Resolving Ecological Validity: A Position Paper for CHI 2015 Workshop on 'Ecological Perspectives in HCI'

Huiyuan Zhou, Kirstie Hawkey, Derek Reilly
Dalhousie University
Halifax, NS, Canada
{hzhou, hawkey, reilly}@cs.dal.ca

Lessons from History

The controversy between "ecologically-valid" approaches and the traditional laboratory approach has a long history in psychology. To avoid rediscovering or repeating the past, it is useful to look back on the history of ecological validity and frame our understanding correspondingly. Brunswik introduced the term "ecological validity" in the 1940s to mean retaining the integrity of environmental causal factors in order to achieve representative design as opposed to systematic design of experiments [4]. Similar debates over the merits of laboratory and field studies occurred in 1960s. Since Neisser's radical criticism of the laboratory approach in cognitive psychology in 1978, a new round of on-going debate has extended to other areas of psychology and related sciences [8]. Over time, a set of common understandings regarding ecological validity have emerged. First, ecological validity is generally defined as the extent to which study methods, materials and setting can approximate the real world. Second, the issue of ecological validity stems from the inherent trade-off between strict controls for causal inference and keeping the essence of context for generalization. Thus laboratory and ecological methods are two points on a continuous spectrum, each with its own strengths and weaknesses. Finally, the core of ecologically valid study design doesn't lie in the physical study environment. Lab studies with artificial tasks and materials have generated universally valid principles while field studies that miss key factors or fail to control irrelevant factors often don't have high ecological validity. Hence naturalistic study methodologies are often desirable but not indispensable in the quest for ecological validity.

The Role of Generalizability

There seems to be clear consensus that ecological validity and generalizability are closely related, but less consensus on what the relationship actually is. For example, Schmuckler [12] identified three dimensions of ecological validity: natural research setting, important and relevant stimuli, and representative task, behavior or response required of the participant. This definition implicitly equates ecological validity to the generalizability of experimental behavior to real-world behavior. However, Kvavilashvili and Ellis [8] explicitly

separate ecological validity into two components: representativeness (naturalness or artificiality of the experimental setting, material and task) and generalizability (transfer of study results to explain similar real-world circumstances). Generalizability was argued to be more critical than representativeness. While Bracht and Glass [1] hold the view that ecological validity is generalization across settings and thus a part of external validity (generalization to other settings and other people), others [5] assert that ecological validity is independent of external validity.

We propose that ecological validity is not a new genre of validity but rather a constructive concept that encourages employing representative setting, stimulus, and task to help achieve generalizability from a particular study to practice. Ecological validity underscores the role of context (e.g. location, time, identity), which is important in framing circumstantial behavior. It doesn't ensure generalizability unless it captures essentially relevant aspects of the real-world for the question under investigation. Nor does it contribute to the validity of the overall study unless combined with careful consideration (if not outright control) of confounding variables that occur in rich contexts. Furthermore, Lund [9] further differentiates generalization from the current study to practical situations and from the current study to the existing knowledge space. The former has more relevance to applied research and the latter has more relevance to basic research. In this vein, ecological validity, while relevant to basic research, is perhaps more central to applied research.

Difficulty of Adopting Ecology in HCI

Ecological validity is especially important when the inherent nature of the laboratory setting (e.g., trustworthy, formal, quiet) fundamentally changes people's behavior (e.g., by becoming more tolerant to risk) [14], or when the topic under investigation is highly context-dependent and/or understanding how systems are used in dynamic, uncontrolled environments is valuable [2]. Nonetheless, there are several general difficulties for the HCI community when applying "ecological approaches" to research:

- *Ambiguity of standards*. For any given study it is usually difficult to determine the critical components of the study setting, and tasks that characterize the real world circumstances. As a result, it is hard to judge whether the study simulation successfully captures those components.
- *Deployment of immature technology*. HCI research that requires bulky equipment to be attached to people (for example) will limit the feasibility to deploy in study settings outside the lab.
- *High cost of field evaluation*. Certain behaviors or interactions are less frequent and highly unforeseeable. Without proper manipulation or control, research could be economically inefficient and unsustainable.
- *Challenges collecting data in the field*. It is difficult for researchers to directly observe behaviors in the field, especially subtle interactions with small mobile devices. Techniques like software logging and experience sampling will miss some contextual details.

In the meantime, challenges to conduct ecologically valid research particular to specific research areas also exist. For example, in the field of usable security and

privacy, the following issues (some of which may be applicable to other areas) are frequently discussed:

- *Recruitment bias*. Due to the sensitivity of the topic of security and privacy, research that requires participants to use real data will likely exclude the participation of people who are highly concerned about security and unwilling to reveal their data and thus yields a biased sample [10].
- *Lack of motivation*. When participating in an experimental simulation, participants will hardly invest themselves to value task data or be responsible for their demonstrated behavior in the manner they do in the real world because those actions won't result in real consequences.
- *Ethical issues*. The secondary nature of security tasks often require researchers to conceal the real purpose of the study. In extreme cases, even with ethics approval, the anger and embarrassment of participants being deceived (by phishing email for example) still raises ethical concerns and can even lead to legal action [7].
- *Legal issues*. Wearable recording devices that disappear into the background could introduce legal concerns. For instance, participants in a life-logging study were forbidden from wearing cameras in many locations [6]. Google Glass has also been banned from a number of places (e.g. hospitals, cinemas, banks).

Potential Solutions

Mitigation approaches for current limitations to increasing ecological validity include adapting laboratory and field study methods, and embracing state-of-the-art technology:

- *Reduced reality in the lab*. By focusing on key factors and using tools creatively, lab studies could also capture the part of reality that matters. For instance, Cho [3] used treadmill and looped track to simulate performing tasks in motion.
- *Role playing and personalization*. Role playing has been demonstrated to be helpful in engaging participants in study tasks and develop a sense of responsibility [15]. Using personal information to tailor scenarios to relevant and plausible ones for each individual would also encourage them to relate real life experiences, and hopefully provide more ecologically valid feedback.
- *Wizard of OZ in the field*. This method is useful to explore new technologies that are hard to deploy in the field. The authors have conducted a study with a Wizard of OZ prototype in the field. Participants were able to actively use the "fake" prototype and relate the feedback to the dynamic context [11].
- *Post-hoc design*. Such methods first log users' natural behavior unknowingly and then ask for their consent to access and use the data afterwards [12].
- *Wearable technology*. Wearable technologies allow the collection of everyday data in a more continuous and less intrusive way. Incorporating more wearable computers into HCI research will be another future trend.
- *Virtual reality*. VR has long been recognized as permitting stimulus environments that potentially carry good ecological validity [13]. With the advancement of immersive VR technology (e.g., Oculus Rift), the enhanced feeling of presence and new ways to interact with the environment will open up a range of possibilities for the HCI community.

Conclusions

Ecological validity serves as a somewhat confusing yet still enlightening concept, not so much guaranteeing study "validity" as highlighting the importance of context to guide sound study design. Despite the practical challenges of applying "ecological approaches" in HCI research, we believe there are real benefits to adapting traditional methodologies in more ecological orientations. Moreover, we believe that being aware of trade-offs and looking beyond ecological validity by itself toward a more holistic view of overall study validity will be most beneficial.

References

[1] Bracht, G. H., & Glass, G. V. (1968). The external validity of experiments. *American educational research journal*, 437-474.

[2] Carter, S., Mankoff, J., Klemmer, S. R., & Matthews, T. (2008). Exiting the cleanroom: On ecological validity and ubiquitous computing. *HCI*, 23(1), 47-99.

[3] Cho, G. (Ed.). (2010). *Smart clothing: technology and applications* (pp.237-240). CRC Press.

[4] Dhami, M. K., Hertwig, R., & Hoffrage, U. (2004). The role of representative design in an ecological approach to cognition. *Psychological bulletin*, 130(6), 959.

[5] Ecological validity. (n.d.). In Wikipedia. Retrieved Dec. 24, 2014, from http://en.wikipedia.org/wiki/Ecological_validity.

[6] Hoyle, R., Templeman, R., Armes, S., Anthony, D., Crandall, D., & Kapadia, A. (2014, September). Privacy behaviors of lifeloggers using wearable cameras. In *Proc. of Ubicomp 2014* (pp. 571-582). ACM.

[7] Jagatic, T. N., Johnson, N. A., Jakobsson, M., & Menczer, F. (2007). Social phishing. *Communications of the ACM*, 50(10), 94-100.

[8] Kvavilashvili, L., & Ellis, J. (2004). Ecological validity and the real-life/laboratory controversy in memory research: A critical and historical review. *History & Philosophy of Psychology*, 6, 59-80.

[9] Lund, T. (2013). Kinds of Generalizations in Educational and Psychological Research. *Scandinavian Journal of Educational Research*, 57(4), 445-456.

[10] Patrick A. (2007). Commentary on Research on New Security Indicators -essay. http://www.andrewpatrick.ca/essays/commentary-on-research-on-new-security-indicators

[11] Reilly, D., Dearman, D., Welsman-Dinelle, M., and Inkpen, K. 2005. Evaluating Early Prototypes in Context: Trade-offs, Challenges, and Successes. *IEEE Pervasive Computing* 4, 4 (Oct. 2005), 42-50.

[12] Schmuckler, M. A. (2001). What is ecological validity? A dimensional analysis. *Infancy*, 2(4), 419-436.

[13] Schultheis, M. T., & Rizzo, A. A. (2001). The application of virtual reality technology in rehabilitation. *Rehabilitation Psychology*, 46(3), 296.

[14] Sotirakopoulos, A., Hawkey, K., & Beznosov, K. (2011, July). On the challenges in usable security lab studies: Lessons learned from replicating a study on SSL warnings. *Proc. SOUPS 2011* (p. 3). ACM.

[15] Vaniea, K., Bauer, L., Cranor, L. F., & Reiter, M. K. (2012, July). Studying access-control usability in the lab: lessons learned from four studies. *Workshop on Learning from Authoritative Security Experiment Results* (pp. 31-40). ACM.