in jealousy: The moderating influence of infidelity experience and sexual orientation of the infidelity. *Evolution and Human Behavior, 24*, 17–23.
- Sagarin, B. J., & Guadagno, R. E. (2004). Sex differences in the contexts of extreme jealousy. *Personal Relationships, 11*, 319–328.
- Shackelford, T. K., Michalski, R. L., & Schmitt, D. P. (2004). Upset in response to a child's partner's infidelities. *European Journal of Social Psychology, 34*, 489–497.
- Sheets, V. L., & Wolfe, M. D. (2001). Sexual jealousy in heterosexuals, lesbians, and gays. *Sex Roles, 44*, 255–276.
- Swearingen, D. L. (1999, April). *Consequences of the midpoint response choice for survey researchers*. Paper presented at the annual meeting of the American Educational Research Association, Montreal, Canada.
- Symons, D. (1979). *The evolution of human sexuality*. New York: Oxford University Press.
- Takahashi, H., Matsuura, M., Yahata, N., Koeda, M., Suhara, T., & Okubo, Y. (2006). Men and women show distinct brain activations during imagery of sexual and emotional infidelity. *NeuroImage, 32*, 1299–1307.
- Tversky, A., & Kahneman, D. (1971). Belief in the law of small numbers. *Psychological Bulletin, 76*, 105–110.
- Weems, G. H. (2004). Impact of the number of response categories on frequency scales. *Research in the Schools, 11*, 41–49.
- Weems, G. H., & Onwuegbuzie, A. J. (2001) The impact of midpoint responses and reverse coding on survey data. *Measurement and Evaluation in Counseling and Development, 34*, 166–176.
- Wells, G. L., & Windschitl, P. D. (1999). Stimulus sampling and social psychological experimentation. *Personality and Social Psychology Bulletin, 25*, 1115–1125.
- Wiederman, M. W., & Allgeier, E. R. (1993). Gender differences in sexual jealousy: Adaptionist or social learning explanation? *Ethology and Sociobiology, 14*, 115–140.