Do lab effects transfer into the real-world? And should we care?

Petr Slovák
HCI Group
Vienna Uni of Technology
Austria
petr@igw.tuwien.ac.at

Paul Tennent
Mixed Reality Lab
University of Nottingham
United Kingdom
paul.tennent@nottingham.ac.uk

Geraldine Fitzpatrick
HCI Group
Vienna Uni of Technology
Austria
geraldine.fitzpatrick@igw.tuwien.ac.at

Introduction

We report on two of our own studies, each of which has built on a laboratory based finding and explored if and how the effects played out in everyday settings. In each, we found effects that in some ways validated the prior lab studies, but each also pointed to very different implications for HCI than those which were suggested by the initial lab work. By lab-based work we mean here empirical studies that are tightly controlled and aim to uncover causal relationships.

We would like to use these examples to open a discussion within RepliCHI on whether or not the transferring of such lab findings into the field is a specific type of replication that is especially important for HCI research, as (i) we generally do want our systems/findings to transfer into field settings; and (ii) it is plausible to expect similar results to the ones in our studies when transferring other lab effects into the real-world. We expect that further discussion of this topic could well complement the existing "into the wild" literature in HCI that now focuses more on open-ended, in-situ exploration (e.g., see [1, 2]).

Moving effects from the lab to the wild

The two studies reported below come from different projects, but each builds on a finding that was previously rigorously analysed in the lab and seemed to be potentially useful in HCI.

Study 1: Feeling connected by sharing heartbeat
The first study picked up on a psychology experiment by Janssen et. al. [4]. Their work showed that sharing heartbeat between people increased feelings of intimacy and social connectedness. This effect was shown not only by statistically significant differences in questionnaire responses, but also by measuring changes in carefully chosen non-verbal aspects of the interaction. Implications of such findings in HCI could be, for example, the use of such an effect to design systems supporting mutual affection in couples living apart or helping to create stronger ties within families and other social groups etc.

To explore the potential for real-world application of the observed effect, we[1] developed a simple technology probe package based on a heart rate monitor belt paired with a standard laptop through a Bluetooth connection. Ten such packages were distributed amongst 5 couples, encouraging each couple to use the probe in any way they wish over the period of two weeks. Furthermore, we invited additional pairs of friends into our lab, let them experience the probe in social scenarios and interviewed them about their reactions and ideas. We analysed the interview and usage data by qualitative means, identifying two distinct effects appearing across most of our sample, and suggested interpretations as to why the effects happen. The results were then presented at CHI'12 [9].

What was interesting in the study with regards to this workshop is that while our results confirmed the initial study in many ways, the implications for HCI were strikingly different. For example, our participants reported feeling much "closer" to another when talking about situations closely resembling the lab study. However, people often felt actually "too close" in these moments, describing the feeling, e.g., "as if a stranger in an elevator was keeping eye-contact for a long time". In other contexts, such as everyday use by the collocated couples, heart rate sharing did not have any effect at all (e.g. "as we are already close enough, this changes nothing"). Such results led us to suggest more specific contexts and situations where the effects of heart rate sharing could be used in positive ways, and better scope the potential applications of the original finding.

Study 2: Linking empathy to synchrony of bio-signals
In another study, we explored work done by Marci et. al. in psychotherapy and psychophysiology [6]. This work focuses on interaction between patient and therapist, and it links moments high in empathy to synchronised changes in skin conductance levels of the therapist/patient pair. For example, if the changes in skin conductance of patient and therapists were synchronised for a particular segment, external raters were more likely to rate such moments as high in empathy, it would also correspond to higher values in self-reported empathy etc.

Such a link could be of interest for HCI, e.g., as a novel indicator to embed in various affective computing systems, creating systems to support teaching of empathy for psychotherapy students, workplace etc. However, the original research was based on a very specific setting (therapy session) and participants with specific skills (therapist with many years of training to become highly empathetic). As such, we were interested to test how robustly the observed effects appear in the types of real-world settings that are of interest of HCI, but which are also often full of distractions and potential intervening variables that could not be controlled in real-world deployment.

[1] That is, Joris Janssen and two of the authors of this workshop paper (Geraldine Fitzpatrick and Petr Slovak)

We designed a study [10] in which pairs of friends discussed a topic of their choice in a public house during normal opening hours. The rationale was to test the robustness of the link in a setting that is more extreme (in terms of potential disruptions and intervening variables) than those needed for the potential applications. In other words, we argued that if the effect is robust enough to appear in a busy pub and for pairs of friends talking about any topics of importance, it is then more likely to appear also in a therapy students class, workplace setting or other potential application contexts.

Our results followed a similar pattern as in the first study: we have seen results that are in line with the original work, but the implications for HCI application of these have changed. For example, when we focused on interactions where participants were instructed to discuss their topics naturally, then thirty-seconds long video snapshots chosen purely on the basis of high synchrony showed also more empathy related non-verbal behaviour (as judged by independent raters). This fits with the prior lab results. However, we also found high synchrony in a condition where we asked one of the participants to ignore the other, i.e., where then little empathy could be expected. Such inconsistencies led us to suggest a re-interpretation of skin conductance synchrony – seeing it not only as an indicator of empathy, but potentially as a more general indicator of "mutual reactivity" (i.e., that people emotionally react to each other). Such reactivity then just happens to correspond well to empathy in the right contexts, such as therapy session or a discussion of two friends about an issue important for one of them. We were able able to further support this hypothesis through other psychological literature such as [5].

Summary
To summarise, each of the two studies have shown that the expected effects can appear also in an uncontrolled, real-world setting, and are thus potentially robust enough for HCI applications. However, and maybe more importantly, each also clarified and better scoped the potential implications of the original finding for HCI.

Do we see a general pattern?

Stepping away from the two examples here, it does seem that, at least for results in psychology, transfer of effects from the lab to the field is far from an obvious claim. For example, Mitchell [7] shows in a recent meta-review that many lab effects either become much weaker when tested in the field, or even change direction entirely. Mitchell also shows how the extent of such "failures to transfer" differs among various sub-fields of psychology.

Can this be expected also of lab-based research in HCI? To our knowledge, there is little literature on this within HCI so far. It is also not discussed in the recent "into the wild" literature, e.g., [1, 3], which seems to have a more "open orientation towards finding out what happens and drawing design principles or recommendations about users' reactions" [1].

We think it would be interesting to discuss in more depth how this focus on lab-to-field transfer of effects differs and complements the existing work on research "in the wild". One immediate difference is the focus, i.e., whether a well-understood lab effect is robust enough to also appear in more realistic (and thus messy) conditions. Among other things, this will probably also raise methodological questions, as the main aim of such work is to test if an effect appears (thus pointing to more quantitative, experimental work), but in a setting where one cannot

control many of the potentially intervening variables. See Oulavirta [8] for an initial discussion of similar topic in the context of Pervasive computing.

Conclusion

We intended to demonstrate that examining whether the results of lab studies appear robustly 'in the wild' may be a specific kind of replication research, and one that could be of significant benefit to the CHI community. Drawing on our two studies, we saw that while the core effect did translate, the implications about how it might be used within HCI were changed markedly. We referenced additional literature in psychology suggesting that such results might also be expected for other lab-based findings.

Short Bio

Petr Slovak is a researcher and PhD student at the HCI Group at Vienna University of Technology. Drawing on his background in both psychology and computer science, his research focuses on support for teaching of empathy in medical and therapeutic settings, with specific interest in the use of biosensors.

Paul Tennent is a researcher at the Mixed Reality Lab at the University of Nottingham. He has worked on a number of systems designed to support the transformation of complex system log data into accountable, queryable objects that can be used in qualitative analysis. More recently he has turned to the analysis and representation of biodata with a particular focus on television and the public understanding of science.

Geraldine Fitzpatrick is Professor at Vienna University of Technology in Austria and heads the Institute of Design and Assessment of Technology. She is interested in how we design pervasive, tangible and ubiquitous technologies to fit in with everyday contexts, with a particular interest in supporting social interaction and collaboration, and health and well being.

References

[1] Brown, B., Reeves, S., and Sherwood, S. Into the wild: Challenges and Opportunities for Field Trial Methods. In *CHI '11*, ACM Press (May 2011), 1657.

[2] Consolvo, S., Harrison, B., Smith, I., Chen, M. Y., Everitt, K., Froehlich, J., and Landay, J. A. Conducting In Situ Evaluations for and With Ubiquitous Computing Technologies. *International Journal of Human-Computer Interaction 22*, 1-2 (Apr. 2007), 103–118.

[3] Hutchinson, H., Hansen, H., Roussel, N., Eiderbäck, B., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., and Evans, H. Technology probes: inspiring design for and with families. In *CHI '03*, ACM Press (New York, New York, USA, 2003), 17—-24.

[4] Janssen, J. H., Bailenson, J. N., IJsselsteijn, W. A., and Westerink, J. H. Intimate Heartbeats: Opportunities for Affective Communication Technology. *IEEE Transactions on Affective Computing 1*, 2 (July 2010), 72–80.

[5] Levenson, R. W., and Ruef, A. M. Physiological aspects of emotional knowledge and rapport. In *Empathic accuracy*. 1997.

[6] Marci, C. D., and Orr, S. P. The effect of emotional distance on psychophysiologic concordance and perceived empathy between patient and interviewer. *Applied psychophysiology and biofeedback 31*, 2 (June 2006), 115–28.

[7] Mitchell, G. Revisiting Truth or Triviality: The External Validity of Research in the Psychological

Laboratory. *Perspectives on Psychological Science 7*, 2 (Mar. 2012), 109–117.

[8] Oulasvirta, A. Rethinking Experimental Designs for Field Evaluations. *IEEE Pervasive Computing 11*, 4 (Oct. 2012), 60–67.

[9] Slovák, P., Janssen, J., and Fitzpatrick, G. Understanding heart rate sharing: towards unpacking physiosocial space. In *CHI '12* (2012), 859–868.

[10] Slovak, P., Tennent, P., and Fitzpatrick, G. Exploring physiological synchrony in everyday meaningful interactions. *Submitted for Interact'13* (2013).