Do Human Menstrual-Cycle Pheromones Exist?

Jeffrey C. Schank
University of California, Davis

Research over the past 15 years indicates, contrary to earlier results, that women do not synchronize their menstrual cycles. If women do not synchronize their cycles, this implies there is no mechanism for synchronizing cycles. Since a pheromone mechanism of synchronization is the only plausible mechanism that has been proposed, it follows that that there are no pheromones that modulate the length of menstrual cycles. To test this hypothesis, eight studies were reviewed that reported pheromone effects on menstrual cycles, other behavior, or physiological correlates in women. The prediction was that serious problems would be found in each of these studies. As predicted, serious problems were found in all eight studies. Taken together, these results cast doubt on the existence of pheromones that modulate the length of menstrual cycles.

KEY WORDS: Human pheromones; Menstrual cycles; Menstrual synchrony; Sexual behavior

Yang and Schank (2006) and Ziomkewicz (2006) in this issue of *Human Nature* found that women did not synchronize their menstrual cycles. Both articles attributed non-synchrony to cycle variability. This is in agreement with recent reviews of menstrual synchrony studies (Arden and Dye 1998; Schank 2000a, 2001b, 2004; Wilson 1992), which revealed a number of problems, including the theoretical problem of cycle variability. Subsequent studies that avoided the methodological errors prevalent in earlier research did not find synchrony precisely because of cycle variability (Strassmann 1997; Trevathan 1993; Wilson et al. 1991). The problem of cycle variability is fundamental because rhythms of different frequencies and variability can never synchronize in the sense of stable phase or state matching over cycles (Winfree 1980).

Received March 1, 2005; revisions requested September 2, 2005; revised version accepted December 22, 2005; final version received May 23, 2006

Address all correspondence to Jeffrey C. Schank, Department of Psychology, University of California, One Shields Ave., Davis, CA 95616 USA. Email: jcschank@ucdavis.edu

If menstrual synchrony does not exist, then this has implications for other research its hypothesized existence has generated. When menstrual synchrony was first reported (McClintock 1971), the leading hypothesis was that synchrony was produced by a pheromone mechanism. Specifically, McClintock (1984) proposed that mutual synchronization could occur if there were ovarian phase-dependent pheromones that shortened and lengthened cycles. In Norway rats, McClintock (1984) reported that odors from the follicular phase shortened estrous cycles in recipients, whereas odors from the ovulatory phase lengthened them. Schank and McClintock (1992) showed, using computer simulation, that a coupled oscillator mechanism mediated by pheromones could theoretically synchronize cycles in rats. A coupled-oscillator mechanism can solve the problem of cycle variability by causing rhythms of variable and different frequencies to converge on the same frequency, resulting in stable phase or state matching (Winfree 1980). Schank and McClintock (1997) tested the pheromone component of the coupled-oscillator mechanism in rats under highly controlled conditions and found that the cycle shortening and lengthening effects reported in McClintock (1984) could not be replicated. Instead, contrary to the earlier study, follicular odors lengthened cycles in recipients and ovulatory odors had no effect (Schank and McClintock 1997). Reanalysis of McClintock's (1978) report of synchrony in groups of rats and a new study attempting to replicate synchrony among female rats found no evidence of synchrony (Schank 2001a, 2001c).

The failure to find evidence of rats synchronizing their estrous cycles raised the question of whether any mammals synchronize their menstrual or estrous cycles. Synchrony in groups of chimpanzees (Wallis 1985), lemurs (French and Stribley 1985; but see Monfort, Bush, and Wildt 1996), and hamsters (Handelmann, Ravizza, and Ray 1980) had been reported in the 1980s, but nothing since then. Using computer simulation analysis, Schank (2000b, 2001d) found that the synchrony reported in chimpanzees, lemurs, and hamsters could be explained by methodological errors and that the data are best explained as chance relationships among cycles. Recent experimental studies have supported these results in chimpanzees (Matsumoto-oda and Kasuya 2005) and hamsters (Gattermann, Ulbrich, and Weinandy 2002). Taken together with the problems of menstrual synchrony, it seems reasonable to conclude that there is no indisputable evidence that mammalian females mutually synchronize their menstrual or estrous cycles.

If menstrual synchrony does not exist, then this casts doubt on the existence of pheromones that modulate the length of menstrual cycles. Human pheromones may exist, but the theoretical motivation for searching for them is now gone. There are two possible explanations for reports of pheromones that may directly or indirectly modulate menstrual cycles. First, they may exist independent of the original reasons for looking for them. There are many examples of serendipitous discoveries in science. Second, they may not exist, and the results of previous studies were in error. If the latter hypothesis is correct, it should be possible to find serious flaws in some or all of the studies reporting menstrual-cycle modulating pheromones.

To test this hypothesis, I reviewed all studies directly or indirectly related to pheromone modulation of the menstrual cycle. This is a very small literature of eight studies spanning 25 years. I divided the studies into four sections: (1) axillary secretions and menstrual cycles, (2) detecting specific compounds, (3) luteinizing hormone, and (4) sexual behavior. Because the literature on human pheromones is small, it is possible to analyze each study in detail. For each study, I present the main conclusions, identify errors, and discuss to what degree the data support the conclusions reached.

AXILLARY SECRETIONS AND MENSTRUAL CYCLES

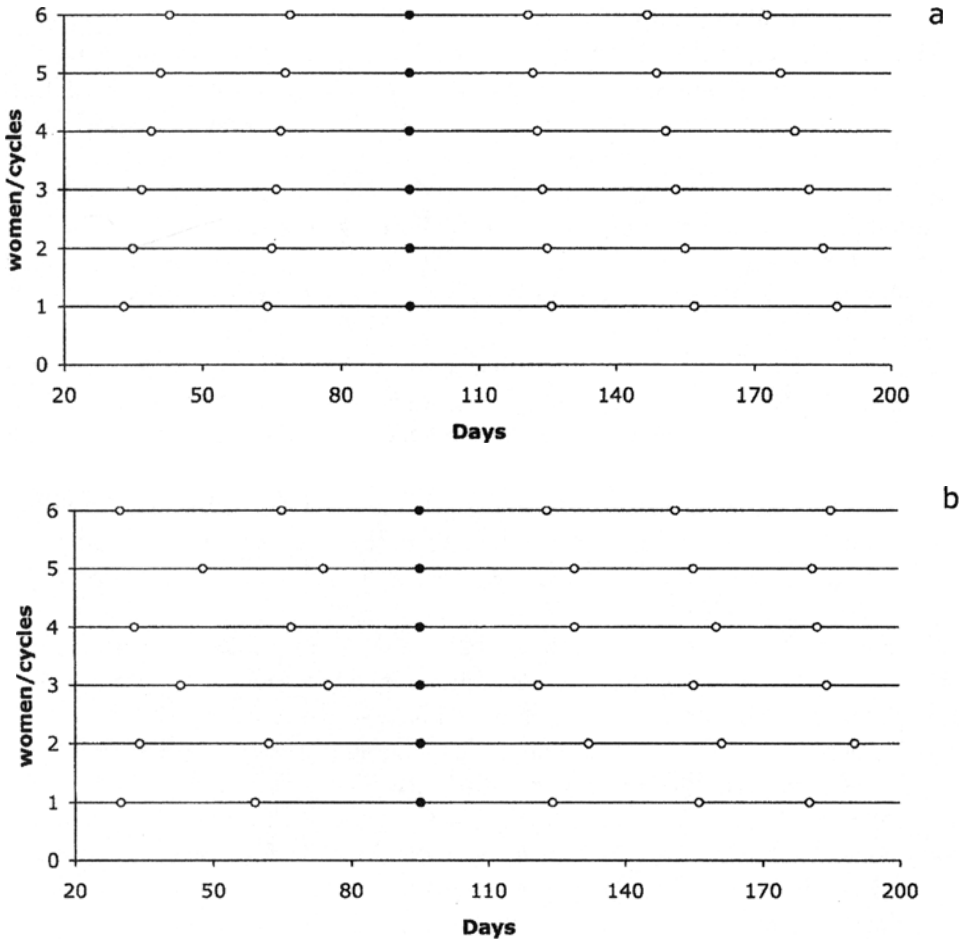
The first study to report a pheromone influence on the menstrual cycle had five experimental subjects, each of whom received axillary secretions from a single donor three times a week for a period spanning five menstrual cycle onsets (Russell et al. 1980). The donor had 28-day cycles throughout the study and was one of the researchers (Wilson 1992). Russell et al. (1980) applied a repeated-measures analysis of variance to differences between the donor and recipients and reported a statistically significant decrease in cycle onset differences indicating that a process of synchronization had occurred.

Wilson (1992) identified several methodological problems with this study. Among the most important was the dropout rate. Initially there were eight women in the experimental condition, and three dropped out for unknown reasons (Wilson 1992). This resulted in a dropout rate of 37.5%. Wilson (1992) also noted that the experimenter conducting the study was the donor, and she was not blind to which participants were in the experimental and control conditions.

Another serious problem was the statistical analysis of the data. Russell et al. (1980) used a repeated measures analysis of variance to determine whether the women synchronized to the donor. However, the data were periodic or quasi-periodic, meaning that the data were not independent, and repeated-measures analysis of variance was not appropriate. We can understand why it is not appropriate by seeing how rhythms with different, consistent cycle lengths repeatedly converge and diverge over time (Figure 1a), as do rhythms with variable cycles (Figure 1b). Thus, repeated-measures analysis of variance applied to such data will inevitably yield statistically significant results owing to repeated convergence and divergence of cycle onsets over time.

To determine if a process of synchronization has occurred, we must first test whether or not onsets are closer together than expected by chance and then observe several consecutive synchronized onsets to rule out periodic or quasi-periodic convergence, which are a property of variable rhythms. To test whether onsets are closer than expected, circular statistical methods can be used (Schank 1997). Stephens (1965) generalized the Kolomogorov-Smirnov test for non-uniform distributions of points around a circle. If points on a circle were closely clustered, then clustering would be an indicator that rhythms are more synchronized than expected by chance.

Figure 1. Two examples of six hypothetical women illustrating the problem of cycle variability. In the first example (*a*), women 1 to 6 have menstrual cycle frequencies of 31, 30, 29, 28, 27, and 26 days, respectively. All cycle onsets match exactly at day 95, but then diverge again. In the second example (*b*), all women have a mean cycle length of 28.5 ± 5 days. Again, all cycle onsets match exactly at day 95 but then diverge owing to cycle variability. Synchronization requires, at a minimum, convergence on the same frequency (i.e., cycle length) and a reduction of cycle variability.



One might also consider the binomial test as an acceptable alternative, but it is not. For example, suppose the entraining rhythm is 30 days; then the expected difference of onsets by chance would be approximately 7.5 days (Schank 2000a). All onsets less than 7.5 days from the entrainer's onset would be counted as synchronized and those greater than 7.5 would not. If there are 10 women in a group, then the onsets of at least 9 must be less than 7.5 days from the entrainer's onset for the binomial test to yield a statistically significant result. The problem with this test is

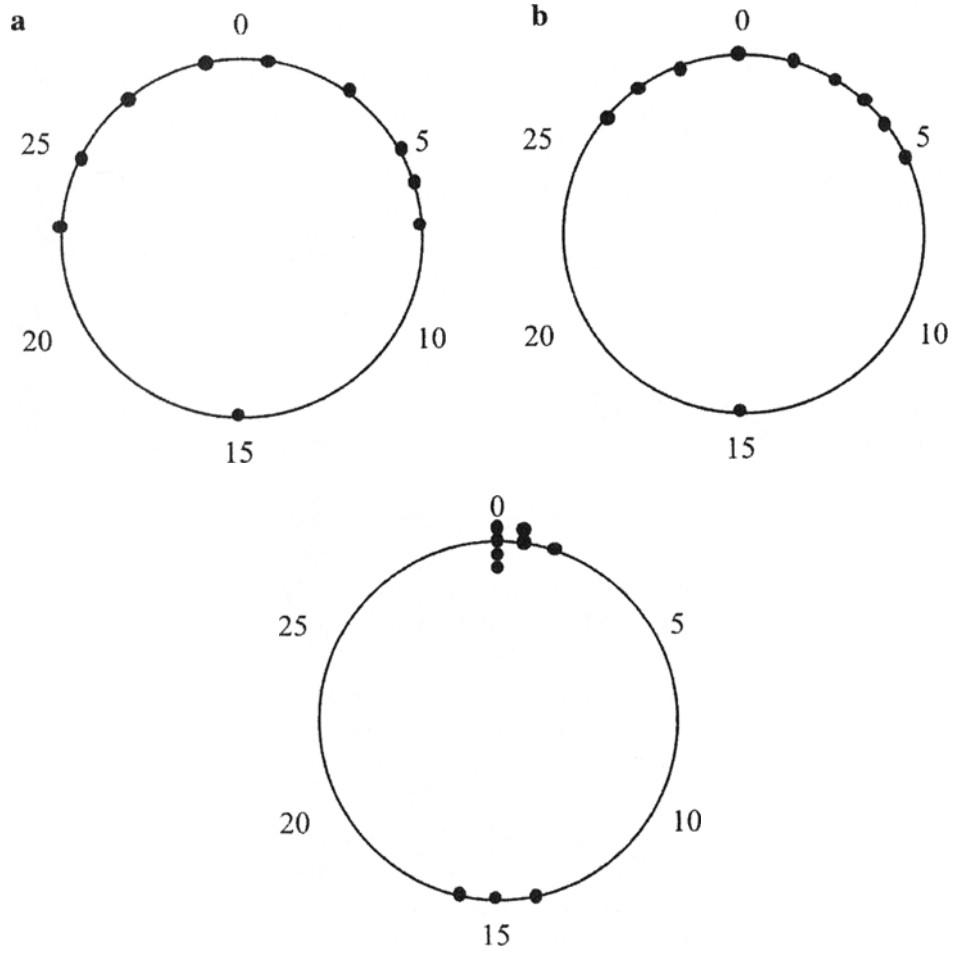
that it does not take into consideration all of the possible phase relationships among the onsets (Figure 2). In Figure 2a, nine onsets are distributed uniformly in the semicircle defined by ±7.5 days from the entrainer, which is significant under a binomial test, but not the Stephens test. When the onsets are closer to the entrainer, both tests yield significant results (Figure 2b), but when most (7 out of 10) are tightly clustered around the entrainer and three are not (Figure 2c), only the Stephens test detects the clustering. This illustrates that analysis of circular data is tricky and requires statistical tests designed for them.

Because participants in the study by Russell et al. (1980) were hypothesized to synchronize to the donor, the appropriate clock for testing synchrony has a period of 28 days since the donor had exactly 28-day cycles. At the end of the study, one participant had a 16-day difference from the donor. This is not possible on a 28-day clock, but a 12-day difference is possible (i.e., 28 – 16 = 12). Using the differences reported for the final month by Russell et. al (1980), a Stephens test ($V_N = .56, p > .15$ for $N = 5$) revealed no significant effect (Figure 3). Thus, the results reported by Russell et al. (1980) were not statistically significant irrespective of other methodological flaws.

An experimental design similar to that used by Russell et al. (1980) was employed in a study by Preti et al. (1986). Ten women were in the experimental condition, nine were in the control condition, and there were four donors. Pads with axillary secretions from three cycles were collected for 12 cycles and extracts from five of these cycles were combined to conform to days 2, 5, 8, 11, 14, 17, 20, 23, 26, and 29 of a 29-day menstrual cycle (Preti et al. 1986). They reported a statistically significant increase in synchrony to donors owing to modulation of recipient cycle length (Preti et al. 1986).

Wilson (1987) pointed out a number of errors in the study. The most serious was that synchronization was impossible given the implementation of the protocol for the experiment. The extract was not delivered in 29-day cycles, or even the actual cycles of donors, but rather in artificial cycles ranging from 21 to 38 days in length (see Table 2A in Preti 1987). Only one of the artificial donor rhythms had a constant cycle length, but it was very short at 24-day cycles. Wilson (1987) concluded the donor extracts were too random in their presentation for synchrony to have occurred. Some of the women purported to have synchronized to the donors did not have changes in menstrual cycle length that accounted for the decrease in the difference between donor cycles and recipient cycles. For example, experimental participant 8 had three consecutive cycles of 28, 29, and 28 days (Table 2A in Preti 1987). The donor extract in this case was applied with cycles of 24, 33, 23, with apparent increase in synchrony of onsets since the onset difference went from +8 to –1. However, the recipient only had a 1-day change in her second cycle, so it was impossible for the recipient to have adjusted her cycles to be closer to the random donor cycles. Indeed, the artificial donor cycle lengths changed on average by 6.3 days as compared to 2 days for donors ($p < 0.02$, df = 9, $t = 2.83$, two-tailed paired t-test). Figure 4 illustrates how the apparent closeness of recipients to donors was

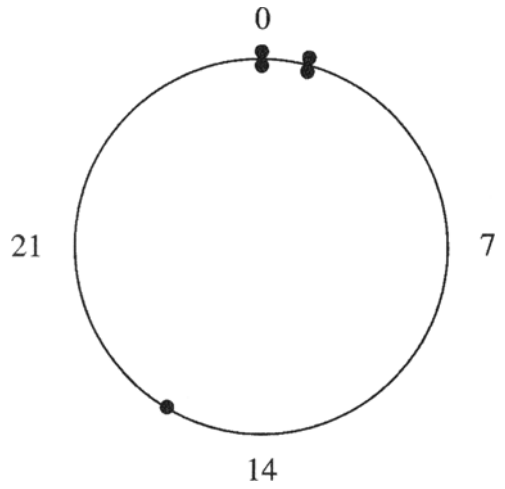
Figure 2. Examples of 10 onsets (indicated by black dots) on a 30-day clock. (*a*) Nine of the 10 onsets are less than 7.5 days from the entrainer (time 0) but are distributed uniformly on the semicircle defined by 7.5 days from the entrainer. Using a binomial test, $p = 0.01$ that 9 of 10 would be within 7.5 days of the entrainer, but $V_N = .33$, $p > .15$ for the Stephens test. (*b*) Nine of the 10 onsets are less than 7.5 days from the entrainer (time 0) but this time they are closer to the entrainer. Again, using a binomial test, $p = 0.01$, but $V_N = .53$, $p < .05$ for the Stephens test. (*c*) Finally, only 7 of the 10 onsets are less than 7.5 days from the entrainer (time 0), but they are very close to the entrainer. Using a binomial test, $p = 0.17$, but $V_N = .53$, $p < .05$ for the Stephens test.



due to changes in the artificial cycles of donors. Thus, the change in the difference was primarily due to nearly random changes in the presentation of donor secretions and not to changes in recipients' menstrual cycle lengths.

The study by Stern and McClintock (1998) was inspired by the first two studies but greatly improved on the protocol methods and analytical procedures. The aim of this study was to test the coupled-oscillator hypothesis of menstrual synchrony

Figure 3. Analysis of Russell et al.'s (1980) data using the Kupier-Stephens (Stephens 1965) circular statistic for testing non-uniform distributions on a circle. The 28-day clock represents the donor and the black dots represent cycle onset for each recipient. The recipient onsets were not statistically significantly clustered at the end of the study.



(McClintock 1984). Based on previous results on estrous synchrony in Norway rats (McClintock 1978, 1984; Schank and McClintock 1992), the authors hypothesized that pheromones released during the follicular phase would shorten the cycles of recipients while pheromones released during the ovulatory phase would lengthen them. As with the previous two studies, axillary secretions were collected from nine donors who wore pads and did not use deodorant soaps or perfumes. To avoid consciously detectable odors as much as possible, putative pheromones were extracted with alcohol (Stern 1992; Stern and McClintock 1998). The nine donors contributed equally to the follicular and ovulatory phase secretions, and these alcohol-extracted secretions were delivered daily to 30 recipients during the experimental phase by wiping a thawed pad on the upper lip (Stern 1992; Stern and McClintock 1998).

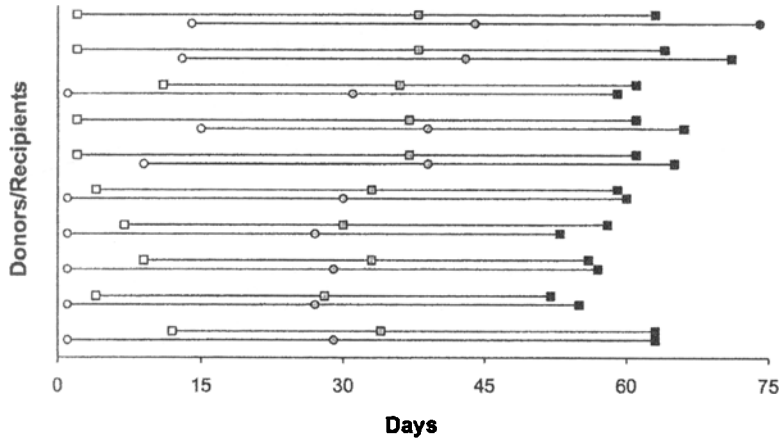
The study spanned five menstrual cycles: one baseline cycle, and two cycles of treatment with one odor followed by two cycles of treatment from the other odor. The data were then transformed into change data using a procedure illustrated in Figure 5 and described briefly in Stern and McClintock (1998:179):

> we created within-subjects controls by measuring the effect on the menstrual cycle in terms of a change in length from each individual subject's cycle preceding each condition. (For experimental subjects this was the cycle that preceded exposure to each type of compound; for control subjects this was the cycle that preceded exposure to the carrier, 70% alcohol).

It is described more specifically in Stern's (1992:85) dissertation:

Figure 4. (*a*) An illustration of the onset relationships between artificial donors' onsets and recipients' onsets for two cycles, where boxes are donors and circles are recipients. Most of the changes in cycle length were in the artificial donor cycles. According to the protocol, all donor cycles should have been 29 days in length. (*b*) If the protocol had been performed correctly, we could plot recipient onsets on a 29-day clock to compare the initial and the final onset relationships. As can be seen, there was little change in phase relationships from start to finish, and the onsets are uniformly distributed on a 29-day clock.

a



b

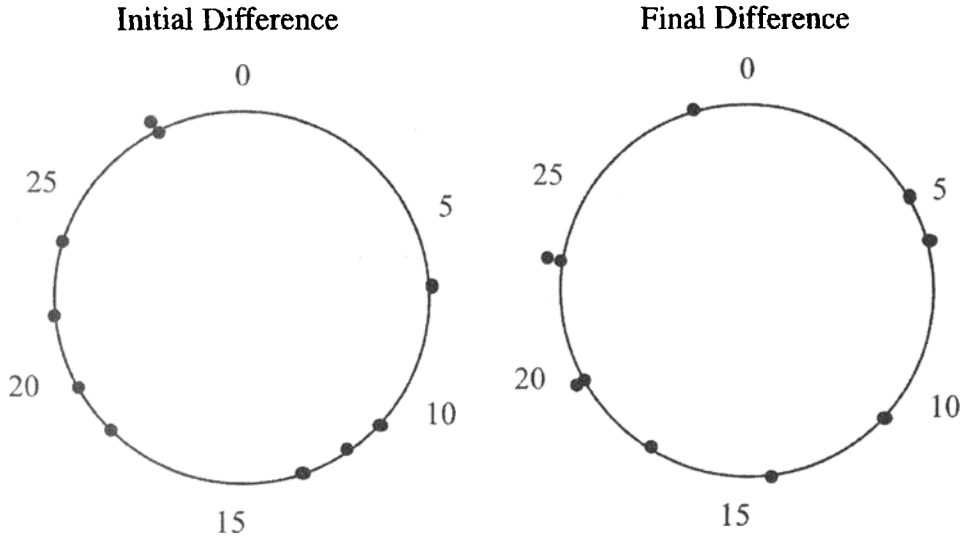
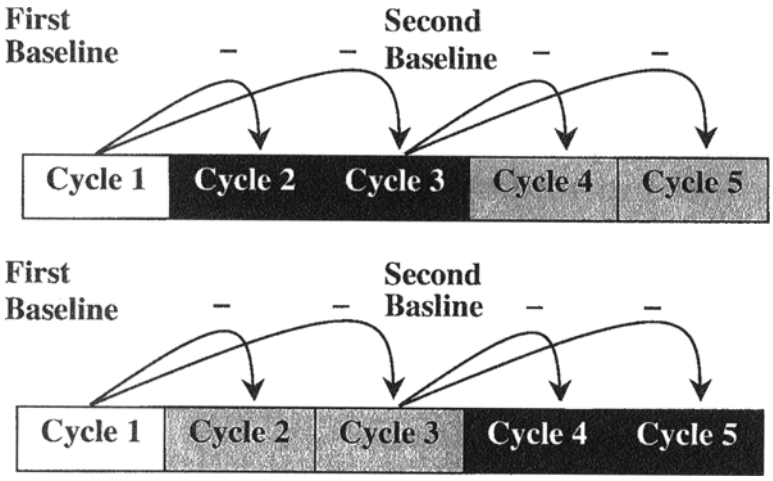


Figure 5. Stern and McClintock's crossover experimental design. Cycle 1 is their first baseline cycle and cycles 2 through 5 are treatment cycles (grey represents follicular odor and black represents ovulatory odor). In the S-M transformation, cycle 1 is subtracted from cycles 2 and 3. Cycle 3 is both a treatment and the second baseline cycle, which is subtracted from cycles 4 and 5.



During the first secretion treatment, the referent cycle length was the baseline cycle length. In this case the cycle length for each of the two cycles during the first treatment was subtracted from the cycle length of the baseline cycle. During the second secretion treatment, the crossover condition, the referent cycle length was the pre-change cycle length. In this case the cycle length for each of the two cycles during the second treatment was subtracted from the cycle length of the pre-change cycle. In other words, the pre-change cycle was the last cycle of the first treatment.

They reported a statistically significant difference in the response of women to ovulatory (1.35 ± 0.50 days) and follicular (-1.43 ± 0.71 days) secretions: $F_{1, 18} = 4.31$ (Stern 1992), $F_{1, 18} = 4.32$ (Stern and McClintock 1998), $p < 0.05$. The slight discrepancy in F-values between Stern (1992) and Stern and McClintock (1998) suggests that the F-value was subsequently rounded up. However, in either case, $F_{1, 18} = 4.31$ or 4.32 is not the critical cutoff for $\alpha = 0.05$. Instead, $F_{1, 18} = 4.33$ is the critical cutoff, which represents the difference between an alpha of 1/19 and 1/20. Thus, their main results represent a statistical trend and not a statistically significant result at the $\alpha = 0.05$ level.

Nevertheless, the trend was close to statistical significance. Do these results suggest a possible effect? There is an error in data transformation used to analyze the data. By subtracting cycle 1 from cycles 2 and 3 and cycle 3 from cycles 4 and 5 (see Figure 5), the treatment conditions were confounded. This can be illustrated by a simple numerical example. Suppose the actual mean cycle length for the women in this study was 28 days, but for each cycle and each group of women the sample

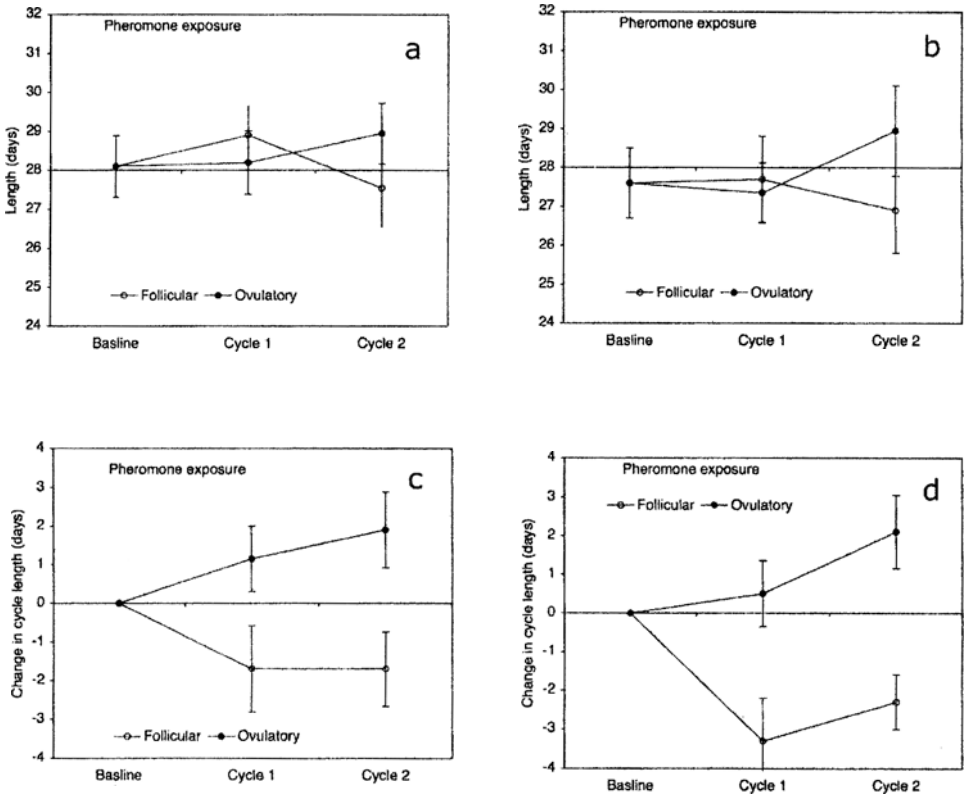
means can vary. For example, suppose for cycle 1 (baseline) the mean is 28 days, for cycle 2 the mean is 29 days, and for cycle 3 it is 26 days. Using the Stern-McClintock (S-M) method, the change from cycle 1 to cycle 2 is 1 day (29 – 28) and for cycle 1 to cycle 3 it is –2 days (26 – 28), for an average change of –1 day. Now, suppose cycle 4 has a mean of 28 days and cycle 5 has a mean of 27 days. If we subtract cycle 3 from cycles 4 and 5 we get 2 (28 – 26) and 1 (27 – 26) for an average increase of 1.5 days. However, if the same baseline (i.e., cycle 1) were used, then we would get 0 (28 – 28) and –1 (27 – 28) for an average decrease of –.5 days. Clearly, depending on which cycle is taken as the baseline, radically different results are obtained. Which cycle should be used? The key is that we want to know how putative pheromones change cycle length from an untreated baseline cycle mean. Cycle 3 is a treatment condition and by definition cannot be a measure of the baseline cycle length. Thus, cycle 1 should have been used as the baseline throughout.

It is also important to determine whether data transformed using the S-M method could explain the results reported in Stern (1992) and Stern and McClintock (1998) as spurious. To answer this question, I constructed random artificial data sets of menstrual cycles using a truncated normal distribution (Schank 2000a, 2001b). Figure 6 illustrates two null data results that when transformed using the S-M method yielded results similar to those of Stern (1992) and Stern and McClintock (1998). In untransformed data (Figure 6a and b), there was no significant effect but after the S-M transformation, the data exhibit considerably different patterns (Figure 6c and d), which are statistically significant. This demonstrates that the S-M transformation can produce significant results from null data.

The study by Jacob et al. (2004) was motivated by the studies reviewed above and by Shinohara et al. (2000, 2001) reviewed below. These authors (Jacob et al. 2004:423) hypothesized that axillary and nipple secretions from breastfeeding mothers would increase the variability in menstrual-cycle length of women receiving these compounds: "Specially, this novel study hypothesized that pheromones from breastfeeding (non-ovulating) women would increase the variability of ovarian cycles, particularly by lengthening them, and also by shortening them, as the effects of pheromones depend on the state of the ovary at the time of pheromone exposure (Schank and McClintock 1992)." To test this hypothesis they chose "regression analyses that would reveal how breastfeeding compounds might perturb cycle length, based on the ovarian cycles' initial state just prior to exposure" (Jacob et al. 2004:423). However, regression analysis only assesses the relationship among cycles. A perfect regression can be achieved when women do not change cycle length from cycle to cycle. If pheromones from breastfeeding women "increase the variability of ovarian cycles, particularly by lengthening them, and also by shortening them, as the effects of pheromones depend on the state of the ovary at the time of pheromone exposure," then change in cycle-length variability must be tested statistically.

Methodologically, this study (Jacob et al. 2004) was very complicated owing to the loss and exclusion of women at different stages of the study (see Figure 1 in

Figure 6. Two simulated data sets of random cycle lengths drawn from a truncated normal distribution (Schank 2000a, 2001b) with mean and standard deviation as reported by Stern (1992) and Stern and McClintock (1998). Graphs *a* and *b* are the untransformed data sets. Neither are statistically significant: (*a*) $F_{1,19} = 0.564$, $p = .462$ and (*b*) $F_{1,19} = 1.721$, $p = .205$. After the S-M data manipulation (see Figure 5), the data now have a radically different pattern and are statistically significant: (*c*) $F_{1,19} = 4.729$, $p = .043$ and (*d*) $F_{1,19} = 4.89$, $p = .04$.



Jacob et al. 2004). Eighty-seven women began the study but only 40 to 42 women (the exact number is uncertain) were included in the final analysis from cycle 1 to cycle 2. To calculate a range for normal cycles, they assumed a mean cycle length of 29 days with a standard deviation of 5 days. Normal cycle lengths were defined as falling in the range of 24 to 34 days (i.e., mean ± s.d.), and it was only women with cycles in this range for the cycle prior to a test condition that were included in the next test condition. The choice of normal range was based on data and analysis presented in Harlow et al. (2000). However, Harlow et al. (2000) define normal cycles as falling in the range of 18 to 40 days. Thus, a range of 24 to 34 days may be too narrow for normal cycles.

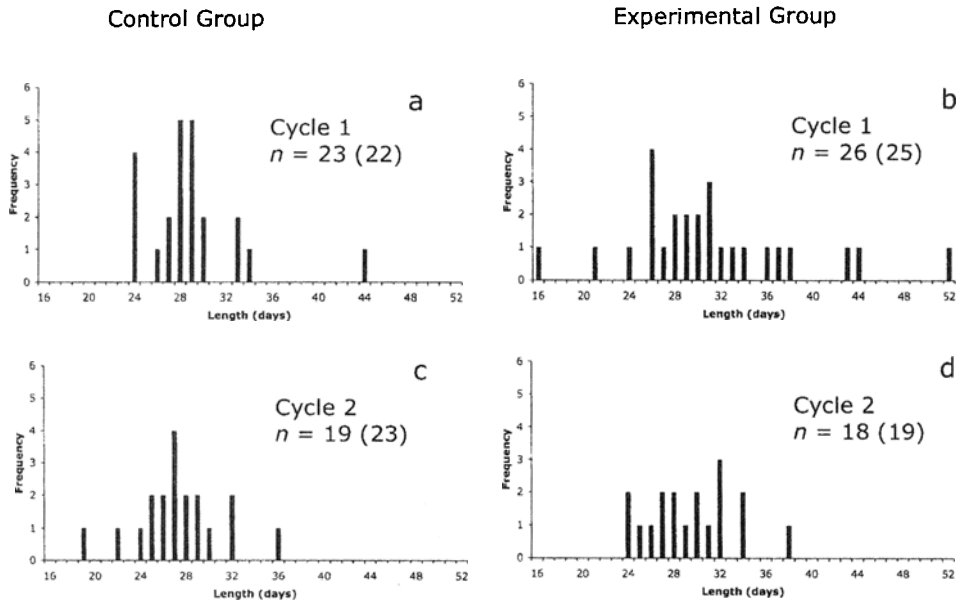
Irrespective of disputes about the range of normal cycles, it is crucial to determine whether the data support the hypothesis that pheromones from breastfeeding increase cycle-length variability. Jacob et al. (2004) reported neither descriptive statistics nor statistical analysis of the cycle variability produced by breastfeeding compounds. Instead, they illustrated an apparent difference in variation for Cycle 1 control vs. experimental groups (Jacob et al. 2004:425, Figure 2). No differences in variability are mentioned for the repeated Cycle 2 condition. To determine whether there was an effect of cycle variability, I reconstructed the cycle-length data sets from Figures 2 and 3 in Jacob et al. (2004:425–426). To do this, I exported the figures from the PDF file of the paper and used computer-generated straight lines to align the data points in the graph. Some points may have identical coordinates, which cannot be detected in any single graph, but by looking at all graphs it was possible to reconstruct a near-complete data set for Cycle 1 and a partial data set for Cycle 2. For Cycle 1, the experimental group, the vertical axis of Figures 2B and 3A and the horizontal axis of Figure 3C in Jacob et al. (2004:425–426) provide information on the number and length of cycles. By counting the number of cycles of a given length along each axis, a better estimate of the actual data was obtained. The same was done for the control group (Figures 2A, 3B, and 3D in Jacob et al. 2004:425–426). The main problems I encountered during this process were inconsistencies among graphs and numbers reported in the text. For example, in Figure 2B (experimental group; in Jacob et al. 2004:425) one woman had a baseline cycle of 27 days followed by a 16-day cycle. However, in the baseline of the Cycle 1 graph in Figure 3A in Jacob et al. (2004:426), this data point is not plotted. Also, there were supposed to be 25 women in Cycle 1 of the experimental condition, and the women in the Cycle 2 experimental condition were supposed to be those whose cycles (in the Cycle 1 condition) were in the range of 24 to 34 days. In Figure 2B in Jacob et al. (2004:425), 8 cycles fell outside that range, so only 17 women should have been included in the Cycle 2 condition, but in the text, Cycle 2 was reported to have 19, and in Figure 3C in Jacob et al. (2004:426), 18 points were plotted. These inconsistencies can be resolved if we assume there were errors in the numbers of women for each condition reported in the text.

For the Cycle 1 condition, Figures 2 and 3 in Jacob et al. (2004:425–426) imply that the control group consisted of 23 women and the experimental group consisted of 26 women. The text reports 22 and 25, respectively. For the experimental condition in Figure 2B (Jacob et al. 2004:425), the graph indicates 8 women were outside the range of 24 to 34 days for Cycle 1. If we subtract 8 from 26, we get 18, which is the number of points plotted in Figure 3A (Jacob et al. 2004:426). Also, in the text, 22 women were reported to be in Cycle 1 and 23 in Cycle 2 (Jacob et al. 2004:425). However, since this was a correlational study, the number of women should be the same for each cycle. This can be resolved if there were actually 23 women in Cycle 1. According to Figure 2A (Jacob et al. 2004:425), one woman in Cycle 1 is outside the range of 24–34 days, which would leave 22 women for the Cycle 2 test condi-

tion. Based on these assumptions for resolving these errors, I was able to determine whether breastfeeding compounds increased cycle-length variability.

The cycle-length distributions are plotted in Figure 7. For the control group (Figure 7a), s.d. = 4.29, and for the experimental group (Figure 7b), s.d. = 7.29. Levine's test for unequal variance yielded a statistical trend but not a significant difference. The cycle distributions for the Cycle 2 condition can only be estimated from Figures 3C and 3D in Jacob et al. (2004:426) because identical pairs of points cannot be detected. For the experimental condition, 18 points could be detected (Figure 3C in Jacob et al. 2004:426), and for the control condition, 19 (Figure 3D in Jacob et al. 2004:426). If $n = 18$ is correct for the experimental group and $n = 19$ is 3 less than in the control group (with three data points identical to the ones plotted), then it is possible to determine whether the putative breastfeeding compounds increase cycle variability. Distributions of these cycles are illustrated in Figure 7c and d. For the Cycle 2 test condition, the s.d. for the control group is 3.77 and for the experimental group, 3.79, which is not significantly different based on Levine's test. There were three data points in the control group that were not in this calculation, but because

Figure 7. Cycle-length distributions from Jacob et al. (2004). For cycle 1, (*a*) s.d. = 4.29 for the control group and (*b*) 7.29 for the experimental group. Levine's test using the median yielded a trend ($F_{1,47} = 3.44$, $p = .07$) toward more variability in the experimental than under control conditions. For cycle 2 (*c* and *d*), s.d.= 3.77 for the control group (*c*) and 3.79 for the experimental group (*d*). Levine's test revealed no difference ($F_{1,35} = 0.29$, $p = .59$) in variability between the experimental and the control conditions. Thus, the data do not support the conclusion that breastfeeding women secrete compounds that increase cycle variability.



they must be identical to data points already plotted, they are not likely to substantially change the calculated standard deviation. Moreover, the standard deviation for the experimental group, 3.79, is less than that typical of women in the reported age range (Harlow et al. 2000). Thus, it does not appear that breastfeeding secretions increased the variability of women's menstrual cycles.

Finally, Jacob et al. (2004) reported that menstrual cycles in Cycle 1 are a better predictor of cycles in Cycle 2 for the experimental condition than for the control condition. However, their protocol for selecting only women in the narrow range of 24- to 34-day cycles confounds the purported difference, since a reduction from 27 women (the number who initially started in the experimental condition) to 18 women is a reduction of one-third of the more variably cycling women. One would expect higher correlations between cycles as intra-woman variability is reduced, leaving women who tend to have the same or similar lengths from cycle to cycle. These results show that the data not only do not support the hypothesis but also are inconsistent with it. Jacob et al. (2004:427) conclude that "breastfeeding compounds maintained the type of cycle length an individual had at the time of exposure. That is, women who were having a long cycle continued to have long cycles; in addition, women who were having a short cycle continued to have short cycles." If breastfeeding compounds caused a woman to have cycles of similar length, then this would appear to be inconsistent with the hypothesis that breastfeeding compounds increased cycle variability. Thus, the data do not appear to support conclusion that breastfeeding compounds modulate women's menstrual cycle lengths.

DETECTING SPECIFIC COMPOUNDS

Morofushi et al. (2000) hypothesized that pheromones that synchronize menstrual cycles are mediated by the main olfactory system. If this were true, then women might be able to detect the presence of pheromones by smell. They tested two chemicals secreted in the armpits (5a-androst–16-en–3-one and 5a-androst–16-en–3a-ol) to test whether women who synchronized their cycles could smell these compounds. Sixty-seven women living in 18 rooms (which implies 13 rooms of four women and 5 rooms of three women) recorded their menstrual onset dates from January to July. To determine whether a woman synchronized with other women, each woman in a room was compared with every other woman in a room to determine the closest onset between pairs in July. For example, if A, B, C, and D are four women in a room with onset dates in July of 4, 11, 17, 23 (there could be more than one onset in July depending on cycle length), then the closest woman for A, B, C, and D are A with B, B with C, C with D, and D with C. They calculated the mean cycle length for each woman and classified each woman as synchronized in July if she was less than one-quarter of her mean cycle length from the closest woman in a room. In the example above, if each woman had a mean cycle length of 28 days, then to be synchronized a woman's onset must be less than 7 days from that of the closest woman. Thus, women B, C, and D are synchronized and A is not. However, this is

not a criterion of synchrony. Synchrony is a relationship among rhythms in which phases or states match over time (Winfree 1980). If A is synchronized with B, then it is mathematically necessary that B must be synchronized with A. With this requirement, only women C and D could be synchronized and not A and B.

It is also incorrect to identify individuals as synchronized. Synchrony is not a property of the individual but the group. In the previous example, a better approach would have been to calculate the average difference for each woman from each other in a group. In this case, we would get 13, 6.3, 8.3, and 12.3 for A, B, C, and D, respectively. In this case, only B appears to be synchronized. However, the group does not deviate from a uniform distribution, nor is the average difference among the women (10.5 days) less than 7 days. Thus, neither A, B, C, nor D can be synchronized because the group is not synchronized.

Using the method just discussed, Morofushi et al. (2000) reported 24 of 64 pairs of women synchronized their cycles in the sense that they had a cycle difference greater than one-quarter of their mean cycle length from January to March and less than one-quarter of their mean cycle length in July. Three women were excluded whose cycle length remained constant from March to July. We cannot determine directly whether synchrony occurred using their method, but we can determine whether cycle onsets were likely randomly related to each other. Using Monte Carlo simulation, I determined using their method that by chance at least 80% of the women in July would have been classified as synchronous. If 50% of those classified in July as synchronous were not synchronous during January and March, then by chance about 40% of the women in this study would have been classified as synchronous using their methods. Thus, by chance, 25.6 women should have been scored as "synchronized." They reported 24 women were synchronized, which is about what would be expected by chance. We can conclude that the menstrual onsets of women were likely randomly related and so were not synchronized.

Morofushi et al. (2000) also reported that the synchronized women had significantly lower olfactory detection thresholds for 5a-androst–16-en–3a-ol than the unsynchronized women, suggesting that this compound may be a pheromone regulating menstrual synchrony. Using the Mann-Whitney test of ranks, they reported $p < 0.01$. They found no significant difference for 5a-androst–16-en–3-one. However, in their Figure 1a (Morofushi et al. 2000:409), there are 23 data points and not 24 with a mean of 2.9 days ± 1.9 as reported in their Table 1 (Morofushi et al. 2000:410). In their Figure 1b (Morofushi et al. 2000:409) there are 24 data points for the synchronized group. When I computed the Mann-Whitney test for the data in Figure 1a (Morofushi et al. 2000:409), I obtained a one-tailed p of 0.046 and two-tailed p of 0.091. If the missing data point is from a woman with a threshold of 6 or greater, then the difference is not significant for even a one-tailed p-value. Thus, the synchronized group did not appear to be synchronized, and data errors render the compound detection results inconclusive.

Luteinizing Hormone

Shinohara et at. (2000) followed up on Stern and McClintock (1998) and reported that follicular-phase axillary compounds increased the frequency of luteinizing hormone (LH) pulses (Table 1), whereas follicular-phase axillary compounds decreased the frequency of LH pulses. Shinohara et al. (2001) subsequently reported that 5a-androst–16-en–3a-ol decreased the frequency of LH pulses just as

Table 1. Results from Shinohara et al. (2000, 2001) with Unexplained Data Anomalies in the Follicular Phase (FP), Ovulatory Phase (OP), Isopropyl Alcohol (IPA), and Androgen Conditions

	Interval (min)			Frequency (pulses/4 h)		
Shinohara et al. 2000	FP	OP	IPA	FP	OP	IPA
n	8	7	5	8	7	5
Before application	66.4	60.3	59‡	3.68	4.02	4.17
s.e.	3.6	2.7	4.9†	.22	.18*	.33*
s.d.	10.18	7.14	10.96	0.62	.48*	.74*
During application	47.9	71.6	58.8‡	5.16	3.4	4.18
s.e.	3.4	3.7	4.7†	.34	.18	.31
s.d.	9.62	9.79	10.51	.96	.48	.69
Example						
Before application	60.00	56.67	63.33			
During application	53.33	70.00	66.67			

Shinohara et al. 2001	Androgen	IPA	Androgen	IPA
n	6	5	6	5
Before application	53.9	54‡	4.57	4.59
s.e	3.9	4.9†	.81*	.88*
s.d.	9.55	10.96	1.98*	1.97*
During application	66.00	53.8‡	3.64	4.59
s.e.	1.5	4.7†	.21	.83
s.d.	3.67	10.51	.51	1.86
Example#				
Before application	70	56.67		
During application	85	60		

† It is extraordinarily unlikely that the standard errors would be the same for controls across experiments.
‡ If the controls are exchanged between studies, the experimental conditions are problematic when compared with the controls.
* The standard errors and deviations are too high for the control groups because the intervals between pulses followed the same standard in both studies.
# The androgen example exceeds 5 s.d. from the mean during application, which is mathematically impossible for this small sample size.

ovulatory phase compounds were reported to do (see Table 1). Shinohara's et al. (2001) study was based on Morofushi et al. (2000) reporting that menstrual synchrony is related to the ability to smell 5a-androst–16-en–3a-ol, but as we saw above, menstrual synchrony probably did not occur in that study.

These results are surprising. The sample sizes were very small, and paired *t*-tests were used to detect changes from baseline. Because the *p*-values were highly significant ($p < 0.01$, $n = 7$ and 8 [Shinohara et al. 2001]; $p < 0.05$, $n = 6$ [Shinohara et al. 2001]), this implies that all of the women responded in the same direction to the putative pheromone. Even if we do not consider problems discussed above for Stern and McClintock (1998) and Morofushi et al. (2000), a 100% response rate in all experimental conditions is hard to explain when Stern and McClintock report that 68% of the women in their study responded to axillary secretions and the rest did not. According to Morofushi et al. (2000), many women in the control condition did not detect 5a-androst–16-en–3a-ol. The women in Shinohara et al. (2000, 2001) were not screened prior to testing for being responders or non-responders and yet all apparently responded.

The Shinohara et al. (2000, 2001) publications were not separate studies, as shown by the fact that the intra-assay coefficients of variation (4.17% and 0.66%) were the same, the age range was the same (19–25 years), and the number of women in the control groups was the same. In Shinohara et al. (2000), the control mean before application was 59 minutes (s.e. = 4.9) between LH pulses and after application it was 58.8 minutes (s.e. = 4.7), but in Shinohara et al. (2001) the control mean before and after application was 5 minutes less than was reported in the previous publication (i.e., 54 and 53.9) with exactly the same standard errors. When the means for the frequency of pulses are calculated, they agree with the inter-pulse intervals for the control conditions across studies, but the standard deviations do not agree (Table 1). Thus, these errors do not appear to be typographical. The control conditions reported in the second publication lead to findings that more strongly support their conclusions (Table 1). In addition, the androgen example subject in Shinohara et al. (2001) is more than 5 standard deviations from the mean during application. The reported standard error is mathematically impossible to calculate with this extreme outlier when $n = 6$. The data reported by Shinohara et al. (2000, 2001) contain unresolved inconsistencies that again render the results inconclusive.[1]

Sexual Behavior

McCoy and Pitino (2002) reported that a putative menstrual-cycle-altering pheromone (Athena Pheromone 10:13) increased the sociosexual behavior of young women. The chemical nature of this putative pheromone will not be revealed until it is patented (McCoy and Pitino 2002). However, the authors claim that the pheromone is a synthetic version of a natural pheromone discovered by Preti et al. (1986; see above) and subsequently chemically identified by Preti et al. (1987).

Women in this study were assigned to either of two conditions by selecting be-

tween one of several identical boxes (on a tray) containing vials with either Athena 10:13 or SD40 alcohol, resulting in 23 women in the pheromone condition and 20 women in the placebo condition. Seven sociosexual variables were measured: sexual intercourse (SI), sleeping next to a romantic partner (SNRP), petting/affection/kissing (PAK), informal dates (ID), formal dates (FD), and male approaches (MA). The data were averaged separately for each variable for the two baseline weeks and the six treatment weeks. The data for each variable were analyzed using 2×2 contingency tables for the number of individuals who increased over baseline versus those who either did not change or decreased. No attempt was made to control for increased Type I error rate resulting from testing six variables for significance. McCoy and Pitino (2002:367) reported that "A significantly greater proportion of pheromone users compared with placebo users increased over baseline in frequency of *sexual intercourse*, *sleeping next to a partner*, *formal dates* and *petting/affection/kissing*."

The main problem is that contingency tables only test the independence of data, they cannot determine whether the users increased over baseline. Since the complete data set was published, these claims can be reanalyzed to test for changes from baseline. For each of the six variables measured, I used paired *t*-tests to assess changes from baseline for both pheromone and placebo conditions for each of the variables. No significant increases occurred over baseline for any of the variables in the pheromone condition (Figure 8), but I did find statistically significant decreases from baseline in the placebo condition for PKA, SNRP, and SI, and a trend for FD, but not for MA. Thus, the experimental hypothesis failed for all six variables; there was no evidence that Athena Pheromone 10:13 increased the sociosexual behavior of young women above baseline.

Why did variables change in the placebo condition but not in the pheromone condition? One explanation is that there were several differences between the two conditions. For the variables of age, height, weight, body mass index, menstrual-cycle length, and relationship status for women, McCoy and Pitino (2002) reported only a statistically significant difference in height: women in the placebo group were about 2 inches taller than the pheromone group on average. In addition, I found that the women in these two groups differed on several other variables not reported by McCoy and Pitino (2002), as shown here in Table 2. Specifically, the placebo-treatment women were more variable in age, the pheromone-treatment women were more variable in body mass index, and perhaps most important, there was a statistically significant difference in initial relationship status (Table 2). The latter may explain the decrease in sociosexual activity in the placebo group. McCoy and Pitino (2002) reported that three women decreased their dating status after baseline. If these women were previously dating steadily, this might be reflected in large decreases in SNRP. If we look at SNRP during the baseline and experimental conditions, we see that four women exhibited a large drop: Spiky (6.0 days to 3.87 days), Green (6.0 to 3.86), Valerie (4.0 to 1.5), and Rhonda (3.5 to 0.5). These four women account for all of the decrease in the placebo condition.

Figure 8. Change from baseline (± s.e.) for variables (*a*) PKA, (*b*) SNRP, (*c*) SI, (*d*) FD, (*e*) MA. None of the changes from baseline were statistically significant (one-tailed) within the pheromone condition using paired *t*-tests: PKA ($t = 1.42$, df = 18, $p = 0.086$), SNRP ($t = 1.24$, df = 18, $p = 0.115$), SI ($t = 1.43$, df = 18, $p = 0.085$), FD ($t = 1.58$, df = 18, $p = 0.066$), and MA ($t = -1.58$, df = 18, $p = 0.066$). However, for the placebo condition, the variables PKA ($t = 2.20$, df = 16, $p = 0.017$), SNRP ($t = 2.11$, df = 16, $p = 0.025$), and SI ($t = 2.09$, df = 16, $p = 0.026$) were statistically significant, FD was close ($t = 1.74$, df = 16, $p = 0.051$), but MA ($t = 1.33$, df = 16, $p = 0.101$) was not statistically significant. An asterisk indicates statistical significance at the $\alpha = .05$ level within treatment conditions. If a Bonferroni correction were applied, then none of these tests would be significant at the $\alpha_6 = 0.0083$ level.

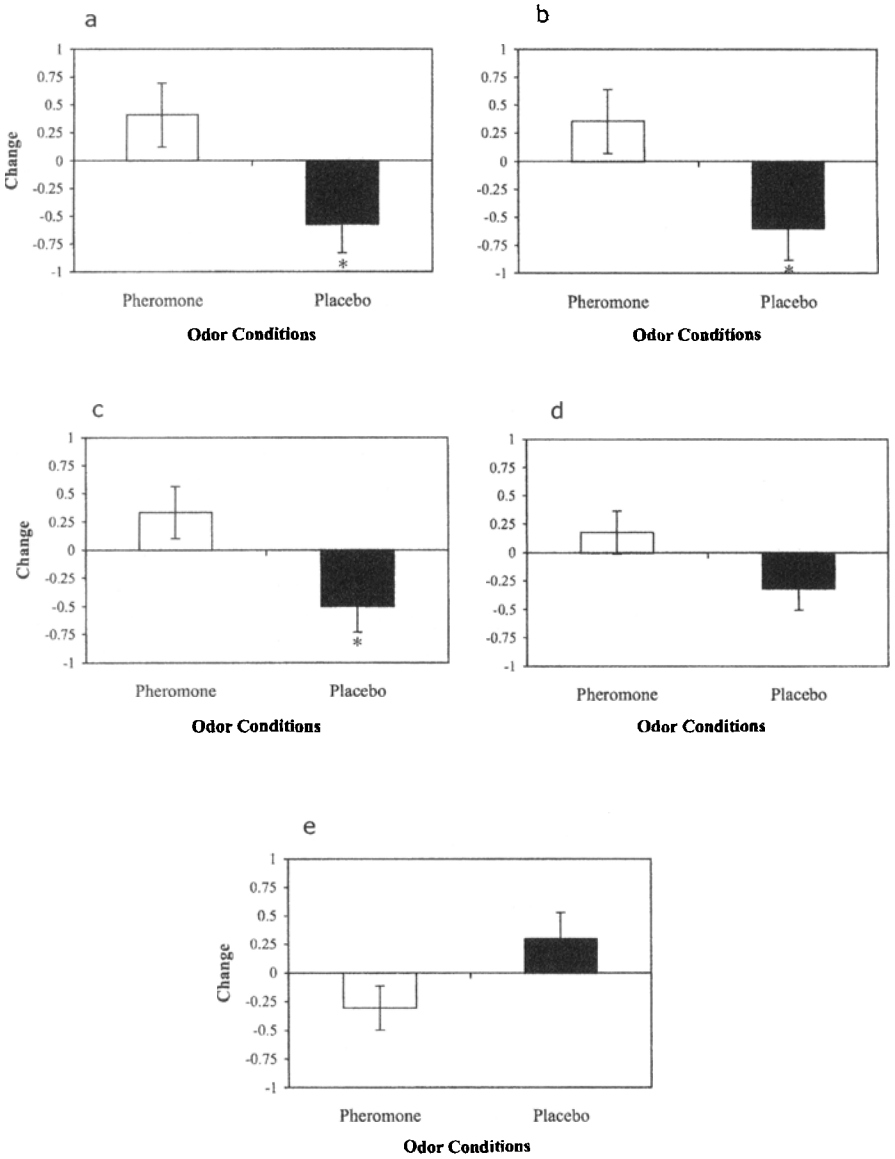


Table 2. Significant Differences in Variability for Age, Body Mass Index, and for Initial Relationship Status for Subjects by Treatment Group in McCoy and Pitino 2002

	Pheromone ($n = 19$)		Placebo ($n = 17$)					
	mean	s.d.	mean	s.d.	$F^{\dagger}$	df_1	df_2	p
Age (years)	25.6	4.6	29.5	8.2	3.18	16	18	.01
Height (in.)	63.7	2.7	65.4[a]	2.2	1.51	18	15	.21
Weight (lb)	134.3	25.1	130.9	21.7	1.34	18	16	.28
Body mass index	23.3	4.4	23.3	2.7	2.66	18	16	.028
Menstrual cycle length	28.8	4.9	29.8	4.9	1.0	18	16	0.5

	Pheromone ($n = 19$)		Placebo ($n = 17$)				
Relationship status	n	%	n	%	c^2	df	p
Not dating	8	42	9	53	7.01	2	.03
Dating	8	42	1	6			
Dating steadily	3	16	7	41			

† F calculated as the ratio of the variances for the two treatment conditions.
[a] $n = 16$

Finally, Athena Pheromone 10:13 should alter menstrual cycle lengths since it was purportedly derived from a pheromone that could (Preti et al. 1987). With regard to menstrual-cycle-length change from baseline, McCoy and Pitino (2002) reported that "For 19 pheromone users, average menstrual cycle length difference between the baseline cycle and the first experimental cycle was –0.11 days [$t(18) < 1$]. For 17 placebo users, this same difference was 1.88 days [$t(16) = 2.24$ (S.E. = 1.37)] and was not significant." However, $t_{16} = 2.24$ (two-tailed) is significant: $p = 0.04$. The mean cycle length at baseline for placebo was 29.8 (s.d. = 4.9) and for pheromone is was 28.8 (s.d. = 4.9), but for the first experimental cycle, the placebo cycle length decreased to 27.6 (s.d. = 4.6) and the pheromone cycle length increased to 28.9 (s.d. = 7.9). The difference for the pheromone condition is 28.9 – 28.8 = 0.1, and for the placebo condition, 27.6 – 29.8 = –2.2, and not –0.1 and 1.88 as reported by McCoy and Pitino (2002). There were too many errors to definitively conclude what happened to menstrual-cycle length.

CONCLUSIONS

If menstrual synchrony does not exist (Arden and Dye 1998; Schank 2000a, 2001b; Strassmann 1997, 1999; Wilson 1992; Yang and Schank 2006; Ziomkiewicz 2006), then it is reasonable to hypothesize that the mechanism proposed to explain it also does not exist. This leads to the prediction the menstrual-cycle-modulating phero-

mones do not exist. To test this hypothesis, eight pheromone studies were reviewed and each had problems as predicted. Russell et al. (1980) made methodological errors and the data did not show an effect of donor. Preti et al. (1986) also made methodological errors, the most serious being that artificial donor cycles varied nearly randomly in cycle length, making entrainment impossible. Stern and McClintock (1998) avoided the methodological errors of the previous studies but their results were a trend and not statistically significant. More important, their study was confounded by using the third cycle, which was a treatment cycle, as a baseline cycle for determining the change in cycle length resulting from ovarian cycle secretions. Jacob et al. (2004) reported that breastfeeding compounds increased cycle variability but reported no statistical test. Subsequent reconstruction of their data revealed no effects of increased or decreased cycle lengths. Morofushi et al. (2000) used an inconsistent measure of individual synchrony, and the data reported do not agree with the statistical analysis reported. Shinohara and colleagues (2000, 2001) appear to have produced methodologically sound studies, but they contain data inconsistencies that draw into question the accuracy of the results. The data of McCoy and Pitino (2002), when reanalyzed, showed only statistically significant changes (decreases) in the placebo condition. One explanation for the placebo change was that the two groups differed on demographic variables, especially relationship status.

Science is not subject to statutes of limitation or prohibitions against double jeopardy. Theories, methods, and data are forever open to critical review. Science only progresses when hypotheses and theories are given the most severe tests possible (Popper 1959). Indeed, even when a theory passes a severe test, errors may be subsequently found in data and methods supporting that theory. This implies that neither theories, data, nor methods can be accepted with absolute certainty. Scientists are fallible, and even the peer review process is no guarantee against error. The fact that errors may occur at all levels of scientific inquiry appears to lead to the skeptical view that all of science is on an equally uncertain footing. However, by repeatedly scrutinizing theories, data, and methods to weed out errors, we can have growing confidence in those that survive. This is a neverending process, but the more we critically scrutinize previous results, the more confident we can be in those theories, data, and methods in which we fail to find errors. Perhaps future studies will find indisputable evidence of pheromones that modulate menstrual cycles, but the studies to date have not.

Jeff Schank is associate professor of psychology at the University of California, Davis. His main research interests are in computational and biorobotic modeling of group behavior and the development of sensorimotor behavior in animals.

NOTE

1. Since the publication of these two articles, I have repeatedly attempted to contact the first author by email to clarify these anomalies but never received a reply.

REFERENCES

Arden, M. A., and L. Dye
1998 The Assessment of Menstrual Synchrony: Comment on Weller and Weller (1997). *Journal of Comparative Psychology* 112:323–324.

French, J. A., and J. A. Stribley
1985 Synchronization of Ovarian Cycles within and between Social Groups in Golden Lion Tamarins (*Leontopithecus rosalia*). *American Journal of Primatology* 12:469–478.

Gattermann, R., K. Ulbrich, and R. Weinandy
2002 Asynchrony in the Estrous Cycles of Golden Hamsters (*Mesocricetus auratus*). *Hormones and Behavior* 242:70–77.

Handelmann, G., R. Ravizza, and W. J. Ray
1980 Social Dominance Determines Estrous Entrainment among Female Hamsters. *Hormones and Behavior* 14:107–115.

Harlow, S. D., X. Lin, and M. J. Ho
2000 Analysis of Menstrual Diary Data across the Reproductive Life Span: Applicability of the Bipartite Model Approach and the Importance of Within-Woman Variance. *Journal of Clinical Epidemiology* 53:722–733.

Jacob S., N. A. Spencer, S. B. Bullivant, S. A. Sellergren, J. A. Mennella, and M. K. McClintock
2004 Effects of Breastfeeding Chemosignals on the Human Menstrual Cycle. *Human Reproduction* 19:422–429.

Matsumoto-oda, A., and E. Kasuya
2005 Proximity and Estrous Synchrony in Mahale Chimpanzees. *American Journal of Primatology* 66:159–166.

McClintock, M. K.
1971 Menstrual Synchrony and Suppression. *Nature* 229:244–245.
1978 Estrous Synchrony and its Mediation by Airborne Chemical Communication (*Rattus norvegicus*). *Hormones and Behavior* 10:264–276.
1984 Estrous Synchrony: Modulation of Ovarian Cycle Length by Female Pheromones. *Physiology and Behavior* 32:701–705.

McCoy, N. L., and L. Pitino
2002 Pheromone Influences on Sociosexual Behavior in Young Women. *Physiology and Behavior* 75:367–375.

Monfort, S. L., M. Bush, and D. E. Wildt
1996 Natural and Induced Ovarian Synchrony in Golden Lion Tamarins (*Leontopithecus rosalia*). *Biology of Reproduction* 55:875–882.

Morofushi, M., K. Shinohara, T. Funabashi, and F. Kimura
2000 Positive Relationship between Menstrual Synchrony and Ability to Smell. *Chemical Senses* 25:407–411.

Popper, K.
1959 *The Logic of Scientific Discovery.* London: Hutchinson.

Preti, G.
1987 Reply to Wilson. *Hormones and Behavior* 20:547–550.

Preti, G., W. B. Cutler, C. M. Christensen, H. Lawley, G. R. Huggins, and C. R. Garcia
1987 Human Axillary Extracts: Analysis of Compounds from Samples Which Influence Menstrual Timing. *Journal Chemical Ecology* 13:717–731.

Preti, G., W. B. Cutler, C. R. Garcia, A. Krieger, G. R. Huggins, and H. J. Lawley
1986 Human Axillary Secretions Influence Women's Menstrual Cycles: The Role of Donor Extract of Females. *Hormones and Behavior* 20:474–482.

Russell, M. J., G. M. Switz, and K. Thompson
1980 Olfactory Influences on the Human Menstrual Cycle. *Pharmacology Biochemistry & Behavior* 13:737–738.

Schank, J. C.
1997 Problems with Dimensionless Measurement Models of Synchrony in Biological Systems. *American Journal Primatololgy* 41:65–85.

2000a Menstrual-Cycle Variability and Measurement: Further Cause for Doubt. *Psychoneuroendocrinology* 25:837–847.
2000b Can Pseudo Entrainment Explain the Synchrony of Estrous Cycles among Golden Hamsters (*Mesocricetus auratus*)? *Hormones and Behavior* 38:94–101.
2001a Do Norway Rats (*Rattus norvegicus*) Synchronize Their Estrous Cycles? *Physiology and Behavior* 72:129–139.
2001b Menstrual-Cycle Synchrony: Problems and New Directions for Research. *Journal of Comparative Psychology* 115:3–15.
2001c Oestrous and Birth Synchrony in Norway Rats, *Rattus norvegicus*. *Animal Behaviour* 62:409–415.
2001d Measurement and Cycle Variability: Reexamining the Case for Ovarian-Cycle Synchrony in Primates. *Behavioural Processes* 56:131–146.
2004 Avoiding Synchrony as a Strategy of Female Mate Choice. *Nonlinear Dynamics, Psychology, and Life Sciences* 8:147–176.

Schank, J. C., and M. K. McClintock
1992 A Coupled-Oscillator Model of Ovarian-Cycle Synchrony among Female Rats. *Journal of Theoretical Biology* 157:317–362.
1997 Ovulatory Pheromone Shortens Ovarian Cycles of Female Rats Living in Olfactory Isolation. *Physiology and Behavior* 62:899–904.

Shinohara K., M. Morofushi, T. Funabashi, D. Mitsushima, and F. Kimura
2000 Effects of 5a-androst–16-en–3a-ol on the Pulsatile Secretion of Luteinizing Hormone in Human Females. *Chemical Senses* 25:465–267.

Shinohara K., M. Morofushi, T. Funabashi, and F. Kimura
2001 Axillary Pheromones Modulate Pulsatile LH Secretion in Humans. *Neuroreport* 12:893–895.

Stephens, M. A.
1965 The Goodness-of-fit Statistic V_N: Distribution and Significance Points. *Biometrika* 52:309–321.

Stern, K. N.
1992 *Ovulation in Women: Methods of Detection, Behavioral Correlates, and Pheromone Regulation.* Ph.D. Thesis, University of Chicago.

Stern, K., and M. K. McClintock
1998 Regulation of Ovulation by Human Pheromones. *Nature* 392:177–178.

Strassmann, B. L.
1997 The Biology of Menstruation in *Homo sapiens:* Total Lifetime Menses, Fecundity, and Nonsynchrony in a Natural Fertility Population. *Current Anthropology* 38:123–129.
1999 Menstrual Synchrony Pheromones: Cause for Doubt. *Human Reproduction* 14:579–580.

Trevathan, W. R., M. H. Burleson, and W. L. Gregory
1993 No Evidence for Menstrual Synchrony in Lesbian Couples. *Psychoneuroendocrinology* 18:425–435.

Wallis, J.
1985 Synchrony of Estrous Swelling in Captive Group-Living Chimpanzees (*Pan troglodytes*). *International Juornal of Primatology* 6:335–350.

Wilson, H. C.
1987 Female Axillary Secretions Influence Women's Menstrual Cycles: A Critique. *Hormones and Behavior* 21:536–550.
1992 A Critical Review of Menstrual Synchrony Research. *Psychoneuroendocrinology* 17:565–591.

Wilson, H. C., S. H. Kiefhaber, and V. Gravel
1991 Two Studies of Menstrual Synchrony: Negative Results. *Psychoneuroendocrinology* 16:353–359.

Winfree, A. T.
1980 *The Geometry of Biological Time.* Berlin: Springer-Verlag.

Yang, Z., and J. C. Schank
2006 Women Do Not Synchronize Their Menstrual Cycles. *Human Nature* 17:433–447.

Ziomkeiwicz, Anna
2006 Menstrual Synchrony: Fact or Artifact? *Human Nature* 17:419–432.