2009 Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing

Synthesis and Analysis in Artificial Intelligence: The Role of Theory in Agent Implementation

Roxanne B. Raine*†, Rieks op den Akker‡, Zhiqiang Cai*§, Arthur C. Graesser*¶ and Danielle S. McNamara*‖
*Institute for Intelligent Systems
The University of Memphis
Memphis, TN, USA 38152
†Email: roxi.benoit@gmail.com
‡Human Media Interaction
Universiteit Twente
Enschede, The Netherlands 7522NP
Email: infrieks@ewi.utwente.nl
§Email: zhiqiang.cai@gmail.com
¶Email: a-graesser@memphis.edu
‖Email: dsmcnamara1@gmail.com

***Abstract*—The domain of artificial intelligence (AI) progresses with extraordinary vicissitude. Whereas prior authors have divided AI into the two categories of analysis and synthesis, Raine and op den Akker [1] distinguish between four types of AI: that of appearance, function, simulation and interpretation. These *subdomains* of AI are differentiated by user goals, creator methodologies, and environmental constraints. In this paper, we focus on how analysis and synthesis could improve the subdomain of Functional-AI.**

***Index Terms*—AI; design; engineering; cognitive science; and philosophy**

I. Introduction

When one considers the creation of an intelligent virtual agent, the term intelligence can be defined as the ability to effectively communicate with a human being, as measured by how convinced the human interlocutor is that the agent is also human [2]. At this point, there is no agent worthy of the title *intelligent* by this standard. In this paper, we present an alternative approach to the design of agents as a possible step towards solving some common problems found in communicative agents.

We introduce a number of psychological findings concerning human communication that should be a resource for artificial intelligence (AI) researchers and designers. The specific theories that will aid those in the field of AI will depend on a number of factors, including whether the agents are to be used as entertainment, tools, or research, whether the professional is interested in the analysis or synthesis of agents, and what the agent's (and its human interlocutor's) environmental constraints are. In this paper, we focus on agents used as tools for specific purposes (aside from simulation of the mind).[1] We discuss the implications of the psychological empirical results and theories as they relate to the analysis and synthesis of these goal-oriented AI machines, and how that an agent's environment could interplay in our proposals.

[1]This is explained further below.

Turing [2] predicted that there would be an agent capable of passing his test of fooling a human judge into believing that it was also human within 50 years of his 1950 paper. However, even today, almost a decade later, the Loebner prize (which will be awarded to the first computer to successfully pass the Turing test) remains unclaimed. Some have argued that Turing's deadline was not met because computers were too quickly capable,[2] we attribute the lag in successful implementation of intelligent agents to a disorganization of focus. In this paper, we hope to contribute to a much-needed process of philosophical organization of the domains within AI.

Many researchers divide AI into a science of analysis and synthesis [4] [5] [6] [7]. By contrast, Raine and op den Akker [1] argue that agent research differs as a function of user goals, creator methodologies, and environmental constraints. Following this assumption, they have recently suggested [1] the division of AI into the following four categories:

1) **Appearance-Based AI.** The agent is embedded in a fictional environment, (e.g., a game, a comic, a novel). The creator (artist) designs the system as entertainment and has control over constraining the gaming environment. This makes the challenge of making the agent compatible with its environment [8] [9] less difficult [10] for this subdomain.[3] Any components of the environment that are difficult for the agent to *know* about the environment can be written out of the environment. Instead of having to create a system in accord with the environment, the Appearance-Based AI designer has the advantage of also designing the AI environment.
2) **Functional-AI.** The agent is used as a tool and appears to produce intelligent behavior. The creator develops the product for practical use (to solve a practical problem, such as an interactive railway, restaurant or travel infor-

[2]See [3] for a discussion.
[3]See [9] for an explanation of this challenge.

978-0-7695-3929-4/09 $26.00 © 2009 IEEE
DOI 10.1109/DASC.2009.141

mation system with a natural dialogue interface), and the system will often be mass produced. The environment is constrained by the intended use of the tool.

3) **Simulation-AI.** The agent is used for discovery. The creator develops and uses the agent to imitate intelligent behavior, often simulating only limited features of intelligence. This type of AI is used to simulate human intelligent behavior in order to test the validity of computational models of that behavior. The agent's environment is defined by reality.
4) **Interpretive-AI.** Instead of creating artificial agents, the philosopher analyzes them, defines their paradigm, and pioneers thought about them. The philosopher is expected to consider all possible goals, methodologies, and agendas in AI. The environment is unconstrained because even that which is not yet (or ever) possible might be considered.

In this paper, we focus on the category of Functional-AI, which creates agents that are intended to be used as purposeful tools [1]. We explore the methods of analysis and synthesis within Functional-AI, and offer suggestions that could improve the performance of artificial agents within this subdomain. By clarifying these procedures, purposes and techniques in Functional-AI research and development, the successful study and implementation of Functional-AI may be drastically expedited.

We will first describe psychological findings that relate to Functional-AI analysis. We will then explore the implications that these findings have in Functional-AI synthesis. Our goal will be to argue for a distinction between an agent's *imitation of* and *compatibility with* humans.

II. Analysis

The fact that we are humans means that we can easily and often effortlessly infer our fellow human interlocutor's perspective based on our own experiences. Often, it is possible to assess ones self as an approximation of the conversational partner [11], [12], [13]. That is, a person might consider their own perspective when making assumptions about another persons perspective [14], even to a fault.

In 1978, Robert Stalnaker [19] defined *presuppositions* as "the propositions whose truth [the speaker] takes for granted as part of the background of the conversation... what is taken by the speaker to be the common ground of the participants in the conversation, what is treated as their common knowledge or mutual knowledge" (p.320). In general, it is these presuppositions that pose challenges in the design of dialogue systems. It is the difficulty in programming all of the necessary presuppositions into a computational system that causes the Loebner prize to remain unclaimed.

Speakers and listeners have been found to initially assume that the features of the environment that are salient and readily available to themselves are also obvious to their conversational partners, even when this is clearly not the case [13], [11], [12], [15]. Although this rarely causes substantial problems in human-human communication, it is likely at the core of the problem for human-computer interaction (HCI).

For an example of the human tendency to base presuppositions on their personal knowledge, Keyser and Henley [13] found that speakers tended to assume addressees understood, but the overhearers to the communication did not tend to think the addressees would understand the speaker. In other words, overhearers, who had all of the information of the communication, were *not* likely to assume that the speaker's addressee would understand the speaker's perspective. It is likely that the speakers' and observers' assessments of the addressees' understanding (or lack thereof) differed because the speakers were assuming that the addressees (and observers) had more communicative grounding Clark than was actually the case. Keyser and Henley's result supports the claim that people tend to assume a higher level of mutual understanding with their interlocutors than they actually have. This false assumption seems to be related to a human heuristic that allows one to infer another's point of view based on one's own knowledge [14]. Results such as these have also been found in numerous other studies of human-human communication [11], [12], [15].

Human interlocutors seem to base their communications largely on a backdrop of their own conscious experiences. This feature of human communication could give the human an advantage in communicating with a fellow human interlocutor. The computer, on the other hand, does not have comparable perceptual experiences on which to base its judgments of its conversational partner. That is, two humans will always have more in common with each other than a human and a computer would have in common, simply because the two humans are of the same species.

This relatability from one human to another human could give human-human interaction a substantial advantage over HCI. Just as one cannot completely understand what it is like to be born of the opposite sex or in a different culture, one cannot easily explain in its entirety what it is like to have the experience of being oneself. In like manner, it is highly unlikely that humans can tell computers everything that needs to be known about the human experience.

It is unreasonable to assume that programming a computer to imitate human behavior would ever yield a successful and impressive interactive agent. In this paper, however, we aim to show that the situation is not as dire as it may seem. Although one should not aim to *imitate* human behavior to achieve successful HCI, the Functional-AI developer has a great advantage because in this subdomain, one's goal is to create an agent that is *compatible with* human behavior.

III. Synthesis

Humans have at least two attributes that make them successful in communicating with humans. The first of these is that we are human, as was explained above. The second is that we have evolved over generations to communicate with humans. The fact that humans *are* humans is definitely an achievement that computers are not going to attain. However, it *is* possible that computers could evolve to be compatible with human interlocutors. Moreover, computers have the additional advantage of being able to *evolve*[4] much more quickly than biological creatures.

The human mind is not as computationally capable as the computer in many (if not all) respects. These computational limitations necessitate the use of many heuristics and short-cuts in order for people to process language and successfully communicate.[5] There are a number of consequences of these heuristics and short-cuts that relate directly to communicating with and creating chatbots. Unfortunately, many of these consequences provide designers with challenges in imitating human interlocutors, including problems with framing, associations, and (debatably) connotations or meaning.

However, there is also an upside to these human tendencies in communicative processes. As mentioned above, humans make a number of assumptions when entering into a communication. Instead of trying to model a chatbot that imitates these assumptions, we suggest that it would be better to model a chatbot that takes advantage of typical assumptions in specific situations. One such example would be the anthropomorphizing of chatbots when the human interlocutors have no reason to believe that the chatbot would be a non-human. If a chatbot is intended to fool humans into believing that it is also human, this anthropomorphization could be an asset.

Chatbots that are designed exactly the same but set in different environments can have very different effects on their human interlocutors [17]. People tend to assume that they are understood by their listeners and that they are on the same page or following along with their speakers, they are fooled easily by chatbots that are set in constrained environments. It follows that humans online who are expecting to communicate with fellow humans will be unlikely to discover that a chatbot is a non-human.

For example, a chatbot named CyberLover has been infiltrating dating chat sites, and luring lovesick interlocutors into providing personal information or visiting malicious websites. Based on a recent news brief from PC Tools News, CyberLovers creators claim that the chatbot is capable of establishing a new relationship with up to ten partners in just 30 minutes [18]. Although CyberLover's creators' goals were unethical, they were a part of Functional-AI. The chatbot is used for a purpose: to steal from unsuspecting human interlocutors by convincing them that it is a trust-worthy human[6]

Enhancing an agent's performance by using computational tools that are known to be effective should produce quicker results in achieving compatibility than imitation of human behaviors (at least for Functional-AI agents). Consider the technique of a bot that is capable of keeping up with the transience of language use by having scanning the internet (e.g., Google's spell-checking, auto-correcting, and search suggestion features), which basically works on the premise that less frequently appearing spellings of a word are likely *mis*spellings. Pandorabot's forthcoming SpellBinder implements a similar tool, and applies this general computer science application to AI. Pandorabot, like Google's search-suggestion feature, automatically learns from transcripts. For Pandorabot, the transcripts are not necessarily web-based, but include movies, television shows, and actual human interactions, among other types of transcripts. The point here is that Pandorabot is not imitating human behavior. The Pandorabot is actually imitating another 'bot' (i.e., Google's system).

Based on the psychological theory presented in the previous section on analysis, the creators of Functional-AI systems could constrain agent environments so as to make their agents more believable. As our first author has mentioned previously [17], we suggest that a successful communicative agent needs to be more than intelligent. It must also be *clever* enough to use what we know about how humans communicate to its advantage. In this way, computers could begin to evolve (so to speak) into agents capable of communicating with humans.

It is possible to create an agent that is sufficiently convincing given that little is required of the agent, its environment is substantially limited, and its remarks can be consistently vague without raising suspicion of its artificiality from its interlocutor. Indeed, the above-mentioned tendency of humans to enter into communications with a set of presuppositions contributes to CyberLover's success. It is also true that the human tendency to attribute meaning and coherence to their experiences also aids in the deception of the interlocutor.[7]

IV. Conclusion

Taking the human tendency to base one's perception off of one's own private knowledge into consideration when constructing a conversational agent should be quite advantageous. There are a number of ways that theory could be considered in implementing Functional-AI chatbots in the future, which should begin during the stages of analysis. When a design group aims to create an agent that will be mass produced as a tool in the future, that group should aim to analyze characteristics that could make the agent more compatible with human interlocutors.

Functional-AI creators should search for ways that they can also use short-cuts in achieving common ground, but not by imitating the human algorithm. Rather, creators should take advantage of human tendencies by using them to the agent's advantage. This can be done with the use of vagueness in language, environmental constraints, and certainly a number of other ways that have yet to be discovered.

[4] Clearly, we are using the term *evolve* loosely here.

[5] Indeed, as Turing mentioned, one could not fault the computer for its tendency to be better at calculation and memorization than the human [2].

[6] It is worth noting that CyberLover is a rare example of a Functional-AI system that is subject to the Turing test standard in AI. CyberLover's success depends on its ability to convince humans that it is also human. Most Functional-AI agents are not subject to this level of scrutiny because their use as tools do not usually require them to convince humans that they are also human.

[7] It also helps if the interlocutor does not know that there is a possibility that the chatbot is not human, which makes this situation and that of CyberLover very different from Turing test situations.

Hopefully, the clarification of AI subdomains, standards, and delineations of analysis and synthesis within these subdomains will improve our understanding of how exactly we can make agents that are every bit as compatible with human interlocutors as humans are themselves. It is our opinion that the first step in this process is to aim towards this goal of compatibility instead of imitation.

ACKNOWLEDGMENTS

This research was supported in part by the Institute for Education Sciences (IES R305A080589 and R305B070349) and the National Science Foundation (IIS-0735682). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES or the NSF.

REFERENCES

[1] R. B. Raine, and R. op den Akker, "Towards a clarification of agent subdomains and standards," Paper submitted to Innovations in Computer Science, Beijing, China, submitted.

[2] A. Turing, "Computing machinery and intelligence," *Mind*, vol. LIX, pp. 46-84, 1950.

[3] L. Crockett, *The Turing Test and the Frame Problem: AI's Mistaken Understanding of Intelligence*. Norwood, NJ: Ablex Publishing Co., 1994.

[4] S. Larsson, "Dialogue systems: Simulations or interfaces?" in *Proceedings of the ninth workshop on the semantics and pragmatics of dialogue*, Gardent and Gaiffe, Eds., Nancy, France, 2005.

[5] A. Newell and S. K. Card, "Straightening out softening up: Response to Carroll and Campbell," *Human-Computer Interaction*, vol. 2, no. 3, pp. 251-67, 1986.

[6] D. Schlangen, "What we can learn from Dialogue Systems that don't work: On dialogue systems as cognitive models," in *Proceedings of DiaHolmia (Semdial 2009)*, Stockholm, Sweden, 2009.

[7] R. Pieraccini and J. Huerta, "Where do we go from here? Research and commercial spoken dialog systems," *Proc. of 6th SIGdial Workshop on Discourse and Dialog*, Lisbon, Portugal, Sept., pp. 1-10, 2005.

[8] J. McCarthy and P. Hayes, "Some philosophical problems from the standpoint of AI," in B. Meltzer and D. Mitchie, Eds. *Machine Intelligence*, vol. 4. Edinburgh: Edinburgh University Press, pp. 473-502, 1969.

[9] J. Fodor, "Modules, frames, fridgeons, sleeping dogs, and the music of the spheres," in *The Robot's Dilemma: The Frame Problem in AI*, Z. W. Pylyshyn, Ed. Norwood, NJ: Ablex Publishing, 1987, pp. 139-49.

[10] D. Waltz, "The prospects for building truly intelligent machines," *Daedalus*, p. 204, Winter 1988.

[11] W. S. Horton and B. Keysar, "When do speakers take into account common ground?" *Cognition*, vol. 59, pp. 91-117, 1996.

[12] B. Keysar, "The illusory transparency of intention: Linguistic perspective taking in text," *Cognitive Psychology*, vol. 26, pp. 165-208, 1994.

[13] B. Keysar and A. S. Henley, "Speakers' overestimation of their effectiveness," *Psychological Science*, vol. 13, pp. 207-212, 2002.

[14] D. Kahneman and A. Tversky, "On the psychology of prediction," *Psychological Review*, vol. 80, pp. 237-257, 1973.

[15] R. B. Raine, "Who dat and where y'at? Acadians and their conventions for consideration of intercultural interlocutors: Towards a disambiguation of perspective taking in communication research," Ph.D. dissertation, Inst. Cog. Sci., Univ. LA at Lafayette, 2009.

[16] H. H. Clark, *Using Language*, Cambridge: Cambridge University Press, 1996.

[17] R. B. Raine, "Making a clever intelligent agent: The theory behind the implementation," IEEE International Conference on Intelligent Computing and Intelligent Systems, Shanghai, China, 2009.

[18] PC Tools issues warning to singles on social networking and online dating sites: Beware of flirting robots. (2007, December 12). PC Tools News. PC Tools, 1998-2008. Retrieved March 22, 2008, from http://www.pctools.com/news/

[19] R. Stalnaker, "Assertion," in P. Cole, Ed. *Syntax and Semantics 9: Pragmatics*, NY: Academic Press, pp. 315-332, 1978.

[20] J. Weizenbaum, "ELIZA A computer program for the study of natural language communication between man and machine" *Communications of the ACM 9*, pp. 36-45, 1966.