



# Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs<sup>☆</sup>



Robert S. Ryan<sup>\*</sup>, Mara Wilde<sup>1</sup>, Samantha Crist<sup>2</sup>

Psychology Department, Kutztown University, Kutztown, PA, USA

## ARTICLE INFO

### Article history:

Available online 1 March 2013

### Keywords:

Internet-based research  
Experimental research

## ABSTRACT

Research on internet-based studies has generally supported their benefits. However, that research sometimes did not directly compare internet-based to traditional delivery, often used non-experimental methods and small samples, and has not used an entirely unknown effect for the comparison to completely rule out demand characteristics. Our lab experiment ( $N = 180$ ), in which participants were supervised by an experimenter, demonstrated previously unexamined effects. Both the frighteningness and disgustingness of insects made people want to kill them, and females wanted to kill the insects more than males did. There were also some interesting patterns of interaction with gender, but they were not statistically significant. However, an unsupervised, but larger, web-based experiment ( $N = 1301$ ) produced the same significant main effects as the lab study, and the same patterns of interaction that had occurred at a non-significant level in the lab study occurred at a statistically significant level in the web-based study. These results add support to the finding that although web-based studies may incur risks by being unsupervised, such as some participants not being genuinely motivated to follow the instructions correctly, the risks are compensated for by the much larger sample size afforded by the web-based approach.

© 2013 Elsevier Ltd. All rights reserved.

## 1. Introduction

As the use of technology increases, many psychologists are now using the internet as a means of conducting their research. This expansion into web-based studies could raise questions about the validity of this method (e.g., Hewson, 2003), but both advantages and disadvantages of internet research have been uncovered. After briefly reviewing those advantages and disadvantages, this paper will point out some gaps in the literature that the current study will address.

### 1.1. Advantages of internet-based research

One of the advantages of internet based research is that at a relatively low cost it can provide large samples that are diverse and

come from underrepresented populations (Birnbaum, 2004; Skitka & Sargis, 2006). Importantly, this efficiency can be obtained while producing similar results (Hewson, 2003; McGraw, Tew, & Williams, 2000). For example Carlbring et al. (2007) found equivalent results for internet-based and paper-and-pencil questionnaires for panic disorder and agoraphobia. Whitaker (2007) found that there was no interaction between gender and method of administration on attitudinal measures in spite of gender differences in computer anxiety. Naus, Philipp, and Samsi (2009) found equivalent responses for measures of quality of life and depression, and also for some subscales (although not for others) of a personality measure. Vadillo and Matute (2009) initially found both similarities and some differences in discrimination learning between lab and internet studies. However, they later wished to test the validity of internet-based experimental research by showing similar results between lab and internet versions of a study on an effect that was not well known in the literature. To that end, they succeeded in showing such a similarity for the *augmentation effect* (a situation in which the usual blocking effect in association learning is reversed).

Another advantage of internet-based research is the lack of researcher presence. This can be beneficial in two ways. First, participants are more apt to be frank in their responses because of a decrease in anxiety over the social consequences (Hewson, 2003). Second, because the procedure can be replicated exactly for each

<sup>☆</sup> The initial results of Experiment 2 were reported as a poster presented at the 79th Annual Meeting of the Eastern Psychological Association, Boston, Massachusetts, March 14–16, 2008. The comparison of the two experiments was reported as a poster presented at the University of Scranton's 25th Annual Psychology Conference, Scranton, PA, April 17, 2010.

<sup>\*</sup> Corresponding author. Address: Box 730, Psychology Department, Kutztown University, Kutztown, PA 19530, USA. Tel.: +1 484 646 4324; fax: +1 610 683 4467.

E-mail address: [rryan@kutztown.edu](mailto:rryan@kutztown.edu) (R.S. Ryan).

<sup>1</sup> An undergraduate student at Kutztown University when the research was conducted, is now a law student at the University of Nebraska, Lincoln.

<sup>2</sup> An undergraduate student at Kutztown University when the research was conducted, is now an alumna residing at 4 Welsh Court, Pottstown, PA 19464.

subject there is no possibility of researcher bias (Birnbaum, 2004). Errors in data entry by a research assistant cannot occur when questionnaire studies are conducted over the internet because the subject enters the data directly (Pettit, 2002).

### 1.2. Potential disadvantages of internet-based research

Although these advantages are appealing, there are a variety of potential disadvantages associated with internet research that warrant caution. According to Hewson (2003), the lack of researcher control poses huge problems. It is impossible to know such things as whether the instructions were followed correctly, the state the subject was in at the time of their participation, and whether they took the study seriously. Also, Birnbaum (2004) found that there was an increased dropout rate in web-based rather than lab studies. Another major disadvantage discussed in both Hewson (2003) and Skitka and Sargis (2006) are the ethical issues raised in internet research. These studies have found problems with the delivery of informed consent and debriefing forms, and with the concern of confidentiality in the experiments.

### 1.3. Gaps in the literature

However, the previous literature on advantages and disadvantages of internet research has some gaps that the present study helps to fill. A first example of such a gap is that much of the literature comparing internet-based and lab studies used studies that, unlike the present study, were not true experiments. Instead, many were based on questionnaire or survey methods (Beldad, de Jong, & Steehouder, 2011; Epstein, Klinkenberg, Wiley, & McKinley, 2001; Gosling, Vazire, Srivastava, & John, 2004; Kays, Gathercoal, & Buhrow, 2012; Lewis, Watson, & White, 2009; Naus et al., 2009; Whittaker, 2007). Because true experiments allow the conclusions to be drawn that, first, a relationship between an independent and a dependent variable is specifically a cause and effect relationship, and, second, the direction of the causality, they afford the possibility of controlling, rather than just predicting, the effects of the independent variable. This gives true experimental research an added value that non-experimental research does not have. Therefore, it is important to show that not only non-experimental research, but also, true experiments can be conducted with just as much confidence in their validity when conducted on the internet as when conducted by a traditional delivery method.

A second gap in the present literature is that many studies used relatively small samples. Among the aforementioned studies, all but one (Gosling et al., 2004) used samples ranging only from 76 to 213 participants. We found a smaller number of comparisons between internet and traditional research in which the methodology of the studies being compared was experimental. However, in some of these studies the sample sizes were also small. For example, the samples used by McGraw et al. (2000) were 261, 128, and 81 participants, and those used by Vadillo and Matute (2009) were 20 and 75 participants. The present study used 1301 participants in the internet experiment. Thus, our conclusion that our internet experiment produced the same result as a traditional delivery experiment is less likely to be a chance result than if it had used fewer participants.

A third gap in the previous literature is that we found some studies that used the internet for an experimental methodology, but they compared their findings to previously conducted studies rather than by either randomly assigning participants to the internet and traditional delivery, or at least conducting the same study again in the traditional manner with a separate sample, but exactly as it had been conducted on the internet (Joinson, Paine, Buchanan, & Reips, 2008; Mitchell, Stanimirovic, Klein, & Vella-Brodrick, 2009; Vadillo & Matute, 2011). In the present study we filled this

gap by not just finding a similar experiment that had been previously conducted by a traditional delivery method and comparing it to our internet experiment, but rather by conducting the exact same experimental study that we had conducted on the internet again, but in a lab and by the traditional face to face delivery method, and then making a direct comparison between the two.

Finally, as mentioned above, one shortcoming we noticed in the current literature is that there have not been as many demonstrations of the equivalence between internet-based and traditional delivery of true experimental results as non-experimental results. Furthermore, we noticed that among the comparisons of non-experimental studies there have been mixed results (e.g., Mitchell et al., 2009; Naus et al., 2009; Vadillo & Matute, 2009). This raises the possibility that mixed results could also occur among the experimental studies, thus arguing for continued attempts to replicate the equivalency finding for experiments.

### 1.4. The motivation for the present study

The present study attempted to provide the needed further replication of the equivalency of internet-based and traditional delivery methods for experiments, as well as addressing a few other issues as well. For example, Vadillo and Matute (2009) pointed out that replicating an established experimental finding has the disadvantage that because it is well known, there is the possibility for demand characteristics to influence the participants. Therefore, in their follow up study (Vadillo & Matute, 2011) they rectified that shortcoming by demonstrating the equivalence of a less well known finding. However, in that study, as noted above, they did not do a direct comparison between the internet-based delivery method and an exactly similar traditional delivery method. They also used a relatively small sample size of only 130 participants. Finally, if a less well known effect helps reduce the probability of demand characteristics, then a completely unknown effect could help even more. Therefore, the present study attempted to (a) demonstrate the equivalence of an experimental manipulation of an entirely new and unknown effect, (b) to do so with a relatively large sample, and (c) to do so by making a direct comparison between the internet-based delivery and an exactly similar traditional delivery. To that end, our study made a direct test of the effect of the presence of experimenter supervision by making the materials and procedures for both delivery methods exactly the same except for the presence of an experimenter.

Another issue that the present study will address is the concern raised by Birnbaum (2004) that whereas it has been shown that internet-based research is equivalent to traditional delivery, whether it is actually *better* because of the larger sample sizes it affords has not been sufficiently demonstrated. The present study also addresses two methodological issues that are not always addressed in comparisons between internet-based and traditional research. First, as suggested by Hewson (2003), in order to remove duplicate responses from data collected from the internet, IP addresses, times, and dates were collected. Second, Birnbaum (2004) raised the concern that internet-based samples may be different from traditional samples in important ways. In order to address this concern in the present study, we collected demographic data which we used to show that our internet sample was in fact quite similar to the sample of college students we used for our traditional delivery.

## 2. Experiment 1

Experiment 1 was conducted in a laboratory under the supervision of an experimenter.

## 2.1. Methods

### 2.1.1. Participants

The participants were 180 university students, all of whom participated in the study in order to receive partial credit for their introductory psychology course.

### 2.1.2. Materials

In a previous study (Ryan, Cipko, & Rizzo, 2006), college students had rated the frighteningness and disgustingness of 44 different insects. These ratings were used to select eight insects, two of which were in each of the four categories resulting from crossing high and low frighteningness with high and low disgustingness. The insects selected for the four categories are shown in Fig. 1.

### 2.1.3. Procedure

Seventy-eight participants participated in the experiment in our lab, but did so by logging onto a computer and completing the experiment exactly as did the participants in the internet-based version of the experiment (described in Section 3), except that they were under the supervision of an experimenter. The other 102 participants completed a paper-and-pencil version of the procedure under the supervision of an experimenter as described below.

The experimenter recorded the gender of each subject, and then told them that they would see a series of pictures of insects and would have to rate their hostility towards them. The rating was described as the extent to which they either wanted to kill, or at least in some way get rid of, that particular insect. The option of just getting rid of the insect was included because, first, some people may be morally opposed to killing, and, second, the disgustingness factor may have prevented some individuals from wanting to kill the insects because they would not want to get close enough to the insects to risk being touched by them. In order to enable the participants to compare the insects with one another before giving their

ratings, all of the insects were initially presented together. Finally, the insects were presented again, this time one at a time in a randomized order, for the participant to rate them on a hostility scale ranging from 0 *not want to kill the insect at all* to 10 *greatest possible desire to kill the insect*.

## 2.2. Results and discussion

For Experiment 1, a three factor ANOVA was used to analyze the hostility ratings with gender as a between subjects factor and both disgustingness and frighteningness as within subjects factors. The results of the ANOVA are shown in Table 1.

We did not include the “experiment” factor (whether the participant completed the experiment by logging onto a computer or using paper-and-pencil) in the analysis shown in Table 1 because an initial analysis showed that the experiment factor did not produce any credible differences. The main effect of the experiment factor was not significant. Also, although the gender by frighteningness by experiment interaction was significant at  $p = .02$  if all of the non-significant main effects and interactions were left in the model, it was not significant ( $p = .25$ ) if they were removed.

As can be seen in Fig. 2, there were main effects of gender, frighteningness, and disgustingness. The females gave higher hostility ratings ( $M = 7.40$ ,  $S.E. = .169$ ) than the males ( $M = 6.21$ ,  $S.E. = .327$ ). The high frightening insects received higher hostility ratings ( $M = 7.17$ ,  $S.E. = .186$ ) than the low frightening insects ( $M = 6.44$ ,  $S.E. = .259$ ). The high disgusting insects received higher hostility ratings ( $M = 7.21$ ,  $S.E. = .182$ ) than the low disgusting insects ( $M = 6.40$ ,  $S.E. = .251$ ).

There were no significant interactions. However, as can again be seen in Fig. 2, gender had a tendency toward an interaction with frighteningness, with disgustingness, and with frighteningness by disgustingness. The effect of frighteningness was numerically greater for males than females. The effect of disgustingness was

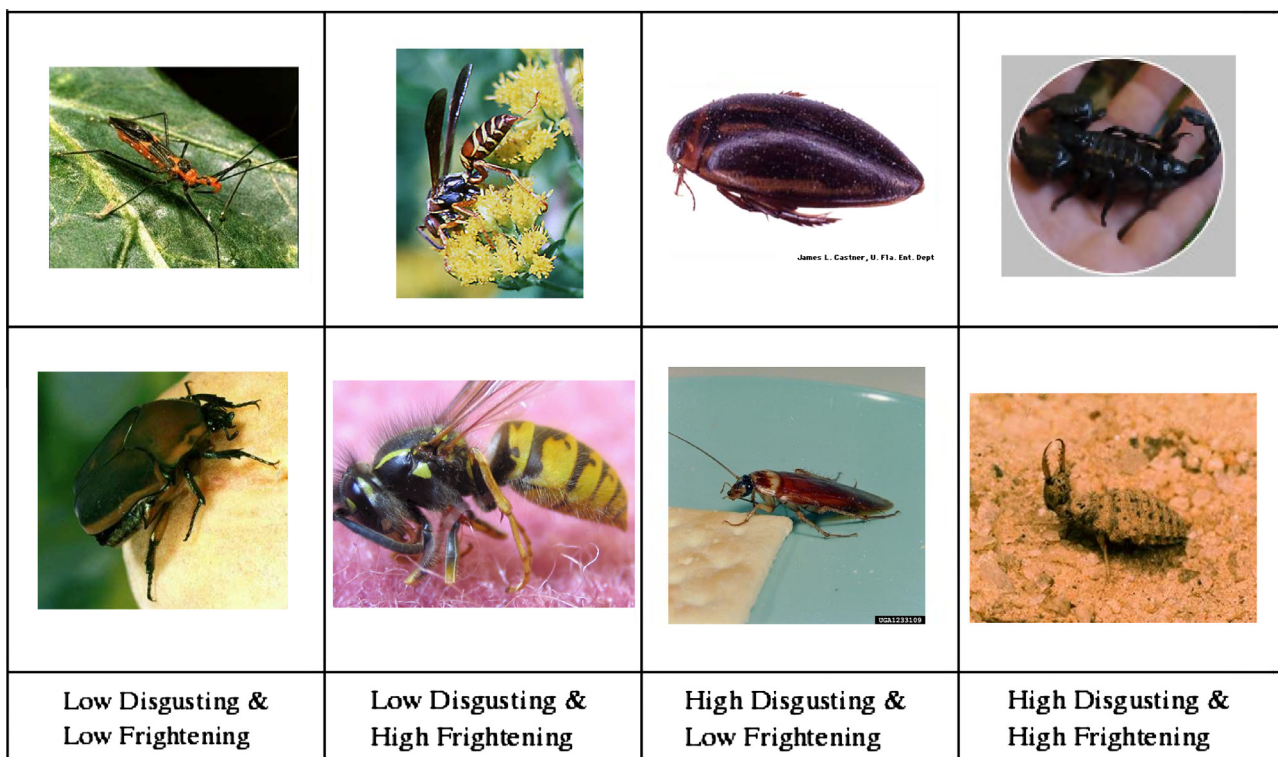
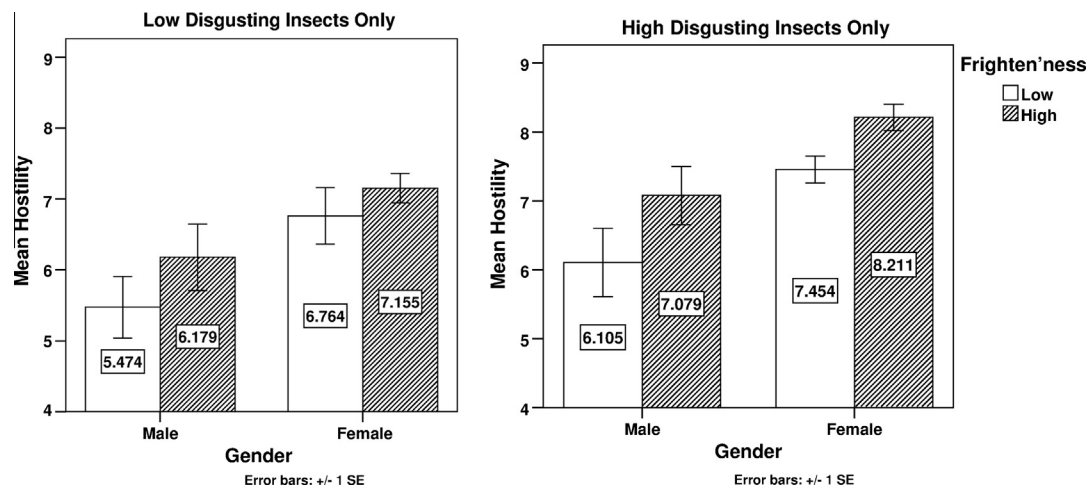


Fig. 1. Pictures of insects depicting low and high disgustingness crossed with low and high frighteningness.

**Table 1**

ANOVA on hostility for gender, frighteningness, and disgustingness for Experiment 1.

Source	SS	df	MS	F	p	$\eta^2$
Gender	168	1	167.6	10.31	.002	.0549
Error (gender)	2893	178	16.3			
Frighteningness	63	1	62.9	7.65	.006	.0412
Fright'ness * gender	3	1	2.7	.33	.566	.0020
Error (frighteningness)	1465	178	8.2			
Disgustingness	80	1	79.6	11.73	.001	.0621
Disgustingness * gender	.4	1	.4	.06	.807	.0003
Error (disgustingness)	1209	178	6.8			
Fright'ness * disgust'ness	2	1	2.4	.36	.547	.0017
Fright'ness * disgust'ness * gender	.2	1	.2	.03	.857	.0002
Error (fear * disgust)	1166	178	6.6			

**Fig. 2.** Mean hostility as a function of gender, frighteningness, and disgustingness in Experiment 1.

numerically greater for females than males. For the low disgusting insects, the effect of frighteningness was numerically greater for males than females, whereas for the high disgusting insects that interaction was smaller. Finally, although also not a significant interaction, the effect of frighteningness was numerically greater for the high disgusting insects than for the low disgusting insects.

### 3. Experiment 2

Experiment 2 was identical to Experiment 1 except that there was no supervision of the participants by an experimenter because the directions were presented on web pages that were found on the public internet. The study was originally posted on a website maintained by Hanover College (<http://psych.hanover.edu/research/exponnet.html>).

#### 3.1. Method

##### 3.1.1. Participants

The participants were 1351 respondents who came from all over the world, although most of them came from North America. After cleaning the data, as described below in the results and discussion section, 1301 respondents' data were included in the analysis.

##### 3.1.2. Materials

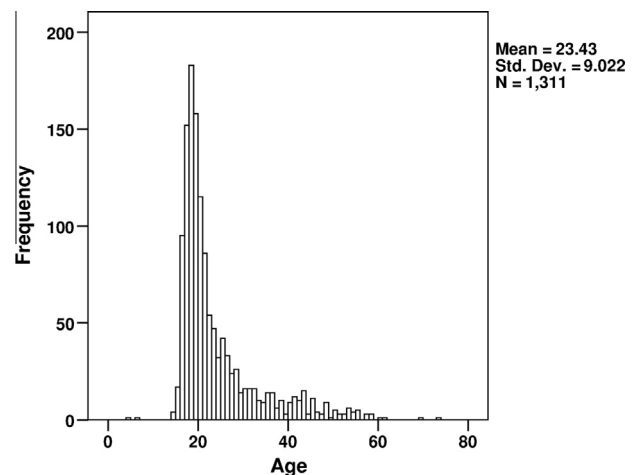
The materials were exactly the same as Experiment 1 except that they were presented on web pages, and more demographic questions than just the gender of the participant were asked (see Table 2).

#### 3.1.3. Procedure

The procedure was exactly the same as Experiment 1 except the instructions were presented on the web pages rather than given by an experimenter. Thus, there was no way to determine how engaged the participants were in the task, and there was no guarantee that the participants were following the instructions.

#### 3.2. Results and discussion

The data set was cleaned by removing responses that contained no ratings, and those that were obvious duplications. A response

**Fig. 3.** Frequency distribution of age of participants in Experiment 2.



**Table 2**

Percent of participants' responses to the demographic questions in Experiment 2 (percentages may not add to 100% due to some non responses).

Question	Possible response	Percent
Your gender?	Male	29.8
	Female	67.3
How old are you?	(see histogram of age in Fig. 3)	
Where are you from?	North America	76.5
	South America	1.9
	Europe	9.8
	Africa	.6
	Asia	3.2
	Australia	.9
	Other	4.5
What is your level of education?	Less than high school	13.7
	High school graduate	15.2
	Some college	50.7
	College graduate	8.4
	Partial advance degree training	2.9
	Advanced degree	6.0
Is English your native language?	Yes	83.0
	No	13.9
Why are you participating in this experiment?	Just for fun	12.3
	Looking for something to do	3.4
	Interested in psychology	21.9
	As an assignment for school	56.5
	Other	3.5
How did you find out about this experiment?	Just browsing the web	23.2
	Was looking for a psychology experiment	40.4
	Heard about it from someone	25.4
	Read about it on another website	8.2
Have you done psychology experiments on the web before?	Never	60.0
	One before this one	15.3
	Several before this one	16.7
	Many before this one	5.2
How much do you use the internet?	Seldom	3.5
	Occasionally	15.0
	Fairly often	26.1
	Every day	52.3
Do you think you will enjoy this experiment?	Yes	33.9
	No	1.2
	We will see	62.5

was considered a duplication if it came from the same IP address only a few seconds after the previous response and contained exactly the same demographic data and ratings.

The sample of participants that we recruited from the internet was very similar to the sample that participated in the lab-based experiment. As shown in Table 2, most of the participants appeared to be college students of traditional college age who were doing the study as an assignment for school. Also, they were predominantly female, most of whom had never done a psychology experiment before, and most of whom who use the internet every day, all of which we would expect to be characteristic of the psychology undergraduate students who participated in the lab-based study.

A three factor ANOVA was used to analyze the hostility ratings in the same way as in Experiment 1. The results of the ANOVA are shown in Table 3.

As in Experiment 1 there were main effects of frighteningness, disgustingness and gender. The females gave higher hostility

ratings ( $M = 6.80$ ,  $S.E. = .073$ ) than the males ( $M = 6.37$ ,  $S.E. = .110$ ). The high frightening insects received higher hostility ratings ( $M = 6.83$ ,  $S.E. = .076$ ) than the low frightening insects ( $M = 6.34$ ,  $S.E. = .073$ ). The high disgusting insects received higher hostility ratings ( $M = 7.15$ ,  $S.E. = .070$ ) than the low disgusting insects ( $M = 6.02$ ,  $S.E. = .077$ ). In addition to these main effects, all of the interactions were significant. As can be seen in Fig. 4, in all of the interactions involving gender, the patterns of the interactions were the same as in Experiment 1. The three way interaction was now especially clear. The males were more affected by frighteningness than the females, but only for the low disgusting insects.

Finally, unlike in Experiment 1, in which the effect of frighteningness was slightly, but not significantly, greater for the high disgusting insects, in Experiment 2 a positive effect of frighteningness was only present for the low disgusting insects. This was the only pattern that was different in Experiment 2 than it had been in Experiment 1.

## 4. General discussion

### 4.1. Findings

This comparison of experimental studies mirrored the comparisons of the many studies that were not experimental (e.g., Gosling et al., 2004; Lewis et al., 2009). It adds to the growing literature supporting the validity of using the internet as a means of collecting data in two ways. First, we obtained similar results across the two methods of administration, even though, in spite of the procedures for these studies being relatively simple, there was still the potential for participants to make responses that did not reflect their actual feelings about the insects. Because most of the participants were college students doing the study as an assignment for school, they could have entered responses only to fulfill the assignment, with little or no intention to make them reflect their actual feelings.

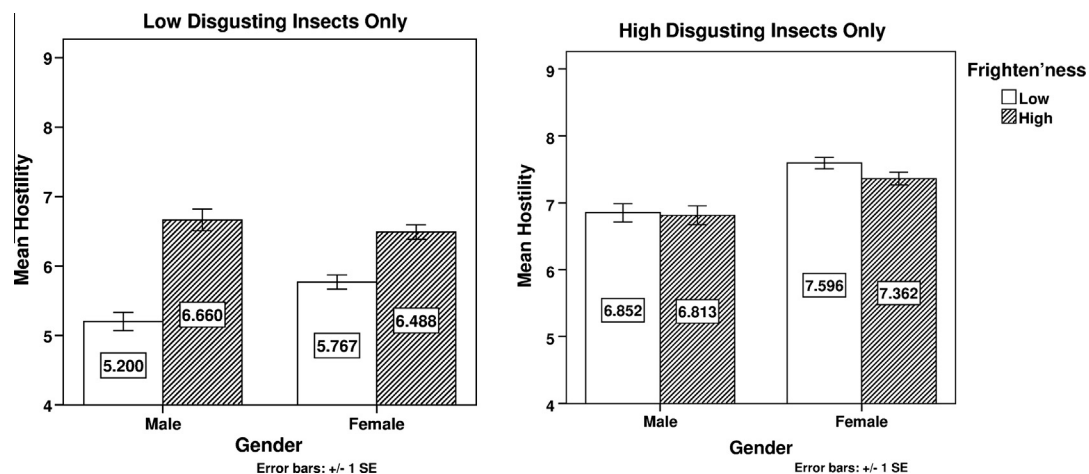
This concern about the motivation of the participants might be compared to the concern about the motivation of students taking assessment tests in higher education. Liu, Bridgeman, and Adler (2012) found that students who were told that their test scores would be used by faculty and potential employers to evaluate their academic ability performed better than control subjects who were told that their test scores would only be used for research purposes. In our comparison study the participants had no reason to believe that their responses were of concern to anyone other than the researchers. It is possible, therefore, that the amount of hostility that they reported would have been different had the participants believed that their responses would be used in some way that affected them personally. Nevertheless, the experiments that we conducted were not about the absolute amounts of hostility reported, but rather, about the relative amounts reported for the insects in the different categories. Therefore, it is not clear what reason participants could have had to have exaggerated or underestimated their hostility towards any particular category of insects dependent on their beliefs about how their data was to be used. Indeed, the data from our present study shows that an unsupervised web-based study produced results similar to those from the supervised lab study. This shows that a well constructed experimental study does not suffer from the lack of experimenter supervision when it is presented on the internet instead of in a lab.

The second way that our study adds support to the validity of using the internet as a means of collecting data is that the results show that data obtained from web research can be not only equal to, but superior to those obtained from traditional lab studies. Despite the lack of experimenter control, the web study, with its greater power, showed significant interaction effects where the lab study only showed patterns that did not reach significance.

**Table 3**

ANOVA on hostility for gender, frighteningness, and disgustingness for Experiment 2.

Source	SS	df	MS	F	p	$\eta^2$
Gender	211	1	211.0	10.90	.001	.0084
Error (gender)	25012	1293	19.3			
Frighteningness	256	1	256.0	48.92	<.001	.0365
Frighteningness * gender	60	1	60.0	11.43	.001	.0088
Error (frighteningness)	6757	1293	5.2			
Disgustingness	1401	1	1401.0	308.60	<.001	.1926
Disgustingness * gender	55	1	55.2	12.16	.001	.0093
Error (disgustingness)	5872	1293	4.5			
Fright'ness * disgust'ness	410	1	410.0	100.57	<.001	.0722
Fright'ness * disgust'ness * gender	18	1	18.0	4.31	.038	.0034
Er. (fright'ness * disgust'ness)	5265	1293	4.1			

**Fig. 4.** Mean hostility as a function of gender, frighteningness, and disgustingness in Experiment 2.

In the lab study all of the interaction patterns were non-significant, whereas in the web-based study, all of the interactions were significant. Among those interactions, all of the interactions that involved gender revealed the same pattern that had appeared at a non-significant level in the lab studies. Only the interaction between frighteningness and disgustingness had a pattern in the web-based study that was different from what it had been in the lab study.

In addition to supporting the validity of internet research, our study also provides further support for [Hewson's \(2003\)](#) recommendation to always collect IP addresses, times, and dates. It may be that part of the reason we were able to obtain the beneficial results of our large sample on the internet was because, having followed Hewson's recommendation, we were able to effectively clean the data of any duplicate responses.

#### 4.2. Possible shortcomings and recommendations for future research

It may be the case that the reason our internet-based and lab-based results were so similar is that the samples were similar. This suggests a possible shortcoming of our study. As a result of the similarity of the samples, in both the internet and the lab study the results may not be generalizable to populations other than college students participating in a study as an assignment. This also leads to the possibility that internet and lab results may not be as similar as they were in the present study if the samples had not been so similar. For example, the internet sample could have included a large proportion of people who were not college students participating as an assignment. Future research should examine this issue. Fortunately, use of the internet may provide

methods of recruiting samples that are more generalizable to broader populations ([Birnbau, 2004](#); [Skitka & Sargis, 2006](#)). Thus, future research could profitably explore the generalizability of the equivalence finding by purposely recruiting broader populations of subjects from the internet and comparing their performance to that of the more restricted population of college students usually used in psychological research.

Following the advice given above, however, would result in perpetuating another possible shortcoming of the present comparison study. Specifically, although the comparison was made between an experimental study conducted on the internet and the exact same study conducted in a lab, the participants in the studies were two separate samples rather than one large sample, the members of which had been randomly assigned to either the internet or the lab presentation. Thus, future research could also be conducted in which participants would be randomly assigned to either the internet or the lab administration of the same experimental study. Doing so, of course, would result in limiting the generalizability of the results as discussed above. Nevertheless, doing both kinds of future research could provide converging evidence about the validity of internet research.

A final and unavoidable shortcoming of the present study was that many of the possible disadvantages of web research were not applicable to this study and therefore, because they did not pose any problem, they were not addressed in our comparison. For example, the problem of increased dropout rate proposed by [Birnbau \(2004\)](#) was avoided in this study. In order for dropout to be a problem, the study must be composed of at least two separate parts, between which the participants have the opportunity to dropout. This study, however, consisted of only one part, and,

as long as the participant completed the study, dropout was impossible. Another potential disadvantage of web research could be ethical issues (Hewson, 2003; Skitka & Sargis, 2006). Again, however, informed consent was not a problem in this study because the first page of the website was the informed consent. By clicking “submit”, the participant was agreeing to the terms of the experiment and “signing” the informed consent. The Institutional Review Board at Kutztown University approved this method for obtaining informed consent.

#### 4.3. Conclusions

Overall, we conclude that the growing field of internet based research is a promising one. For the most part, any disadvantages can be kept to a minimum, and those that cannot are certainly outweighed by the potential advantages that this method can offer. The increased power of a web study enables researchers to provide evidence for subtle effects that are only suggested by trends in a smaller study, and the lack of supervision does not usually change the direction of effects or hide them with excess variability in the data. Thus, although we recommend future research as described above, we are optimistic about the future of internet based research.

#### Acknowledgments

Special thanks to Joseph Cipko, Alyssa Rizzo, Nikita Driscoll, Melissa Gilroy, Trisha Parker, Brittany Robison, Lisa Scala, Lora Seiverling, and Sarah Windfelder for help in data collection.

#### References

- Beldad, A., de Jong, M., & Steehouder, M. (2011). I trust not therefore it must be risky: Determinants of the perceived risks of disclosing personal data for e-government transactions. *Computers in Human Behavior*, 27(6), 2233–2242. <http://dx.doi.org/10.1016/j.chb.2011.07.002>.
- Birnbaum, M. H. (2004). Human research and data collection via the internet. *Annual Review of Psychology*, 55, 803–832.
- Carlbring, P., Brunt, S., Bohman, S., Austin, D., Richards, J., Öst, L., et al. (2007). Internet vs. paper and pencil administration of questionnaires commonly used in panic/agoraphobia research. *Computers in Human Behavior*, 23(3), 1421–1434. <http://dx.doi.org/10.1016/j.chb.2005.05.002>.
- Epstein, J., Klinkenberg, W. D., Wiley, D., & McKinley, L. (2001). Insuring sample equivalence across internet and paper-and-pencil assessments. *Computers in Human Behavior*, 17(3), 339–346. [http://dx.doi.org/10.1016/S0747-5632\(01\)00002-4](http://dx.doi.org/10.1016/S0747-5632(01)00002-4).
- Gosling, S. D., Vazire, S., Srivastava, S., & John, O. P. (2004). Should we trust web-based studies? A comparative analysis of six preconceptions about internet questionnaires. *American Psychologist*, 59(2), 93–104.
- Hewson, C. (2003). Conducting research on the internet. *Psychologist*, 16(6), 290–293.
- Joinson, A. N., Paine, C., Buchanan, T., & Reips, U. (2008). Measuring self-disclosure online: Blurring and non-response to sensitive items in web-based surveys. *Computers in Human Behavior*, 24(5), 2158–2171. <http://dx.doi.org/10.1016/j.chb.2007.10.005>.
- Kays, K., Gathercoal, K., & Buhrow, W. (2012). Does survey format influence self-disclosure on sensitive question items? *Computers in Human Behavior*, 28(1), 251–256. <http://dx.doi.org/10.1016/j.chb.2011.09.007>.
- Lewis, I., Watson, B., & White, K. M. (2009). Internet versus paper-and-pencil survey methods in psychological experiments: Equivalence testing of participant responses to health-related messages. *Australian Journal of Psychology*, 61(2), 107–116.
- Liu, O. L., Bridgeman, B., & Adler, R. M. (2012). Measuring learning outcomes in higher education: Motivation matters. *Educational Researcher*, 41(9), 352–362.
- McGraw, K. O., Tew, M. D., & Williams, J. E. (2000). The integrity of web-delivered experiments: Can you trust the data? *Psychological Science*, 11(6), 502–506.
- Mitchell, J., Stanimirovic, R., Klein, B., & Vella-Brodrick, D. (2009). A randomised controlled trial of a self-guided internet intervention promoting well-being. *Computers in Human Behavior*, 25(3), 749–760. <http://dx.doi.org/10.1016/j.chb.2009.02.003>.
- Naus, M. J., Philipp, L. M., & Samsi, M. (2009). From paper to pixels: A comparison of paper and computer formats in psychological assessment. *Computers in Human Behavior*, 25(1), 1–7. <http://dx.doi.org/10.1016/j.chb.2008.05.012>.
- Pettit, F. A. (2002). A comparison of world-wide web and paper-and-pencil personality questionnaires. *Behavior Research Methods, Instruments, & Computers*, 34(1), 50–54.
- Ryan, R. S., Cipko, J., & Rizzo, A. (2006, March). Characteristics of pictures of insects associated with fear and disgust. In *Poster presentation at the 77th annual meeting of the Eastern Psychological Association*, Baltimore, MD.
- Skitka, L. J., & Sargis, E. G. (2006). The internet as a psychological laboratory. *Annual Review of Psychology*, 57, 529–555.
- Vadillo, M. A., & Matute, H. (2009). Learning in virtual environments: Some discrepancies between laboratory and internet-based research on associative learning. *Computers in Human Behavior*, 25(2), 402–406. <http://dx.doi.org/10.1016/j.chb.2008.08.009>.
- Vadillo, M. A., & Matute, H. (2011). Further evidence on the validity of web-based research on associative learning: Augmentation in a predictive learning task. *Computers in Human Behavior*, 27(2), 750–754. <http://dx.doi.org/10.1016/j.chb.2010.10.020>.
- Whitaker, B. G. (2007). Internet-based attitude assessment: Does gender affect measurement equivalence? *Computers in Human Behavior*, 23(3), 1183–1194. <http://dx.doi.org/10.1016/j.chb.2004.11.016>.