agent's decisions, and testing a technique for leveraging user assessments to test similar agent predictions [9].

I plan to use qualitative studies to explore RQs 1.2, 2.3, 3.1, and 3.2, which I hope to complete in the next 18 months. These results will iteratively inform an explanation-centric debugging and assessment approach and prototype. This approach and prototype will be evaluated by a final summative study approximately two years from now.

OPEN QUESTIONS

I have three primary questions to discuss with the researchers at the IUI Doctoral Consortium.

First, capturing mental models is not trivial [15]. I would like to discuss other researchers' experiences capturing mental models, including methods others have found useful and potential drawbacks to be aware of.

Second, I am still uncertain of how to design formative studies to answer RQ2, and would like to discuss design ideas with researchers who have performed similar formative work.

Finally, I am looking for different levels of abstraction to use when investigating RQ 3.1. The two I have in mind are general decision boundaries and individual predictions, but would like to hear ideas from researchers who have more extensive machine learning backgrounds than my own.

My dissertation research is gathering momentum, and with approximately two years left, I feel this is an ideal time to discuss my work at the IUI Doctoral Consortium. I have enough experience with human-computer interaction research to constructively discuss the work of fellow students, but am not yet too far along to integrate their advice into my own research.

REFERENCES

1. Amershi, S., Fogarty, J., Kapoor, A. and Tan, D. 2010. Examining multiple potential models in end-user interactive concept learning. *Proc. CHI*, 1357–1360.
2. Ashcraft, M.H. 1994. *Human memory and cognition*. Harpercollins College Div.
3. Billsus, D. and Hilbert, D. 2005. Improving proactive information systems. *Proc. IUI*, 159-166.
4. Bruner, J.S. and Tagiuri, R. 1954. The Perception of People. *Handbook of Social Psychology*. G. Lindzey, ed. Addison-Wesley.
5. Dzindolet, M.T., Peterson, S.A., Pomranky, R.A., and Pierce, L.G. 2003. The role of trust in automation reliance. *International Journal of Human-Computer Studies*. 58, 6 (Jun. 2003), 697–718.
6. Glass, A., McGuinness, D. and Wolverton, M. 2008. Toward establishing trust in adaptive agents. *Proc. IUI*, 227-236.
7. Johnson-Laird, P.N. 1983. *Mental Models: Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press.
8. Kapoor, A., Lee, B., Tan, D. and Horvitz, E. 2010. Interactive optimization for steering machine classification. *Proc. CHI*, 1343-1352.
9. Kulesza, T., Burnett, M., Stumpf, S., Wong, W., Das, S., Groce, A., Shinsel, A., Bice, F. and McIntosh, K. 2011. Where Are My Intelligent Assistant's Mistakes? A Systematic Testing Approach. *Proc. IS-EUD*, 171–186.
10. Kulesza, T., Stumpf, S., Burnett, M., Wong, W.-K., Riche, Y., Moore, T., Oberst, I., Shinsel, A. and McIntosh, K. 2010. Explanatory Debugging: Supporting End-User Debugging of Machine-Learned Programs. *Proc. VL/HCC* (2010), 41–48.
11. Kulesza, T., Stumpf, S., Wong, W.-K., Burnett, M., Perona, S., Ko, A. and Obsert, I. 2011. Why-Oriented End-User Debugging of Naive Bayes Text Classification. *ACM Transactions on Interactive Intelligent Systems*. 1, 1 (Oct. 2011).
12. Kulesza, T., Wong, W.-K., Stumpf, S., Perona, S., White, R., Burnett, M., Oberst, I. and Ko, A. 2009. Fixing the program my computer learned: barriers for end users, challenges for the machine. *Proc. IUI*.
13. Lim, B. and Dey, A. 2010. Toolkit to support intelligibility in context-aware applications. *Proc. Ubicomp*, 13-22.
14. Lim, B., Dey, A. and Avrahami, D. 2009. Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Proc. CHI*, 2119-2128.
15. Norman, D. 1983. Some Observations on Mental Models. *Mental Models*. D. Gentner and A. Stevens, eds. Psychology Press.
16. Rosson, M. and Carrol, J. 1990. Smalltalk scaffolding: a case study of minimalist instruction. *Proc. CHI*, 423-429.
17. Settles, B. 2009. *Active learning literature survey*. University of Wisconsin-Madison.
18. Sharp, H., Rogers, Y. and Preece, J. 2007. *Interaction Design*. John Wiley & Sons Inc.
19. Stumpf, S., Rajaram, V., Li, L., Wong, W., Burnett, M., Dietterich, T., Sullivan, E. and Herlocker, J. 2009. Interacting meaningfully with machine learning systems: Three experiments. *International Journal of Human-Computer Studies*. 67, 8 (Aug. 2009), 639–662.
20. Talbot, J., Lee, B., Kapoor, A. and Tan, D. 2009. EnsembleMatrix: Interactive visualization to support machine learning with multiple classifiers. *Proc. CHI*, 1283-1292.
21. Tullio, J., Dey, A., Chalecki, J. and Fogarty, J. 2007. How it works: a field study of non-technical users interacting with an intelligent system. *Proc. CHI*, 31-40.
22. Vig, J., Sen, S. and Riedl, J. 2009. Tagsplanations: explaining recommendations using tags. *Proc IUI*, 47-56.