User Satisfaction *without* Task Completion

Peter Wallis
Centre for Policy Modelling
Manchester Metropolitan University
All Saints Campus, Oxford Road
Manchester, M15 6BH, UK
pwallis@acm.org

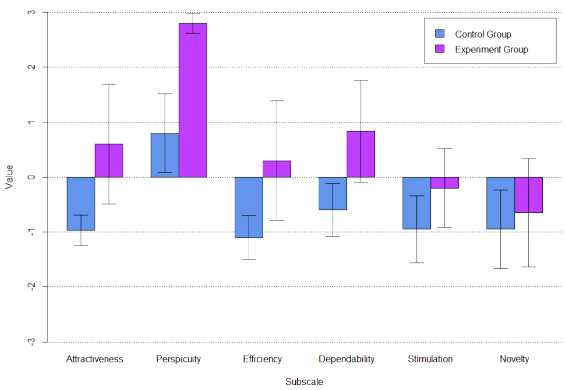

Figure 1: User satisfaction as measured using the UEQ questionnaire (Laugwitz et al., 2008) for two spoken dialog systems, one of which has added **social intelligence**.

Figure 1 presents an interesting result. We implemented a spoken dialog system that provided directory assistance, gave some subjects a set of tasks to do using the system, and then measured their user satisfaction with a standard HCI tool. We then gave the system some 'social intelligence' and re-ran the trial with the same tasks and with subjects from the same population. The graph shows a result that is not only significant (0.95) but also dramatic in that the experience was generally seen as positive, contrasting with the reaction to the control case and with popular opinion of IVR systems in general. A good result but, more importantly, note the experimental system did not work any better. The tasks were chosen such that only 20% were achievable no matter how good the interface. It seems better user satisfaction can be attained without making the system work better.

In previous papers we have given the background and motivation for our notion of social intelligence (Wallis, 2013), and described in detail the experimental setup (Wallis et al., 2014). In this short paper we describe the work underway to implement something more than a demonstrator.

1 The Theory

The idea that a computer could understand and use language has been with us from the very beginnings of computer science. Despite massive effort and considerable commercial potential, developments in the area have met with limited success. Historically the focus has been on the information conveyed by language but we are developing the idea that language is primarily social in purpose and function. Rather than focus on language and meaning, we focus on issues such as power and distance, roles and obligations, all within the context of normative relations and human/cultural expectations.

Moving down a level, we embrace Tomasello's claim that human communication is **intentional** and **cooperative** (Tomasello, 2008). This move is however the cumulation of 15 years looking at language as action in a social setting ranging from work on politeness (Wallis et al., 2001) and abuse (de Angeli et al., 2005) to conversational strategies (Wallis, 2008) and engagement (Wallis, 2010). In summary the key to language *as humans use it* is their surprisingly effective (but hard to notice) skills at recognising the intent of others. To use an example from Dennett, seeing two children tugging at a teddy bear, the human observer will be quite certain they both *want* it (Dennett, 1987). Even if people don't actually reason in terms of beliefs, desires and goals, the intentional stance we take is how we think others think when we communicate with them and is hence key to the recipient design of our utterances.

What is more we humans are (socially) compelled to cooperate in the process. If we have reached the point of being *engaged* (Wallis, 2010) in a conversation with someone, then we work hard to *account for* (Seedhouse, 2004) what he or she says. Some actions – especially classic speech acts – are intended to be interpreted, other communicative acts are just ‘radiated’ (such as smiles) and others, such as unconsciously scratching your nose, are just acts. A chimpanzee according to Tomasello, is perfectly capable of recognising intent, but has no social compunction to interpret actions as those of communication.

2 The Mechanism

Given this is the true nature of the ‘language instinct,’ there are two challenges for those who want to engineer better conversational agents. How do we create a conversational agent that can recognise the intent of its interlocutor, and what intentions should the agent have and when? The current work in this area at CPM is using a Belief Desire and Intention (BDI) architecture to implement a dialogue manager. BDI has been used for this many times before (Ardissono and Boella, 1998; Wallis et al., 2001; Kopp et al., 2005; Wong et al., 2007) and such use is often, it seems, conflated with Good Old Fashioned AI models of conversation based on planning (Allen et al., 1995). In the rest of this article I will use the term BDI to mean a Rao and Georgeff (Rao and Georgeff, 1995) style BDI system which does not *do* planning but rather selects and manages plans from a static plan-library. Such architectures were introduced to explicitly address the issue of situated action associated with traditional planning systems; the advantage it has over more popular approaches to the issue such as Behaviour Based Robotics (Arkin, 1998) is that it maintains a notion of working to a recipe. A BDI architecture in the sense used here explicitly balances reactive and deliberative behaviour, managing plans rather than creating them.

Intention recognition is a task that poses some interesting and challenging problems for AI research but not all it poses are insurmountable. A large slab of the general problem can be handled using a BDI architecture and treating intention recognition as a variant of plan selection (Heinze, 2003). No doubt a human would do it better, but the interactive nature of the dialog problem means that, as long as the system can account for its failings in an understandable way, the human will forgive it in much the same way we accommodate children *without blame* for their lack of knowledge.

A bigger challenge is the question of what the system ought to intend (to do) and when. Ultimately of course machine learning should be able to mimic human (intentional) behaviour in a social setting and so identifying explicit intentions would become redundant. The problem is that unacknowledged theory tends to be embedded in the training data (Hovy, 2010). The recent Dialog State Tracking Challenge (Hen, 2014) is, although an excellent and exciting development, a case in point with notions of dialog state being based on Information State Update (Kreutel and Matheson, 2000). As a model of human communication ISU puts, we believe, too much emphasis on the information ‘carried’ (Reddy, 1993) by speech acts and pays insufficient attention to the larger structures within dialog we think of as intentional (Wallis, 2008).

Studying language has of course been the work of many for centuries if not millennia but such work tends to be seen as ‘unscientific’ by many with a physical sciences background and confined to the dusty shelves of forgotten libraries. Once one has a model of intention however, descriptions of people wanting X, believing Y, and Z being normative become concrete enough to implement. Creating models of causality in such relations is hard however because it is all so *obvious* to us humans - too obvious to notice. It takes special skills and training to do the noticing and what is needed is some cross disciplinary work to identify a set of intentional structures (defined as a BDI plans) that might be used by a synthetic social actor filling a particular set of roles. We have had past successes with researchers from Applied Linguistics using the ethnomethodological variant of Conversation Analysis (ten Have, 1999) and with Grounded Theory (Urquhart et al., 2010), but being at core computer scientists, we are open to suggestions.

3 Conclusion

The system used to produce the data in Figure 1 was a demonstrator that only worked for the tasks given to the subjects and used ‘canned’ expressions much like the classic chat-bot mechanism (Ali, 2001). Our aim now is to implement the system fully and deploy it with members of the public with real information needs.

References

2001. The ALICE artificial intelligence foundation, January. http://206.184.206.210/alice_page.htm.

J. F. Allen, L. K. Schubert, G. Ferguson, P. Heeman, C. H. Hwang, T. Kato, M. Light, N. G. Martin, B. W. Miller, M. Poesio, and D. R. Traum. 1995. The TRAINS project: A case study in defining a conversational planning agent. *Journal of Experimental and Theoretical AI*, 7(7):7–48.

L Ardissono and G. Boella. 1998. An agent architecture for NL dialog modeling. In *Artificial Intelligence: Methodology, Systems and Applications*, Berlin. Springer (LNAI 2256).

Ronald C. Arkin, editor. 1998. *Behavior-Based Robotics*. MIT Press, Cambridge, MA.

Antonella de Angeli, Sheryl Brahnam, and Peter Wallis (eds). 2005. *Abuse: the darker side of Human-Computer Interaction*. INTERACT, Rome, September. http://www.agentabuse.org/.

Daniel C. Dennett. 1987. *The Intentional Stance*. The MIT Press, Cambridge, MA.

Clinton Heinze. 2003. *Modelling Intention Recognition for Intelligent Agent Systems*. Ph.D. thesis, Department of Computing and Information Systems.

2014. Dialog state tracking challenge 2&3. http://camdial.org/ mh521/dstc/.

Eduard Hovy. 2010. Injecting linguistics into nlp by annotation, July. Invited talk, ACL Workshop 6, NLP and Linguistics: Finding the Common Ground.

Stefan Kopp, Lars Gesellensetter, Nicole Kramer, and Ipke Wachsmuth. 2005. A conversational agent as museum guide - design and evaluation of a real-world application. In *5th International working conference on Intelligent Virtual Characters*. http://iva05.unipi.gr/index.html.

Jörn Kreutel and Colin Matheson. 2000. Modelling dialogue using multiple inferences over information states. In *Proceedings of ICOS-2, 2nd Workshop on Inference in Computational Semantics*, Dagstuhl.

B. Laugwitz, T. Held, and M. Schrepp. 2008. Construction and evaluation of a user experience questionnaire. In A. Holzinger, editor, *USAB 2008 (LNCS 5298)*, pages 63–76.

A. Rao and M. Georgeff. 1995. BDI agents: from theory to practice. Technical Report TR-56, Australian Artificial Intelligence Institute, Melbourne, Australia.

Michael J. Reddy. 1993. The conduit metaphor: A case of frame conflict in our language about language. In Andrew Ortony, editor, *Metaphor and Thought*. Cambridge University Press.

Paul Seedhouse. 2004. *The Interactional Architecture of the Language Classroom: A Conversation Analysis Perspective*. Blackwell, September.

Paul ten Have. 1999. *Doing Conversation Analysis: A Practical Guide (Introducing Qualitative Methods)*. SAGE Publications.

Michael Tomasello. 2008. *Origins of Human Communication*. The MIT Press, Cambridge, Massachusetts.

C. Urquhart, H. Lehmann, and M. Myers. 2010. Putting the theory back into grounded theory: Guidelines for grounded theory studies in information systems. *Information Systems Journal*, 20(4):357–381.

Peter Wallis, Helen Mitchard, Damian O'Dea, and Jyotsna Das. 2001. Dialogue modelling for a conversational agent. In Markus Stumptner, Dan Corbett, and Mike Brooks, editors, *AI2001: Advances in Artificial Intelligence, 14th Australian Joint Conference on Artificial Intelligence*, Adelaide, Australia. Springer (LNAI 2256).

Peter Wallis, Keeley Crockett, and Clare Little. 2014. When things go wrong. In Sarvapali D. Ramchurn, Joel Fisher, Avi Rosenfeld, Long Tran-Thanh, and Kobi Gal, editors, *Human-Agent Interaction Design and Models (HAIDM)*, Paris.

Peter Wallis. 2008. Revisiting the DARPA communicator data using Conversation Analysis. *Interaction Studies*, 9(3), October.

Peter Wallis. 2010. A robot in the kitchen. In *ACL Workshop WS12: Companionable Dialogue Systems*, Uppsala.

Peter Wallis. 2013. The intentional interface. In Mario Conci, Virginia Dignum, Mathias Funk, and Dirk Heylen, editors, *Proceedings of the Workshop on Computers as Social Actors*. CEUR, September. http://ceur-ws.org/Vol-1119/.

A. Wong, A. Nguyen, and W.R. Wobcke. 2007. Robustness of a spoken dialogue interface for a personal assistant. In *Proceedings of the 2007 IEEE/WIC/ACM International Conference on Intelligent Agent Technology*, pages 123–127.