Web-based image norming: How do object familiarity and visual complexity ratings compare when collected in-lab versus online?

DWAYNE E. PARÉ AND GEORGE S. CREE

University of Toronto Scarborough, Toronto, Ontario, Canada

It has been demonstrated previously that, for some experimental paradigms, Web-based research can reliably replicate lab-based results. Yet questions remain as to what types of research can be reproduced, and where differences arise when they cannot be. The present article examines the effect of research location (laboratory vs. online) on normative data collection tasks. Specifically, participants were randomly assigned to a laboratory or online condition and were asked to rate 593 photorealistic images on the basis of object familiarity ($N = 103$) and object visual complexity ($N = 98$). Dependent measures were compared across location conditions, including response latencies and image rating agreement. Our results suggest that norming data collected online are reliable, but an interesting interplay between task type and research location was observed. Specifically, we found that participating online (i.e., a more familiar environment) leads to systematically higher familiarity ratings than in the lab (i.e., an unfamiliar environment). These differences are not found when the alternate complexity rating task is used.

As Web-based psychological research increases with the proliferation of access to the Internet, benefits and limitations associated with conducting traditional laboratory research online must be considered (see Reips, 2000, for an overview). Some benefits include the ability to have data collected continuously without the need for constant experimenter supervision, as well as having access to large heterogeneous samples. A potential major limitation is loss of laboratory control. Although this limitation would appear to be an obvious concern, numerous studies have found that Web-based research can often replicate lab-based results with little drawback (Birnbaum, 2004; Dandurand, Shultz, & Onishi, 2008; Huang, 2006; Krantz & Dalal, 2000; McGraw, Tew, & Williams, 2000; Metzger, Kristof, & Yoest, 2003; Musch & Klauer, 2002).

In uncommon cases where Web-based experiments do not replicate their laboratory counterparts, it has been suggested that the differences themselves can be informative (e.g., Krantz & Dalal, 2000). In most instances, the differences are not due to lack of laboratory control, but instead can be attributed to more meaningful causes, such as changes in social attitudes (Buchanan, 1998–1999), or an interaction between task type and research location (Stern & Faber, 1997). Therefore, examination of why some types of research do not replicate online is as important to the growing area of Web-based research as determining which types can be reproduced.

The present article adds to this line of investigation by examining the effect of research location (laboratory vs. online) in two image norming experiments in which

object familiarity ratings and object visual complexity ratings were collected. Although use of the Web for normative data collection afforded us the benefits associated with online studies—namely, automated data collection and a large sample of participants—it also begged the question whether or not normative data collected online would look like data collected in the lab. We believed the data would be similar, but thought other factors, such as the type of normative task, would have to be considered. Therefore, the purpose for this research is twofold: (1) to show that online data collection is a viable means for obtaining image norming data, and (2) to determine whether any differences exist in the data, and if so, to provide an explanation for those differences.

The first experiment examined the effect of research location on object familiarity ratings. Participants were randomly assigned to participate either in the lab or online,¹ and were asked to rate the familiarity of objects presented as photorealistic images on a computer screen. Both locations used a Web-based version of the experiment to ensure that all details of the user experience were identical (e.g., user interface, instructions, stimulus display). On the basis of the breadth of previous findings replicating lab-based research online, we predicted that overall object familiarity ratings collected via the Web would be comparable to the ratings collected in-lab (e.g., objects rated as highly familiar in the lab would also be rated as highly familiar online). However, we also hypothesized that we would find differences in response latencies, and environmental context effects of research location, on the familiarity rating task.

D. E. Paré, dpare@psychexperiments.com

First, we believed that participants' response latencies would be greater when objects are rated online. With researcher scrutiny absent in the online condition, it was assumed that participants would be more likely to take their time or divide their attention while rating objects; we expected, therefore, to see a significant difference in the time participants took to rate object familiarity between conditions.

Second, familiarity ratings were predicted to be comparable between conditions, but not identical, due to the effect of environmental context. Specifically, we predicted that the participants' location would influence how they rated the objects. To understand this prediction, it is important to discuss the processes involved in making a judgment of familiarity.

Familiarity ratings are likely based on several types of information, some of which are consciously available, some not. For example, participants very likely assess the number of examples they can consciously bring to mind, from memory, in which they have encountered the object being displayed. In the developmental literature, it has been shown that children can make use of environmental context—such as a familiar environment—to assist in task performance (Acredolo, 1979) and to employ sophisticated cognitive strategies to their task (Ceci & Bronfenbrenner, 1985). There is also the possibility that participants in the online condition could make use of their familiar environments to help consciously recall more examples of the objects being presented.

Other types of information not as consciously accessible probably also contribute to the familiarity assessment (e.g., the ease with which one can process the object in the image). In the memory literature, it has been demonstrated that people's feelings of familiarity are influenced by the fluency of processing (Jacoby & Dallas, 1981; Jacoby & Whitehouse, 1989; Joordens & Merikle, 1992; Whittlesea, 1993). In some of these studies, manipulations are designed to increase fluency, leading participants to be more likely to rate a stimulus as "old" in recognition memory tasks. It is quite possible that something similar happens with familiarity ratings. More recently, Whittlesea and Williams (2000, 2001a, 2001b) have illustrated a discrepancy-attribution hypothesis, which posits that feelings of familiarity can be influenced less by fluency and more by the discrepancy detected between the expected and actual processing of a stimulus. Although these theories of how familiarity works may not fully agree, they do suggest that feelings of familiarity are based on expectations and can be dependent on context. A familiar environment could give the participant a global feeling of familiarity, which could color the rating computed for each object, leading to ratings that differ in a familiar environment (online) from those in an unfamiliar environment (the lab).

To summarize: If familiarity data are comparable but influenced by environmental context effects between conditions, we would expect to see data highly correlated across conditions but with a significant shift in the mean familiarity rating (e.g., an image of a dog is rated as the most familiar object online, with an average rating of 8.5,

and also rated as the most familiar object in-lab, but with a significantly different average rating of 7.5). Alternatively, if differences were found not attributable to environmental context, but to something else—say, noise due to lack of laboratory control—no correlation and random differences in the ratings between conditions would be expected (e.g., the dog image would be rated as most familiar in one location but not in the other, with significantly different mean ratings). A high correlation and no difference between means would suggest that the data were identical in both conditions (e.g., the image of the dog is rated as most familiar, with a rating of 9 in both locations).

EXPERIMENT 1

The first experiment was designed to collect and compare normative data in two research locations: online and in-lab. Specifically, object familiarity ratings were collected for 593 photorealistic images.

Method

Participants. Students from an introductory psychology class at the University of Toronto Scarborough participated in the experiment for course credit ($N = 103$) by signing up for an image norming experiment through the University of Toronto's online Experimental Participation System (EPS:online).

Design. The interface of the experiment was created using Adobe Flash (previously Macromedia Flash), with the majority of the programming done in ActionScript 2.0. All files were hosted on a Linux-based computer running an Apache Web server and MySQL database. The PHP scripting language was used for communication between Flash and MySQL to facilitate organization and storage of the data. Both the laboratory and online version of the experiment ran within a Web browser using the Flash plug-in, and communicated with the same database (i.e., both versions of the experiment were identical, with physical location being the only manipulation). Latency was measured client-side by the Flash program for both location conditions.

Lab participants used PCs running Microsoft Windows XP as the operating system and Internet Explorer as the Web browser with the Flash Player plug-in. Online participants were free to use any computer they chose, as long as it was Internet-capable and had the Flash Player plug-in for the Web browser of their choice.

Location of the experiment was not advertised within the experiment details. Instead, participants received one of two possible e-mails following sign-up. The "lab-based" version of the e-mail informed participants of the room where the experiment would take place on campus, and the "Web-based" version of the e-mail provided students with a link to the experiment Web site. Location condition was randomly assigned with an approximate 3-to-1 ratio in favor of the online condition.²

Procedure. Participants were told that images of a variety of objects would appear individually on the screen and that they should rate each object's familiarity on a scale from 1 to 9, to answer the question "How often do you come in contact with or think about the concept referred to by this picture?" A rating of "1" indicated that the concept was not familiar and a rating of "9" indicated that the concept was extremely familiar. A participant who did not recognize an object was told to press the "Don't Know" button on the familiarity scale. The numbers 1–9 were arranged in a circle with the "Don't Know" button at the bottom; the mouse pointer always began in the center of the circle. Participants responded by moving the mouse pointer to one of the numbers, clicking to select the value, then moving back to the center of the circle to click "submit." This ensured that it required an equal amount of effort to respond with each value,

while also avoiding the possibility of the same value being repeatedly pressed as a means of responding faster.

In the practice phase, 10 sample images were presented to acclimate participants with the interface used for rating image familiarity. In the experimental phase, 593 images were presented individually to participants in a successive, nonspeeded manner. After a familiarity rating was submitted for an image, the next image to be rated would appear. The task took approximately 1 h to complete.

Results

Response latencies. Participants in the online condition were on average significantly slower at submitting image familiarity ratings ($M = 3.08$ sec, $SD = 0.40$) than their laboratory counterparts ($M = 2.84$ sec, $SD = 0.50$) [$t(1184) = 10.10, p < .001$].

Familiarity ratings. Ratings for images correlated highly between laboratory and online conditions [$r(591) = .97, p < .001$; see Figure 1A]. Mean familiarity ratings for images were significantly different between location conditions [$t(1184) = -7.06, p < .001$], with online familiarity ratings higher ($M = 5.50, SD = 1.60$) than lab-based familiarity ratings ($M = 4.80, SD = 1.80$; see Figure 1B).

“Don’t know” responses. No difference was found between the laboratory condition ($M = 3.82\%$, $SD = 8.60$) and the online condition ($M = 3.38\%$, $SD = 6.60$) in making a “don’t know” response [$t(1184) = 0.99, n.s.$].

Discussion

As predicted, participants were slower at rating images online than in-lab. We expected this finding because we thought unsupervised participants in the online condition would be more likely to have their attention divided between the task at hand and other distractions while running through the experiment (e.g., surfing the Web, instant messaging), thus increasing their overall rating time. It should be noted, however, that although significant, the difference in response latencies between conditions was quite small (about 24 msec, on average). Millisecond differences hardly seem attributable to participants’ multitasking, but it is possible that larger time differences would be noticeable only at the slow end of the response

latency distributions, where obvious signs of divided attention might be expected. To explore this further, we examined the bottom 10% of the response latency distributions for both conditions and still found nothing to suggest that participants were behaving differently in the online condition. It is possible, therefore, that the differences we found were caused by something other than differences in participant attention.

In the laboratory condition we kept computer hardware, Web browser, and operating system constant while ensuring no programs other than the experiment were running on the computers. In the online condition, participants were able to use any hardware/software combination they wanted. In addition, even if participants fully attended to the experiment, they may have had programs running in the background. Consequently, the differences found in timing may be due to differences in physical computer setup (e.g., differences in hardware, software, or computer load) rather than to diminished participant attention. Reimers and Stewart (2007) also found small timing differences in their examination of online versus in-lab Adobe Flash studies, and suggested that the differences might be attributed to physical differences between locations.

Interestingly, we found that the correlation for familiarity ratings between the two conditions was near perfect. Rank order of object familiarity was almost identical between locations. This finding supports our hypothesis that familiarity data collected online are comparable to data collected in-lab. Neither condition differed in the number of “don’t know” responses made by participants, lending further support to the notion that both groups were responding comparably to the objects presented.

Finally, we found that object familiarity ratings in the online condition were significantly higher (at a rate of 0.7 units on a scale of 1–9) than those in the lab. This finding, combined with our near-perfect correlation of familiarity ratings between conditions, demonstrates that there were no discernible random differences in the data, only a systematic increase in object familiarity in the online condition. This supports the prediction that ratings would be

Figure 1. (A) Average familiarity ratings for 593 images ranked from least familiar object (gnu) to most familiar object (clock), showing a highly consistent relationship between ratings assigned in the lab-based and online conditions. (B) Although consistent, average familiarity ratings were systematically higher in the online condition. $^{**}p < .001$.

different across location conditions. This difference could be attributable to participants' ability to make use of their environment and consciously recall more examples of the objects displayed in the images. Alternatively, there may be some ease-of-processing effect in a familiar environment that is misattributed to the familiarity of the objects displayed, leading to higher ratings in the online condition. In either case, we believe that the resulting context effects are unique to a task dependent on familiarity, and would not result for tasks that do not involve familiarity.

There are, of course, alternative explanations for the significant differences found. It is possible that there is something different about participating online that generalizes to any task performed. Consequently, any rating task being carried out online would always lead to systematically higher ratings. One could imagine that a participant sitting at home while rating items would be much more comfortable than one sitting in the lab. Accordingly, that participant might be more inclined to give higher ratings based on comfort level (consciously or unconsciously) in any task that required ratings.

If being in a familiar location increased baseline familiarity ratings for online participants because of the familiarity task itself, changing the task from one dependent on familiarity to another task type unrelated to familiarity should eliminate the mean differences found in ratings across condition. On the other hand, if there were something about being in a familiar location that generalized to any rating task, increased mean ratings would be expected, even on a task unrelated to familiarity.

To test this hypothesis, a second experiment was run, using a similar procedure to the familiarity task but using visual complexity rating as the task. We predicted that the results from the second experiment would be similar to those found in the first (i.e., a high correlation of ratings between conditions and a slower response latency in the online condition). But it was also predicted that the mean rating differences found in the familiarity task would be eliminated because of the change in task type.

EXPERIMENT 2

The second experiment was designed to collect and compare normative data in two research locations: online and in-lab. Specifically, object visual complexity ratings were collected for 593 photorealistic images. Moreover, change in normative task from the first experiment (familiarity to visual complexity) was implemented as a hypothesized way of eliminating context effects dependent on familiarity.

Method

Participants. Students from an introductory psychology class at the University of Toronto Scarborough participated in the experiment for course credit ($N = 98$) by signing up for an image norming experiment through the University of Toronto's online Experimental Participation System (EPS:online).

Design. The interface and implementation of the second experiment were nearly identical to those of the first experiment, with only minor modifications to reflect the changes in task instructions and procedures. Participants were randomly assigned to either the

laboratory or online condition, and the same Adobe Flash-based user interface for the experiment was run across conditions.

Procedure. Participants were told that images of objects would appear on the screen and that they should rate the images' visual complexity on a scale from 1 to 9 around the question "How visually complex is this picture?" Instructions stated that visual complexity was referring to the amount of detail, or the intricacy of the detail in the image, and that complexity of the actual image itself should be rated, not the real-life object that the image represents. A rating of "1" indicated that the image was very simple, and a rating of "9" indicated it to be very complex.

In the practice phase, 10 sample images were presented to acclimate participants with the interface used for rating image visual complexity. In the experimental phase, 593 images were presented individually to participants in a nonspeeded manner. After a complexity rating was submitted for an image, the next image to be rated would appear. The task took approximately 1 h to complete.

A final question was added to the second experiment that asked participants to rate the amount of attention they gave throughout the experiment on a scale of 0% (no attention) to 100% (full attention). The addition of this question was intended as a way to account for potential differences between location conditions. Of particular interest was whether or not reported attention was related to differences in response latencies.

Participants were given the following example to help them calculate the amount of attention they should report:

If you gave your full attention to the experiment the entire time, your estimate would be 100%—if you split your attention between the experiment and watching TV, you might estimate 50%. Or if you gave your full attention during half of the experiment, and then divided your attention between Instant Messaging and the experiment for the other half of the time, your estimate would be around 75%.

It was also emphasized that attention reports would not adversely affect course credit, so participants should be as honest as possible.

Results

Response latencies. Participants in the online condition were significantly slower at submitting image visual complexity ratings ($M = 3.10$ sec, $SD = 0.40$) than their laboratory counterparts ($M = 2.90$ sec, $SD = 0.50$) [$t(1184) = 10.10$, $p < .001$].

Complexity ratings. Ratings for images correlated highly between laboratory and online conditions [$r(591) = .97$, $p < .001$; see Figure 2A]. There was no significant difference between mean complexity ratings for online ($M = 4.32$, $SD = 1.67$) and lab-based ($M = 4.56$, $SD = 1.66$) location conditions [$t(1184) = 2.51$, n.s.; see Figure 2B].

Reported attention. The mean attention level reported in the lab-based condition was 87.35% ($SD = 11.61$) and the mean attention level reported in the online condition was 87.40% ($SD = 11.76$).

Discussion

Once again we found that response latency was significantly greater in the online condition than in the lab-based condition by an average of 20 msec. This was expected, on the basis of the findings from the first experiment. It was suggested earlier that this significant difference could be due either to different levels of participant attention across conditions or to something else, such as physical computer setup. To discriminate between these alternative solutions,

Figure 2. (A) Average visual complexity ratings for 593 images ranked from least visually complex object (Allen key) to most visually complex object (cathedral), showing a highly consistent relationship between complexity ratings assigned in lab-based and online conditions. (B) There was no difference found in average visual complexity ratings between conditions. $^{**}p < .001$.

a question was added in the second experiment that asked participants to rate on a scale from 0% to 100% the amount of attention they gave during the experiment. Given that both groups reported astonishingly identical attention levels of just over 87%, plus the fact that the difference in latencies remained quite small, we are inclined to believe that physical computer setup was the primary cause of the differences found in response latencies between conditions.

The correlation between image visual complexity ratings in the lab-based and online conditions was once again nearly perfect. This suggests that rank order of object visual complexity was almost identical between locations. Thus, as was also seen in the familiarity task, an object's position on the visual complexity scale relative to the other objects remained the same in both online and in-lab conditions; this supports our prediction that the visual complexity ratings collected online were comparable to those collected in the lab.

Finally, we found that visual complexity ratings in the online condition were not significantly different from those in the lab. This finding, combined with the near-perfect correlation of visual complexity ratings between conditions, demonstrates that the data collected in the lab and online were indistinguishable. This also supports our prediction that changing the type of task to one not dependent on familiarity would eliminate the systematic difference in ratings found between conditions in the familiarity-dependent task. With these results, we can confidently conclude that it was the task dependency on familiarity that elicited the environmental context effect of research location. Alternative explanations suggested for an increase in baseline ratings such as comfort level are not likely to have been the cause, since they would have also been expected in any online experiment, which clearly was not the case in this experiment.

GENERAL DISCUSSION

Initially, we set out to show that image norming data could be collected on the Web and that it was compa-

rable to data collected in-lab. We also intended to find an explanation for any differences found in the obtained data between location conditions. We believe we have accomplished both; our two experiments have shown not only that collecting image norming data from the Web is reliable, but also that a simple change in a research question can significantly affect the outcome of results. Specifically, we found that familiarity ratings are affected by environmental contexts such as research location, whereas visual complexity ratings are not.

Alternative explanations for the rating differences found between tasks (e.g., comfort level) were considered, but in the end it seems clear that the task dependent on familiarity was sensitive to the more familiar environmental context. Exactly how this sensitivity arises is something that should be examined further. We have suggested that familiarity of environment may have helped participants come up with examples of objects when rating familiarity, or it may have produced an ease-of-processing effect, but a more detailed explanation is needed. Knowing what causes this interplay between familiarity and research location might be useful in determining whether, and what, other tasks would be susceptible to these effects. This understanding would be valuable to anyone planning to run experiments outside of the lab—more specifically, over the Web.

Differences were also found in response latencies across conditions for both experiments. Specifically, we found a small delay in responses made online. We explored the possibility that the cause of these differences was participants' dividing their attention in the online condition, but after careful examination of the response latency distribution, and with matching reports of participant attention in both conditions, we concluded that the millisecond differences found between the lab and online conditions were most likely due to physical factors of the computer systems, such as computer load, that were not controlled.

The present study's use of a randomized design allowed for greater experimental control and gave us the ability to find reliable differences between data obtained online and in-lab. But this design might also be a limiting factor for

generalizing these findings to participants who self-select where they participate. There is some potential that people who choose to participate in experiments online are different from those who do not (Birnbau, 2004). Although this is a possible concern, research has also shown that with careful experimental setup, even experiments posted online and open to large heterogeneous samples can produce valid data (Krantz & Dalal, 2000).

Although many methodologies remain to be tested online, it seems clear that the evidence is growing for the validity and reliability of the data collected. The present findings highlight the importance of considering the type of task and possible effects of research location when Web-based research is planned. This study also emphasizes the point that finding differences in data obtained online versus data obtained in-lab should not necessarily be regarded as a problem; instead, there might be good reason for the differences. They can assist in designing new methodologies, interpreting data, and may even shed light on processes previously unnoticed.

AUTHOR NOTE

This work was supported by Natural Sciences and Engineering Research Council Discovery Grant 72024642 to G.S.C. The authors thank Steve Joordens for helpful discussions related to the manuscript, and Anisha Chandra for help with data collection. Correspondence concerning this article should be addressed to D. E. Paré, University of Toronto Scarborough, 1265 Military Trail, Toronto, ON, M1C 1A4 Canada (e-mail: dpare@psychexperiments.com).

REFERENCES

- ACREDOLO, L. P. (1979). Laboratory versus home: The effect of environment on the 9-month-old infant's choice of spatial reference system. *Developmental Psychology*, **15**, 666-667. doi:10.1037/0012-1649.15.6.666
- BIRNBAUM, M. H. (2004). Human research and data collection via the Internet. *Annual Review of Psychology*, **55**, 803-832. doi:10.1146/annurev.psych.55.090902.141601
- BUCHANAN, T. (1998-1999). Using the World Wide Web for formative assessment. *Journal of Educational Technology Systems*, **27**, 71-79. doi:10.2190/167K-BQHU-UGGF-HX75
- CECI, S. J., & BRONFENBRENNER, U. (1985). "Don't forget to take the cupcakes out of the oven": Prospective memory, strategic time-monitoring, and context. *Child Development*, **56**, 152-164. doi:10.2307/1130182
- DANDURAND, F., SHULTZ, T. R., & ONISHI, K. H. (2008). Comparing online and lab methods in a problem-solving experiment. *Behavior Research Methods*, **40**, 428-434. doi:10.3758/BRM.40.2.428
- HUANG, H.-M. (2006). Do print and Web surveys provide the same results? *Computers in Human Behavior*, **22**, 334-350.
- JACOBY, L. L., & DALLAS, M. (1981). On the relationship between autobiographical memory and perceptual learning. *Journal of Experimental Psychology: General*, **110**, 306-340. doi:10.1037/0096-3445.110.3.306
- JACOBY, L. L., & WHITEHOUSE, K. (1989). An illusion of memory: False recognition influenced by unconscious perception. *Journal of Experimental Psychology: General*, **118**, 126-135. doi:10.1037/0096-3445.118.2.126
- JOORDENS, S., & MERIKLE, P. M. (1992). False recognition and perception without awareness. *Memory & Cognition*, **20**, 151-159.
- KRANTZ, J. H., & DALAL, R. (2000). Validity of Web-based psychological research. In M. H. Birnbau (Ed.), *Psychological experiments on the Internet* (pp. 35-60). San Diego: Academic Press.
- MCGRAW, K. O., TEW, M. D., & WILLIAMS, J. E. (2000). The integrity of Web-delivered experiments: Can you trust the data? *Psychological Science*, **11**, 502-506. doi:10.1111/1467-9280.00296
- METZGER, M. M., KRISTOF, V. L., & YOEST, D. J., JR. (2003). The World Wide Web and the laboratory: A comparison using face recognition. *CyberPsychology & Behavior*, **6**, 613-621. doi:10.1089/109493103322725388
- MUSCH, J., & KLAUER, K. C. (2002). Psychological experimenting on the World Wide Web: Investigating content effects in syllogistic reasoning. In B. Batinic, U.-D. Reips, & M. Bosnjak (Eds.), *Online social sciences* (pp. 181-212). Ashland, OH: Hogrefe & Huber.
- REIMERS, S., & STEWART, N. (2007). Adobe Flash as a medium for online experimentation: A test of reaction time measurement capabilities. *Behavior Research Methods*, **39**, 365-370.
- REIPS, U.-D. (2000). The Web experiment method: Advantages, disadvantages, and solutions. In M. H. Birnbau (Ed.), *Psychological experiments on the Internet* (pp. 89-117). San Diego: Academic Press.
- STERN, S. E., & FABER, J. E. (1997). The lost e-mail method: Milgram's lost-letter technique in the age of the Internet. *Behavior Research Methods, Instruments, & Computers*, **29**, 260-263.
- WHITTLESEA, B. W. A. (1993). Illusions of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 1235-1253. doi:10.1037/0278-7393.19.6.1235
- WHITTLESEA, B. W. A., & WILLIAMS, L. D. (2000). The source of feelings of familiarity: The discrepancy-attribution hypothesis. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **26**, 547-565. doi:10.1037/0278-7393.26.3.547
- WHITTLESEA, B. W. A., & WILLIAMS, L. D. (2001a). The discrepancy-attribution hypothesis: I. The heuristic basis of feelings and familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 3-13. doi:10.1037/0278-7393.27.1.3
- WHITTLESEA, B. W. A., & WILLIAMS, L. D. (2001b). The discrepancy-attribution hypothesis: II. Expectation, uncertainty, surprise, and feelings of familiarity. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 14-33. doi:10.1037/0278-7393.27.1.14

NOTES

1. Although a Web browser was used to run the experiment for both the lab-based and online conditions, throughout the remainder of this article the terms *Web-based* and *online* will refer exclusively to participation outside of the laboratory.

2. Because of the ease of online data collection, more participants were assigned to the online condition than to the lab-based condition. Initially, we set out to compare a subset of the online data with the lab-based data to maintain consistency between conditions, but we found no statistical differences between the reduced subset of data and the larger full sample of our online data. Therefore, throughout this article, we decided to make use of the entire data set in all of the analyses.

(Manuscript received November 24, 2008;
revision accepted for publication March 24, 2009.)