Human Factors and Simulation, Vol. 30, 2022, 11–17
https://doi.org/10.54941/ahfe1001486

The Importance of Assessing Both Expert and Non-Expert Populations to Inform Expert Performance

Stephen R. Mitroff[1], Emma M. Siritzky[1], Samoni Nag[1], Patrick H. Cox[1], Chloe Callahan-Flintoft[2], Andrew Tweedell[2], Dwight J. Kravitz[1], and Kelvin S. Oie[2]

[1]The George Washington University Washington, DC 20052, USA
[2]DEVCOM Army Research Laboratory Aberdeen Proving Ground, MD 21005, USA

ABSTRACT

Human factors research is best accomplished when basic science theories, often derived from studying non-expert populations, and applied principles, often derived from studying expert populations, inform each other. Mechanistic cognitive theories gain when informed by practical applications, and operational implementations are best optimized by understanding the basic nature of human operators. An interplay between research involving data from expert and non-expert populations holds great promise, but is often thwarted by information from each side not flowing to the other. The argument here is that both types of data are fundamentally important, and explicit efforts should bring them together into unified and integrated research programs. Moreover, effectively understanding expert performance requires assessing non-expert populations.

Keywords: Applied research, Non-expert populations, Translating basic science

INTRODUCTION

Realizing the benefits of human factors research requires basic science theory and applied research in operational environments to work in tandem, each informing the other (e.g., Treviño et al., 2021). On one hand, incorporating information from practical applications provides insight into mechanistic theories about cognitive processing. On the other hand, the application of human knowledge in specific implementations requires understanding the nature of the human operators that will be using those very implementations. This interplay holds great promise, but is too often thwarted by information from each side not flowing to the other. Basic science researchers are often reluctant to accept findings from complex environments and from what is typically a relatively small number of highly-specialized participants. Similarly, industry decision makers can be reluctant to believe in the applicability of results from simplified testing environments using non-expert research participants with non-operational stimuli. The argument we put forward here is that both types of data are fundamentally important to understanding expert performance, and explicit efforts should be made to

© 2022. Published by AHFE Open Access. All rights reserved.

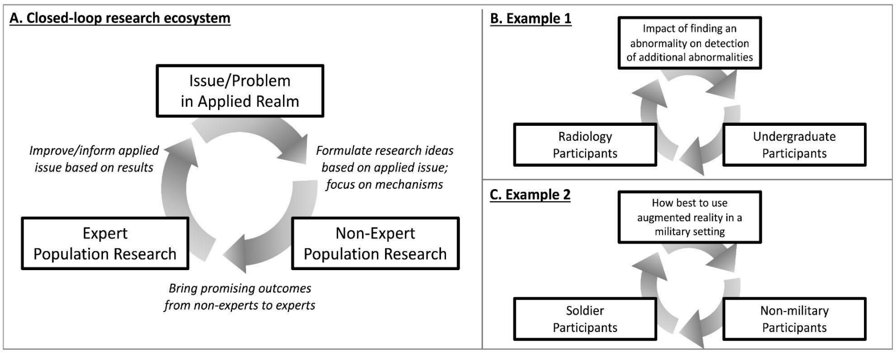

Figure 1: (A) Depiction of a closed-loop research ecosystem to use expert and non-expert participants, with examples for (B) radiology and (C) for military research.

bring them together into unified and integrated research programs. Moreover, effectively understanding expert performance requires assessing non-expert populations.

CLOSED-LOOPED ECOSYSTEM

It is often critically important to understand how operators (e.g., aviation security officers, radiologists, military personnel) perform in their professional setting. Many professions have life-or-death consequences, making it vital to examine exactly what factors affect performance. While technological aids can potentially improve professional performance (e.g., the use of computer-aided detection in radiology; Kunar, 2022; Lehman et al., 2015), at the end of the day a human operator must act. Given the fundamental role of the human in the process, extensive research across a range of fields (e.g., radiology, military operations, air traffic control, cytology) has explored a breadth of factors that can improve, or hinder, operators' success. However, the majority of these research endeavors ultimately hit the same roadblock—it is practically difficult to test specialized operators. Expert participants can be hard to gain access to, have limited availability, and sometimes there just are not enough of them to conduct the needed research. Beyond such practical hurdles, it is also, arguably, theoretically limiting to solely assess a specialized, expert population as it may obfuscate meaningful factors.

Given the potential limitations of solely assessing experts, non-expert populations can provide a much-needed resource. Specifically, it can be highly useful to create a closed-loop ecosystem wherein an idea rooted in an applied realm (e.g., radiologists are more likely to miss an abnormality if they just found another abnormality) is explored with non-experts (e.g., undergraduate students) to affordably and extensively explore a number of theoretical and mechanistic possibilities (see Figure 1). Then, the most promising candidate outcomes can be brought back to the expert population for further testing. With such a process, researchers can explore possible ideas with the more accessible population and then only assess the specialized population with vetted research paradigms and well-honed hypotheses.

As an example of this closed-loop ecosystem, military researchers often draw upon the wealth of basic research conducted with non-expert samples. Military goals can be informed by leveraging basic research to better understand how complex mechanisms of attention operate and affect performance. For example, prior military research has relied on data from non-military personnel to examine scene information extraction, as well as general situational awareness, to make recommendations for developing new technologies such as augmented reality used in the field (e.g., Larkin et al., 2020). To align this work with the proposed closed-loop ecosystem, the next critical step would be to test these recommendations on an expert population that would actually use such technologies in practice; in this case, Soldiers. In short, the overarching goal is to use the research with non-experts to narrow the scope of exploration and then test the most promising results with the relevant experts to improve its applicability in the field.

IMPORTANCE OF ASSESSING NON-EXPERTS

While such closed-looped ecosystem research practices offer a way to best use available resources, the argument here is that it is *necessary* to assess non-experts to fully understand expert performance. That is, even if researchers have full access to a large number of experts, they still need to test non-experts. Specifically, assessing non-experts allows for quantifying fundamentally important factors, such as strategic versus perceptual drivers of performance, or the time course of learning. Many of the potential gains in the applied sphere come from selecting the best people to train into becoming experts; without non-expert performance it is impossible to know how to enact that selection process or to divorce the effects of extensive practice from expertise with the specific operational environment.

As an example of the benefits of assessing non-experts, consider the process of employee recruitment. Factors such as spatial ability have been shown to play a critical role in developing aptitude in a variety of STEM fields and could have important implications for recruitment (e.g., Wai et al., 2009). Non-experts, through not only their availability but specifically their lack of training, provide researchers the ability to tease apart mechanisms of spatial attention and working memory in the visual system (Xu & Franconeri, 2015). These results can inform not only which individuals should be targeted for recruitment, but also what training efforts may serve to widen the potential pool of applicants.

As another example, in the field of chemistry, experts have been shown to not only employ spatial reasoning strategies, but also to heavily rely on algorithmic problem solving which develops through experience (Stieff & Raje, 2010). Accordingly, such strategies are unavailable to non-experts, making non-experts great candidates to understand mechanistic accounts of the development of such spatial abilities. These results show that only through studying both expert and non-expert populations can researchers get a clear picture of the skills needed to enter the field of chemistry, as well as

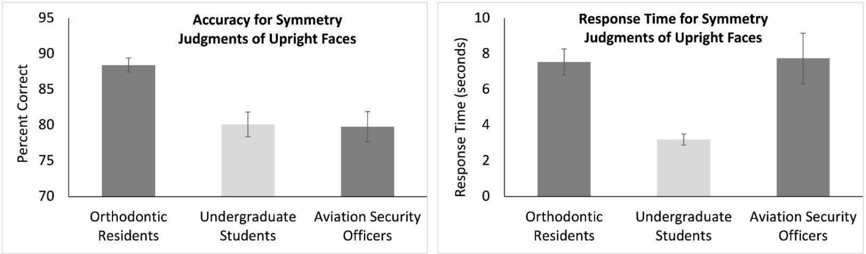

Figure 2: Reproduction of data from Jackson et al. (2013). Orthodontists were more accurate than non-experts (undergraduates) and experts from another field (aviation security officers), while being slower than the non-experts but not the other expert group.

those skills ultimately used after entry. This example is specific to chemistry, but the logic can play out for any number of expert fields.

Similarly, studies that employ expert and non-expert comparisons have provided insight into the generalizable advantage that experts with professional training can have beyond their domain-specific tasks. In a study comparing visual search performance between professional aviation security officers to that of nonprofessional undergraduate students, the use of a domain-general visual search task for both groups revealed fundamental differences in performance, potentially rooted in increased practice and expertise, that benefited the professional searchers (Biggs et al., 2013). These findings also revealed the generalizability of domain-specific training to universal visual search tasks, highlighting the necessity of non-expert comparisons in studying the scope of learning and the decay rate of practice benefits for those with professional training.

It can also be informative to not only study expert and non-expert participants, but also multiple groups of experts from different domains. For example, one study explored if orthodontists have heightened sensitivity to facial symmetry (Jackson et al, 2013) given that orthodontists assess facial symmetry as part of patient care, with asymmetries being signs of potential health problems. Orthodontic residents, non-expert undergraduate students, and professional aviation security officers all completed a facial symmetry judgment task with both upright and inverted faces and a non-facial symmetry judgment task with dot patterns. As reproduced in Figure 2, the orthodontist participants were significantly more accurate at both upright and inverted facial symmetry judgments compared to the undergraduate students and the aviation security officers (Jackson et al., 2013). Critically, this most likely did not arise from a simple speed-accuracy tradeoff or change in effort and motivation since both the orthodontists and aviation security officers were significantly slower than the undergraduate students. Likewise, there was no difference between the groups for the non-facial symmetry judgment task. This study highlights how insights can be gained into expert performance through targeted comparisons with other participant groups who do not have specific domain knowledge.

IMPLEMENTATION HURDLES

To actually accomplish the process suggested here—a closed-loop ecosystem wherein a research question is explored cyclically with both expert and non-expert populations—all relevant players of the research enterprise have to be on board. Academic researchers must be willing and excited to engage with expert populations and understand the intricacies of the mechanisms in practice, and the practitioners must be open to learning from simplified studies and providing insights to inform theories. Critically, these two camps must be receptive at every stage of the process, including generating research ideas, conducting the research, reviewing the generated products, and reviewing grant submissions. If academic researchers are reluctant to put value on applied work or if practitioners and applied researchers are reluctant to accept findings from non-expert populations, then this system falls apart as progress can be blocked by "gatekeepers" at any stage.

To illustrate the precariousness of the process and how it can be easily derailed, consider a recent research endeavor from our team. In collaboration with a radiologist, a research program was designed to address an important issue in breast imaging—the impact of using 3D technology for breast cancer detection (Adamo et al., 2018). For decades, the tool of choice for breast imaging has been mammography—a 2D imaging technique (e.g., Bleyer & Welch, 2012). However, the field is shifting towards using tomosynthesis—a segmented 3D imaging technique. The key advantages of tomosynthesis are that radiologists tend to make fewer false alarms (e.g., Durand et al., 2015; Friedewald et al., 2014) and detect more cancers (e.g., Ciatto et al., 2013). However, there is a major downside in that it takes significantly longer to evaluate a patient with tomosynthesis than with mammography alone (e.g., Bernardi et al., 212, Dang et al., 2014). This added time is not merely an inconvenience as it creates an implementation problem that has placed extreme pressure on the workload of radiologists.

Our research team sought to understand what factors slow search in tomosynthesis to identify ways to speed up the process while maintaining the accuracy benefits (Adamo et al., 2018). The first step of the closed-loop ecosystem (Figure 1) was identifying the real-world problem that practitioners face (i.e., added time of conducting search in tomosynthesis), and the second step was designing a protocol that could be used with both experts and non-experts. This was successfully done, as a simplified research protocol was created that compared 2D and 3D search in both radiologists and university students, and both populations demonstrated reduced false alarms, increased target detections, and slower response times in 3D search compared to 2D search (Adamo et al., 2018). The next step was to use the tool with easy-to-access undergraduate students to assess a range of factors. The most promising outcomes would then be tested in the more limited expert population (radiologists). The timing was ideal as the National Cancer Institute put out a call for grant proposals that was well-aligned. At this stage of this particular example, there were engaged academic and practitioner researchers and a funding agency looking to support such work. However, the grant panel of practitioners was not on board, producing comments such as

"*The use of simplified stimuli and unrealistic target probabilities might be problematic for translating the results from the proposed laboratory research to clinical practice*" and "*it is questionable whether results from novice participants will bear direct relevance and be translatable to tomosynthesis in radiologists*." Surely there were other aspects that affected funded decisions, but such reviewer comments highlight how easily this precarious balance of an academic-practitioner partnership can be disrupted.

It is also worth highlighting a practical hurdle that may arise within this closed-loop system stemming from the opaque definition of expertise. Even if publication and grant proposal reviewers agree on the need to assess both expert and non-expert participants, the next step would be to gain consensus on the criteria or definition of the samples. This taxonomy may not be as straightforward as imagined. For example, a meta-analysis of 91 unique studies from sport psychology research (Swann et al., 2015) found eight different categorical definitions for "elite athletes." These generally involve sport-specific measures, quality or quantity of training or experience, or attained level of competition (e.g. international or professional competition). Although the extreme positions of the nature versus nurture debate in expertise are largely dismissed (Ackerman 2014), the relationships between the previously mentioned metrics for expertise are often tenuous. For example, the amount of time spent in deliberate practice or high scores on laboratory performance metrics may never translate to high enough 'on-field' performance to reach international or professional levels for a given individual. This disconnect may also contribute to applied researchers and practitioners being reluctant to accept some basic research methodologies. This example is obviously quite specific, but it highlights the many layers of hurdles that stand in the way of effectively combining expert and non-expert populations to both inform theories and practice.

CONCLUSION

While there has been an, at times, adversarial relationship between research practices that use expert versus non-expert participants, the current proposal is that embracing both is vital for fully understanding the nature of expert performance. Basic research often relies on non-expert populations both because of the ease in data collection, and because the "non-expert" is essential for understanding how human behavior develops and operates without extensive training in application-specific tasks. However, if one goal in understanding human behavior is to draw inferences on how an individual would perform at a given task or make recommendations to aid operations or application development, without being tethered to some applied goal or real-world applicability, basic research can lack purpose. On the applied end of the research spectrum, however, it is easy to inadvertently get lost in nuance or a highly-specialized space where, without basic research to map out fundamental cognitive principles, one could study an expert group for years without generating generalized results that will apply to the next group of experts. The aim of this paper, therefore, is to illustrate how embracing both expert and non-expert participant research is essential, and especially powerful when the two are explored collaboratively.

ACKNOWLEDGMENT

Research supported was provided by US Army Research Office grant #W911NF-16-1-0274 and US Army Research Laboratory Cooperative Agreements #W911NF-19-2-0260 & #W911NF-21-2-0179. P.H.C. was supported by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at The George Washington University, administered by Oak Ridge Institute for Science and Education through an interagency agreement between the U.S. Department of Energy and the Office of the Director of National Intelligence.

REFERENCES

Ackerman, P. (2014) Nonsense, common sense, and science of expert performance: Talent and individual differences. *Intelligence*, 45, 6-17. https://doi.org/10.1016/j.intell.2013.04.009

Biggs, A. T., Cain, M. S., Clark, K., Darling, E. F., & Mitroff, S. R. (2013). Assessing visual search performance differences between Transportation Security Administration Officers and nonprofessional visual searchers. *Visual Cognition*, *21*(3), 330–352. https://doi.org/10.1080/13506285.2013.790329

Jackson, T. H., Clark, K., & Mitroff, S. R. (2013). Enhanced facial symmetry assessment in orthodontists. *Visual Cognition, 21*(7), 838-852. DOI:10.1080/13506285.2013.832450

Kunar, M.A. (2022). The optimal use of computer aided detection to find low prevalence cancers. *Cognitive Research: Principles & Implications*, *7*, 13. https://doi.org/10.1186/s41235-022-00361-1

Larkin, G. B., Geuss, M. N., Yu, A., Callahan-Flintoft, C., Rexwinkle, J., Hung, C. P., & Lance, B. J. (2020, July). Human Performance with Complex Technology: How Visual Cognition Is Critical to Enhanced Performance with Aided Target Recognition (AiTR). In *International Conference on Applied Human Factors and Ergonomics* (pp. 134–140). Springer, Cham.

Lehman, C. D., Wellman, R. D., Buist, D. S., Kerlikowske, K., Tosteson, A. N., Miglioretti, D. L., & Breast Cancer Surveillance Consortium. (2015). Diagnostic accuracy of digital screening mammography with and without computer-aided detection. *JAMA internal medicine*, *175*(11), 1828–1837.

Stieff, M., & Raje, S. (2010). Expert algorithmic and imagistic problem solving strategies in advanced chemistry. *Spatial Cognition & Computation*, *10*(1), 53–81.

Swann, C., Moran, A., and Piggott, D. (2015). Defining elite athletes: Issues in the study of expert performance in sport psychology. *Psychology of Sport and Exercise*, 16, 3–14. https://doi.org/10.1016/j.psychsport.2014.07.004

Treviño, M., Birdsong, G., Carrigan, A., Choyke, P., Drew, T., Eckstein, M., ... & Wolfe, J. M. (2021). Advancing Research on Medical Image Perception by Strengthening Multidisciplinary Collaboration. *JNCI Cancer Spectrum*.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*(4), 817.

Xu, Y., & Franconeri, S. L. (2015). Capacity for visual features in mental rotation. *Psychological Science*, *26*(8), 1241–1251.