

Improving Human Interaction with Autonomous Systems: Supporting Intentionality through Increased Awareness

Peter Goodison
University of Bath
Claverton Down, Bath, UK
p.goodison@bath.ac.uk

Peter Johnson
University of Bath
Claverton Down, Bath, UK
p.johnson@bath.ac.uk

Joanne Thoms
BAE Systems Integrated System
Technologies Limited,
Lyon Way, Frimley, Surrey, UK
Jo.thoms@baesystems.com

ABSTRACT

Understanding intentionality is a necessary feature of joint activities involving interdependent agents. This challenge has increased alongside the deployment of autonomous systems that are to some degree unsupervised. This research aims to reduce the number of intentionality recognition breakdowns between people and autonomous systems by designing systems to support the awareness of information cues used in those decisions. The paper outlines theoretical foundations for this approach using simulation theory and process models of intention. The notion of breakdowns is then applied to mistaken intentions in a diary study to gain insight into the phenomena.

Author Keywords

Awareness, Autonomous Systems, Intention, Design

ACM Classification Keywords

H5.2 User Interfaces

1. INTRODUCTION

The application and growth of increasingly independent and autonomous systems (AS) working with and alongside people is leading to the development of an ‘awareness gap.’ Klein, Woods, Bradshaw, Hoffman and Feltovich state “to be a team player, an agent—like a human—has to be reasonably predictable, and has to have a reasonable ability to predict the actions of others.... currently, however, the ‘intelligence’ and autonomy of agents directly works against the confidence that people have in their predictability” [14]. People make decisions about future actions, in part, according to their observations of other actors (human or machine). This decision process is made increasingly problematic because these systems are currently not designed to aid awareness or understanding by others of the purpose of their actions. Researchers have noted that actors working jointly coordinate and integrate tasks seamlessly in collocated environments [11]. Such recognition, understanding and prediction of future states, events, and outcomes have been described as awareness that enhances a mutual understanding of the changing situation. Awareness has been a term used elastically to describe different processes of gaining understanding. Peripheral awareness [8], context awareness [16], situation

awareness [5] and workspace awareness [9] are some of the research threads identified where people benefit from obtaining information about their environment. Endsley’s definition of awareness is, “knowing what is going on” [5] in the sense of understanding why an event is occurring e.g. its purpose.

Unlike approaches that seek to understand how awareness of the current and past states of a scenario can be delivered and supported [15], the research presented in this paper seeks to understand how future predictions of other actors are made and how such predictions might be aided and enhanced. This will benefit instances where actors (human or machine) must interact to achieve a desired outcome.

2. AWARENESS AND AUTONOMOUS SYSTEMS

There are situations when people do or will need to work jointly with AS. AS are already being deployed in various contexts such as maritime, aerospace, medical healthcare and industrial [22]. In this research AS are systems that “make and execute a decision to achieve a goal without full, direct human control” [2].

The task of maintaining control over a system that decides its own actions can be a difficult one. The main benefit of an AS is that it provides an additional effort and capability for certain activities. Supervisory approaches to Human-AS interaction try to mitigate the human awareness of the AS intended actions problem by constraining AS behaviour. This method requires AS to operate according to a predefined plan library that is deterministic and restricts possible actions [17]. It demands detailed knowledge and analysis of the plans, procedures and actions needed to complete known and envisioned tasks. This is compounded by the fact that often the future situations of operation are dynamic and unpredictable. Consequently it can be difficult to make AS activity adaptive and flexible and devolve much decision-making and problem solving to the AS. When different models of Human-AS control are used that allow the AS to have greater flexibility in planning and choosing its own actions, problem solving and decision-making, then the problem of predictability by the human of what the AS will do next or why it is doing it is increased. This presents challenges for human-AS interaction design.

One avenue in which to pursue this problem is to understand how people cope with the uncertainty and unpredictability found in person-person interaction and joint activity. This could then be used as a basis for considering the less understood human-AS interaction requirements.

As social beings we learn about the world around us through other people, both actively and passively. People have evolved information gathering, decision making, problem solving and anticipatory judgment processes and mechanisms to understand other people's likely actions and intentions because it is advantageous to their own survival and well-being to anticipate the future actions of others. These processes and mechanisms require knowledge of social action and behaviour, the perception of relevant features and attributes contributing to the formation of hypothesized intentions and the subsequent testing and reshaping of those hypotheses. However despite this evolved knowledge processes and mechanisms maintaining awareness of an actor's intentions can remain problematic for people. They strive to find meaning from an actor's behaviour and transpose this into the supposed intentions to produce possible implications that then in-turn may impact upon their own intentions, goals, plans and future actions. Interpreting the situation correctly relies on perceiving the features and attributes of available cues, a rich, experiential knowledge base and rapid and effective decision-making and judgment. When successful potential conflict or deviation is foreseen, future actions are dynamically altered, breakdowns are avoided, repair costs are minimized and successfully coordinated joint activity occurs.

In collocated settings intention recognition is aided by an actor having a greater perceptual awareness of the other agents and environment. But systems of remote distributed components require additional support to achieve even reduced levels of awareness and to enable coordinated work efforts. Currently when performing shared and/or joint tasks that require interacting with others or technology in distributed environments systems provide support for a users' awareness of state and action. For example, a system allowing participants to monitor each other's tasks in a collaborative activity [15]. Hourizi & Johnson [12] have shown that this awareness can be significantly enhanced when information about future actions, intentions and implications are provided. In addition they have proposed and applied a framework to aid the design of technology to support this.

In the psychological research domain studies have invoked the 'theory of mind' [1] to explain how people predict other people's intentions. However, since AS are not people can we attribute their behaviour to a causal theory that assumes they use the same mechanism as people? While the reviewed literature that follows provides an explanation of how intentions are recognised, this theory has not been

tested or attempted to be applied. By considering how this might be applied to system design we can then go on to investigate how applicable it is and how well it enables awareness of intention in human-AS interactions. This can be framed as the questions; what information is required for judgments of intentionality? How can that be applied to human-AS interaction? Does supplying that information have any affect upon human awareness of AS behaviours and their interactions with AS?

3. INTENTION RECOGNITION

The theory of mind allows humans to understand that other agents (human or machine) have different goals from our own through recognising intention. This is termed a judgment of intentionality. We recognise that motivation for action can be internally and externally generated which we can interpret in others as reasons or causes and make attributions about their motives. Intention has been defined as "a plan of action the organism chooses and commits itself to the pursuit of a goal - an intention thus includes both a means (action plan) as well as a goal" [21]. In dissecting this definition we can identify that actions have different types of purpose. Intentions can represent the end accomplishment of a plan and the tasks within the plan that enable the high level goal. This is recognising that predictions can be either short-term (the next action) or longer term (the end goal).

Considering the intention of an action as a prediction can serve two main benefits. From a hindsight perspective it can be used to understand and account for the reason for one particular action being performed over another i.e. where an action fitted into a plan after a goal has been achieved. In foresight understanding an intention can create a prediction and anticipation of the future possible states that other agents will create. This is consistent with [18] who propose different types of intention - 'intention to' and an 'intention that' referring to an intention as an end goal or an intention as the purpose of an action enabling a state within a plan.

A key challenge in designing AS as effective team players is for people co-working with AS to be able to understand the system's intentions allowing human and machine interaction to be predictable and reliable [14]. The evolution of command style computing to current and future AS, means systems are now making an increasing number of decisions independently of human supervision and authorisation. An imperative solution is needed to address the problem that people are becoming the isolated observers trying to understand what is happening and what is the most appropriate action required, in a situation where computers can choose to effect how a situation develops.

Judging intentionality is a decision-making process. The product of the process is dependent upon how the information about a scenario is perceived and cognitively processed. Due to decisions of intentionality being

dependent on the information available about an AS, false or lack of a critical level of information will result in an incorrect assumption about intentionality.

The research position taken here is that intentionality decisions in people are applied heuristics that cognitively “pattern match.” Pattern matching uses observable information as cues to match to plans and actions of how to achieve goals in a given context stored as knowledge structures. Pattern matching occurs constantly as we observe a sequence of actions and construct an explanatory chain for each action to understand the end goal. Much of the motion and action processing occurs at an autonomic level [20] although we can consciously reason about intention using more effortful higher level reasoning. For observers this means a lack of information required for accurate matching, or incorrect information, will result in an incorrect match. The theoretical account adopted of how people are able to distinguish intentions is simulation theory [7]. As a synopsis the theory accounts for third-person mental attributions by the attributor imaginatively putting themselves in the ‘targets shoes’. They pretend to have the same initial states – such as desires and beliefs – and then simulate the decision being made. People do this by using the same mechanisms necessary to decide their own actions and running the ‘target agents’ inputs into these mechanisms. The output from these mechanisms will then produce the ‘targets intent’.

We find the types of information inputted into such a mechanism in process models of intentionality such as Read and Miller’s social dynamics model of person perception and dispositional inference [19]. They propose that inferred traits are frame-based structures or schemas where the goals of the agent being inferred play a central representational role. Read and Miller liken trait inference to the mind constructing “stories” or “narratives” where the trait is used to encode the perceived information as an exemplar to judge if the observed behaviour is a match. The encoding can include the agents, objects and actions perceived in an environment that are then assembled into a scenario. This scenario will activate in reasoning and memory possible goals for the action operating as a bottom-up and top-down process.

Together these theories suggest user human-AS interaction needs to support intentionality recognition by making the input cues required perceivable, salient and semantically pertinent. These points can be described in the following questions. What cues are required for intentionality awareness? How can those cues ensure attention at the most appropriate time and when should that be? How can the cues be represented in a meaningful way? Since there has been little research in this area the present research has begun by examining which cues do or do not improve intentionality awareness.

4. BREAKDOWNS

To enable the investigation of supporting intentionality this approach examines what information requirements are used in intention recognition by analysing breakdowns to reveal the information that was missing or identifying what additional information may have prevented the error. The notion of a breakdown used by Easterbrook allowed the identification of such failures. Easterbrook defines a breakdown as “a mismatch between the internal mental models of people, such as the expectations of one participant in a collaborative activity and the actions of another participant in the same activity” [3]. In adapting this for the specific use of intentionality in this research, a breakdown would be identified if observers assumed an AS had a particular goal (expectation), but then turned out to be mistaken (mismatch).

Hourizi and Johnson’s awareness model predicts that for information in an environment to generate awareness it must be subject to certain cognitive processes - (1) available (2) perceived, (3) attended to and (4) evaluated for implications [12]. Applying this general model of awareness, leads to the identification of challenges to improve intentionality awareness. Identifying what needs to be perceived to make an intentionality decision, when should the information be made available (attended to) and how to represent semantically implicational information remain pertinent questions?

This model can also be used for evaluating post breakdown situations to identify a likely cause such as what information was unavailable to perception or was available but was not attended to? This approach was used in the analysis of intentionality breakdowns captured through a diary study.

5. UNDERSTANDING INTENTIONALITY BREAKDOWNS

A diary study was designed to capture a breadth of examples of intentionality breakdowns in naturalistic real-world settings. This was conducted by giving 4 participants notebooks where they recorded breakdowns as they went about their day-to-day life. Participants kept the notebooks for 2 weeks. Participants were instructed to record the time, location and a description of the error along with why they thought it happened. The research was interested in capturing errors where the participants had misinterpreted someone else’s intention incorrectly, or someone else had interpreted their intention incorrectly. The results provided 23 real life examples of people misreading someone’s intention. They raised a number of issues and guidance in the intentionality recognition needs to be supported.

Firstly the study informed an understanding of when and why intentions are judged at all. Many of the examples recorded by participants were included as breakdowns because behaviour was performed by an actor as a signal directed towards other actors. This signal was expected to

prompt a required action by the other, but in the recorded cases did not. These cases demonstrate explicit demands for actors to judge intentionality. According to Kanno, Nakata and Furuta [13] intended recognition is a favourable method of reducing breakdowns by the actor being aware that it is being observed and actively cooperating in the recognition. The present research assumes this would represent an AS supplying the appropriate information to enable a correct inference.

The diary study also recorded implicit breakdowns referred to as keyhole recognition [13]. In these cases the observed actor is unaware of the other actor, and executes their actions with no special consideration for the other. Implicit and explicit breakdown types reflect which actor has recognised that actions may be necessary. Some tasks require actors to aggregate their combined activities together to achieve the desired outcomes prompting explicit intentionality. Other tasks are more loosely coupled and simply require actors to avoid conflicts with other actors' activities, relying on more implicit recognition.

Types of intentionality breakdowns can be further categorised into the response they provoke. Where an explicit signal is performed, an actor has 2 possible breakdown outcomes. Perform the wrong behaviour or fail to respond at all to the stimulus. Similar breakdowns can be labeled of implicit examples. Here an actor may perform unnecessary actions prompted by judgments about another actor, or fail to anticipate that actions were needed.

There are at least 3 minimum types of constructors that are used by people for cue information – the actor, the action(s) and the target object(s). The actor represents the agent being taking action. The action represents the activity that the observed actor is performing. The target object/state is the component that the action is directed towards in order to change it in a desired way. The availability of each of these informational elements suggests it will affect an actor's ability to attribute intentionality accurately.

Making information cues available is important, but the cues must also be correctly categorised. Categorising the identity of elements correctly is crucial to creating an accurate interpretation of intention. If an element is miscategorised then the likely goals according to that misrepresented intention will be activated as expectations of likely goals, instead of the correct set of possibilities. This can apply to the agent, actions and target object. A likely explanation for miscategorising is a lack of "visibility" of an element resulting in it being obscured or obstructed. Also people's mechanism of intentionality recognition operates by expectation and the identification of one component increases the likelihood of a closely related element being identified. This leaves highly routine/rule based activities where familiarity, need and high frequency cause expectation effects, and tasks that require plans of similar/matching action sequences, to be vulnerable to

greater numbers of breakdowns.

Categorising the fundamental elements of a scenario is important because possible goals for observed actions will be generated by knowledge about what the actions may be trying to accomplish in that specific circumstance. Variables such as the role of the actor, the values of the actor, the capability of the actor, the enabled goals, the culture and the context can produce radically different motives for similar actions.

Understanding the location of elements is also important for actors to generate possible effects and relationships of proximity between them, which aid the generation of causal links of motivation.

In the case of interpreting the intention of an actor a judgment of their attention is necessary to generate the correct goals that could be applied to the attended component. Between people this is often accomplished by following the actor's gaze to understand where their attention is directed. In AS this feature will require a new representation because radar based sensors are omnidirectional as can be artificial computer vision systems. Highlighting which element an action is directed towards will aid why an action was needed and what that action is trying to change.

It should be remembered that although there are many possible motivations for human behaviour, the motivations of AS are constrained, so the awareness aids discussed may potentially narrow the possible goals considerably when observing.

6. CONCLUSION

Supporting intentionality recognition to aid human-AS interaction requires the delivery of appropriate cues and context within which to process those cues. This will enable people to reduce the possible number of goals in a search space when making a judgment of intentionality thereby reducing the number of opportunities for incorrect interpretation and resultant repairs or breakdowns. While certain types of breakdowns can be prevented others cannot, and may never be avoided. Since an action can be caused by more than one intention, certain behaviours will always display more than one goal. For example when someone opens a window is it because they are too hot or because there is a smell in the room? The only way to increase certainty is to supply more contextual or explicit intentional information, but this can become impractical. Secondly even supplying all the correct information, making people attend to it and understanding it will still result in different interpretations because people possess different knowledge and preferences of how to achieve goals. With further studies we aim to determine the effectiveness of supplying information based on current models of intentionality. Systems trying to support intentionality first need to address what information to

provide before trying to negate the effect of individual interpretative decisions on judgments of intentionality.

7. ACKNOWLEDGMENTS

I would like to thank the anonymous participants of the diary studies and my sponsors the EPSRC and BAE Systems, Integrated Systems and Technologies.

8. REFERENCES

1. Baron-Cohen, S. (1995) *Mindblindness*, MIT Press
2. Connelly, J., Hong, W. S., Mahoney, R. B. Jnr., Sparrow, D. A., Current challenges in autonomous vehicle development, *Unmanned Systems Technology VIII, Proc. SPIE*, 6230, 62300D, (2006).
3. Easterbrook, S.M., Coordination Breakdowns: How flexible is collaborative work? In P. Thomas (Ed.), *CSCW: Requirements and Evaluation*, London: Springer-Verlag, (1996), pp. 91-106.
4. Epley, N., Keysar, B., Van Boven, L., Gilovich, T., Perspective taking as egocentric anchoring and adjustment. *J. Pers. Soc. Psychol.* 87 (3), (2004), 327–339.
5. Endsley, M.R. Theoretical underpinnings of situation awareness: a critical review. In: Endsley, M.R. and Garland, D.J., Editors, 2000. *Situation Awareness Analysis and Measurement*, Lawrence Erlbaum, Mahwah, NJ, (2000), 3–32.
6. Endsley, M., Toward a Theory of Situation Awareness in Dynamic Systems, *Human Factors*, vol. 37, no. 1, Endsley, M. (1995): pp. 32–64.
7. Gordon, R.M., Folk Psychology as Simulation, *Mind and Language* 1, (1986), 158–71
8. Grudin, J, Partitioning Digital Worlds: Focal And Peripheral Awareness In Multiple Monitor Use. In *Proc. CHI 2001*
9. Gutwin, C. and Greenberg, S. A descriptive framework of workspace awareness for real-time groupware. *Computer Supported Cooperative Work. The Journal of Collaborative Computing*, vol. 11, nos. 3–4, (2002)
10. Gutwin, C., and Greenberg, S. The importance of awareness for team cognition in distributed collaboration. In E. Salas, S. Fiore (Eds) *Team Cognition: Understanding the Factors that Drive Process and Performance*, APA Press, (2004), 177-201.
11. Heath, C., & Luff, P., Collaborative activity and technological design: Task coordination in London Underground control rooms. *Proc. ECSCW 91, Kluwer Academic Publishers, (1991), 65-80*
12. Hourizi, R. and Johnson, P. Designing to support awareness: a predictive, composite model. In *Proc. SIGCHI Conference on Human Factors in Computing Systems, CHI '04*. ACM, (2004), 159-166.
13. Kanno, T., Nakata, K., and Furuta, K. 2003. A method for team intention inference. *Int. J. Hum.-Comput. Stud.* 58, 4 (Apr. 2003), 393-413
14. Klein G, Woods D, Bradshaw J, Hoffman R, Feltovich P., Ten challenges for making automation a “team player” in joint human-agent activity. *Computer 2004*,:91–95.
15. Kraut, R. E., Fussell, S. R., & Siegel, J., Visual information as a conversational resource in collaborative physical tasks. *Human-Computer Interaction*, 18, (2003), 13-49
16. Mankoff, J., Dey A. K., Hsieh G., Lederer J., Ames S., M., Heuristic Evaluation Of Ambient Displays. In *Proc. CHI 2003*.
17. Miller, C., H. Funk, P. Wu, R. Goldman, J. Meisner, and M. Chapman, The playbook approach to adaptive automation. In *Proc. Human Factors and Ergonomics Society's 49th Annual Meeting*, (2005)
18. Panzarasa, P., N. R. Jennings, and T. J. Norman, Formalizing Collaborative Decision-Making and Practical Reasoning in Multi-Agent Systems, *Journal of Logic and Computation* 12 (1), (2002), 55–117.
19. Read, S. J., & Miller, L. C., Explanatory coherence and goal-based knowledge structures in making dispositional inferences, In B. Malle & S. Hodges. *Other Minds: How humans bridge the divide between self and others*, New York: Guilford Press, (2005), pp.124-142.
20. Scholl, B. J., & Tremoulet, P. D., Perceptual causality and animacy, *Trends in Cognitive Sciences*, (2000), 4 (8), 299-309.
21. Tomasello, M., Carpenter, M., Call, J., Behne T., and Moll, H., Understanding and sharing intentions: the origins of cultural cognition, *Behav Brain Sci* 28 (2005), 675–691
22. Watson, D. P. and Scheidt, D. H., Autonomous Systems, *Johns Hopkins APL Technical Digest*, 26(4), (2005).