Mimicking Human-Human Interaction: A Framework for Perceiving the User's Self-Expression in HCI

Authors blanked for blind review.

ABSTRACT

Most social interaction between humans can be conceptualized as a complex network of expression, perception and cognition. Through these means, we are able to manifest and communicate our inner states. The growing number of systems which learn and adapt to user attributes, states and preferences implies the usefulness of applying social communication theories to human-computer interaction. This argument is exemplified with the comprehensive application of an accepted theory of self-expression to existing systems. The result is a conceptual framework which defines a two-dimensional space, enabling the classification of specific interactions based on which type of social interaction is being mimicked. After exploring this space by interpretively mapping a few representative interactions, we suggest that the design of systems which simulate the social cognition of involuntary self-expression remains relatively unexplored.

Author Keywords

Frameworks, HCI, social psychology, communication.

ACM Classification Keywords

H1.2. User/Machine Systems: human factors, human information processing, software psychology.

A TAXONOMY FOR SOCIAL MACHINES

Our interests in engendering intelligence in machines have manifested in many forms of interactive media: examples include video games which simulate intelligent agents, robots which communicate with human users, and even mobile devices which adapt to context. A review of work in artificial intelligence demonstrates that computers are now capable of many processes often associated with intelligence, including awareness of context [3], autonomic self-management [9] and even large-scale detection and cognition of distributed phenomenon using sensor networks [16]. Work in affective computing has pushed a new understanding of how to model and portray emotion within interactive media [11, 12, 13]. The idea that machines may be able to think and feel like humans naturally leads one to question whether machines can interact on a social level, a curiousity which has persisted since the foundations of artificial intelligence and robotics (the Turing Test [15], and Masahiro Mori's "uncanny valley"[10]).

Critics of artificial intelligence tend to draw attention to its unfounded assumptions on the nature of human intelligence [6]). However, exploring practical applications does not make such an assumption, as the intention is merely to simulate intelligence. Whether truly intelligent or not, technological artifacts and research prototypes are able to learn more and more about users in their attempts to customize, personalize, recommend and predict, creating adaptive systems which tailor themselves for the contemporary user [14]. Development of robots which engage directly with humans has led to the success of large scale projects which center on creating robots designed specifically for social interaction [2, 7]. As opposed to demonstrating that interactive media can be truly intelligent and social on the same level as humans, the presented work simply suggests that computer systems are growing the ability to directly elicit social anthropomorphic reactions from us by simulating the perception of, and reaction to, our inner states.

Since the early literature on machine intelligence, our understanding of human social interaction and social cognition has developed substantially. Applying a conceptual framework of how humans interact with each other to the design of interactions between humans and technology is in accord with current thought on how to bring interactive media to the next level. Dourish places the embedding of computing into our social reality alongside that of our physical reality (tangibles) in his definition of embodied interaction [5]. Also, proponents of cognitive robotics suggest the "modeling of human-level cognitive faculties in robots" [4], which includes work on robots which exploit social-cognitive strategies [4]. Exploiting an understanding of our social reality is an important step toward new applications of technology which are embodied in our social environment.

CHI 2011, May 7–12, 2011, Vancouver, BC, Canada.
Copyright 2011 ACM 978-1-4503-0267-8/11/05...$5.00.

While many systems have already begun to interact with users socially, we have yet to establish a conceptual framework for how those interactions are able to detect and determine our inner states: a process humans constantly perform while interacting with each other. A taxonomy defining which types of social interaction are being simulated between technology and its users could identify gaps in our exploration of technological application, inspire correlative studies which define how users react to systems with respect to their social interaction mode, and define a new perspective for designers.

DETECTING HUMAN SELF-EXPRESSION

Systems which attempt to work in the realm of human social interaction tend to simulate perception and draw conclusions about our inner states through our behaviour. They are, in effect, detecting and responding to a type of signal known as "self-expression": behaviour which intentionally or unintentionally signals or shows an agent's thought, affect and experience [8]. The purpose of the present work is to apply the comprehensive framework of self-expression developed by Mitchell Green [8], a specialist in philosophy of mind, language and aesthetics who is focused on human-human interactions, to the classification of human-computer interactions (not on systems as a whole, but on specific interactions). Finally, by directionally limiting our analysis to user's self-expression toward the machine, we avoid discussing the inner states of machines. We shall proceed with the assumption that, while a human user may use strategies designed to express experience to other humans, the system's reception and reaction is simulated (as opposed to consciously experienced).

The application of Green's theories of self-expression to human-computer interaction provides us with the ability to analyze any user behaviour directed toward an interactive medium and map it in a two-dimensional space: one axis representing its how users express themsleves, and one axis representing how the user's inner state is being detected. For the present work, we have selected a few representative examples to be used for an exploration of how user interactions with existing systems fit on this space. As an introduction, consider our resulting two-dimensional space (see figure 1). The rest of this paper outlines how each dimension is defined and how each example was mapped to this space. Finally, we will discuss the implications of this tool.

X-AXIS: HOW USERS EXPRESS THEMSELVES

Green's first applicable conception of self-expression accounts for **involuntary**, **voluntary** and **voluntary-and-willed** self-expression [8]; a distinction which draws a clear dimension on which we can place technological interactions. Enabling the user to make explicit choices, either as system preferences or as decision points, is akin to eliciting users' *voluntary-and-willed* self-expression within the system's environment. Conversely, keeping track of regular system use and implicitly drawing conclusions about users' inner states is akin to exploiting a user's *involuntary* self-expression. These cases demonstrate how human-computer interactions can be mapped to the x-axis.

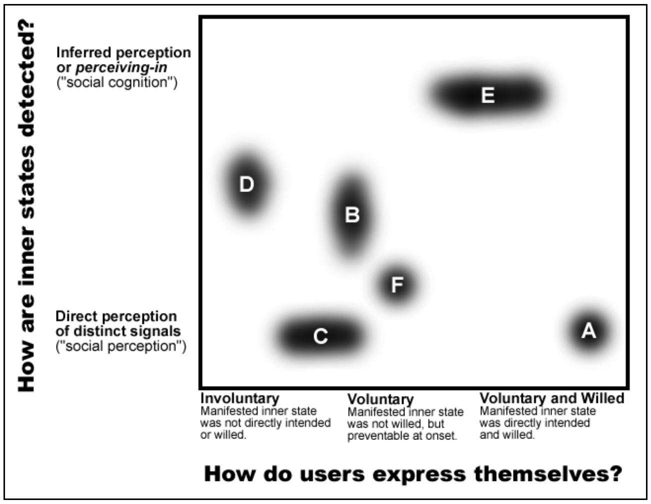
How are inner states detected?
Inferred perception or *perceiving-in* ("social cognition")
Direct perception of distinct signals ("social perception")
E
D
B
F
C
A
Involuntary
Manifested inner state was not directly intended or willed.
Voluntary
Manifested inner state was not willed, but preventable at onset.
Voluntary and Willed
Manifested inner state was directly intended and willed.
How do users express themselves?

Figure 1. A 2D space to which specific human-computer interactions can be mapped based on which types of human-human social interactions are being simulated.

Green draws the line between *involuntary* and *voluntary* self-expression by suggesting that, although self-expressions which we manifest without conscious thought may seem *involuntary*, they can still be considered *voluntary* if we can prevent them at the time of onset (for example, we naturally make loud noises when we walk, but this self-expression is still *voluntary* as we can choose to walk quietly if we do not wish to be heard) [8]. It is important to note that, while we adopt this distinction for our purposes, we do not consider the idea that one can disable features or turn off systems as rendering an expression preventable. As an example, consider a biofeedback system which tracks a user's heartbeat. This expression should be considered *involuntary* as the user can't really prevent his heart from divulging his inner state, despite the fact that he could turn off the system or remove the sensor at any time.

Y-AXIS: SYSTEM'S DETECTION OF INNER STATES

Green's second useful ontology of self-expression concerns his discussion of how those self-expressions are perceived. Green suggests that human emotional states are often directly perceivable by others [8]. He argues that humans construct opinions of other humans' emotion through the perception of manifest indices (body postures, behaviours, facial expressions, etc.), a process he claims is not much different from our perception of the material world:

"Someone who presents to me the surface of an apple from one angle has thereby shown me the apple even if I do not inspect its interior or its other side. The reason is that a sufficiently large portion of a side of an apple is, for normal human observers, not only itself perceptible but also a characteristic component of that apple. This amounts to the fact that under normal conditions, perception of part of an apple's surface is enough to justify me in inferring (if only unconsciously) the existence of the entire apple." [8]

It stands to reason that we may perceive emotion in much the same way, using characteristic behaviour to infer the existence of a particular emotion. Green calls this process *part-whole perception.* He continues by juxtaposing part-whole perception with *perceiving-in*: being "shown a thing A by sensing a distinct object B in such a way that we see (or... sense) A in B" [8]. The key difference here lies in the satisfaction criteria for a given inner state to be considered as perceived. In *part-whole perception*, the minute a characteristic component of the inner state is perceived, the observer determines the existence of said inner state. In *perceiving-in*, the medium B must provide evidence for the existence of said inner state. This type of perception can take place through a single signalling medium, as in the mailman raising the flag on a mailbox, or even be distributed in representation, as in determining that a colleague is in another room after hearing a non-distinct, distant sound, seeing a shadow, and catching a reflection in a mirror. While *part-whole perception* enables us to detect phenomena in our environment directly, *perceiving-in* facilitates the inference of conclusions based on comparative analysis of a wide range of factors (for instance, detecting an emotional state by a combination of tone of voice, choice of words and actions, body posture, vigilance, etc.). This distinction can be directly applied to human-computer interactions.

While some interactive media simulate the direct detection of a single element which is a characteristic component of an inner state (similar to *part-whole perception*), others infer user states through a parametrized analysis of user interactions in a more abstract and symbolic way (using cognitive processing to perceive inner states in the analysis of human-computer interactions: *perceiving-in*). Here, the focus is not on the sensors being used, but the relationship between the input data streams and the conclusions being drawn about the user's inner state. Systems can span the range from those which detect a characteristic component of the user's inner state (simulating **social perception**), to those which require the multivariate analysis of user interactions to interpret distributed inner states (simulating **social cognition**). For example, consider the difference between a user clapping to trigger an audio-activated light switch and a user speaking into a voice recognition system on his mobile phone to trigger a call. While the former simply detects the user's inner state (the desire for light) by the presence of a loud audio trigger, the voice recognition system carries out complex information processes and parametrization of an auditory data stream with a goal of extracting whose name is being spoken and removing noise, and then tying that information to the user's contact list in order to manifest his inner thoughts regarding who he wants to call. Similarly, any interaction where a computer system attempts to determine its users' inner states can be placed on a continuum between those which rely more on simulating *social perception*, and those which rely on simulating *social cognition.*

MAPPING EXAMPLE INTERACTIONS TO THE SPACE

A rigorous descriptive framework should include fairly regimented rules with respect to mapping, classification and categorization. However, the proposed taxonomy is in its infancy, presented to provoke discussion and break ground toward applying socio-cognitive theories to the classification of specific interactions. As such, our attempts to map example interactions are mostly interpretive, based on a heuristic interpretation of the definitions of each dimension presented above. What follows are brief descriptions of the reasoning behind how specific examples were considered with respect to our 2D space:

A: Changing message display preferences on Google's Gmail is an intended action which reflects a user's inner state. This interaction is clearly *voluntary-and-willed.* When the system detects the user changing the pagination option, for example, this interaction is a characteristic component of his/her inner need to change the number of messages on each screen. Thus, this interaction leans more toward *social perception.* One can imagine many standard desktop computing interactions mapping to this corner of the space.

B: Barrington et. al. [1] developed an interface which detects head and face motion to apply ambient effects to music. While the system's continuous observation of a user's motion suggests that this interaction lies closer to *involuntary*, the fact that head and face motion is preventable holds it near the center (*voluntary*). While the system uses an abstract extrapolation of user behaviour to determine an arousal parameter, the behaviours measured are simply motion; one can envision determining arousal to much more detail. As such, this system lies between *social perception* and *social cognition.*

C: Consider the example of a user approaching a door with the intent of passing through it. The user is intending to push the door when a sensor interprets his/her presence as a desire to open the door and opens it. This interaction lies between *involuntary* and *voluntary*, as this seemingly preventable behaviour would be difficult to prevent since the user is unaware of the sensor's existence. The user's presence in front of the door is a characteristic component of his intent to pass through the door, mapping this example toward *social perception.*

D: Apple's iTunes "Genius" feature is a system which analyzes all users' music libraries to determine a set of semantic metadata which can be used to generate new playlists based on a single chosen song. The system's gathering of data from a user's music library can be considered *involuntary*; once opted in, the system analyzes general use of the system to draw conclusions about users' musical tastes and preferences. Presumably, the system detects the co-occurence of songs and builds semantic links. This detection of a characteristic component of users' inner state of enjoying both songs must be further abstracted through a deeper analysis of variables such as the frequency of co-occurence, the time between both occurences, and how often this co-occurence is shared with other song pairs and other users. This balances the interaction between *social perception* and *social cognition*

E: Yohanan et. al. have been working on the Haptic Creature, a furry robotic creature which reacts to data from a wide variety of sensors to control animal-like behaviour, including purring, ear movement, and breathing [18]. The robot's sensors perpetually detect the user's affect through physical sensors and provide haptic feedback. The current version of the creature clearly requires direct petting; however, touch is a modality which will be naturally exploited by lifting and moving the creature, as well as resting a hand upon it. This suggests that physical interaction with the creature would lay somewhere between *voluntary* and *voluntary-and-willed*. The robot "uses a combination of the recognized gesture, recent time history and its own model to formulate a response" [18], suggesting that it clearly lies on the *social cognition* side of our y-axis.

F: Microsoft's anticipated Kinect platform uses physical and audio sensors to detect the user's motion and behaviour in front of the Xbox gaming console and screen. While interaction with many of the games are clearly *voluntary-and-willed*, the system's video chat program detects the user's position in order to keep the camera on track [17], which lies closer to the *voluntary* center of the spectrum in that it is automatic but preventable. This video chat interaction may use a complex algorithm, but it is essentially looking for a characteristic component of the user's inner desire to move, thus pushing it to the *social perception* side of the spectrum.

CONCLUSION

We can already clearly see advantages of applying theories of human-human interaction to HCI. Through the mapping of key examples, it seems that the majority of existing interactions would likely fill the bottom-left, bottom-right, and top-right areas, leaving the top-left quite open. The model is suggesting that existing systems are not yet taking advantage of inner states which can be determined through simulated social cognition of users' involuntary self-expression. It seems that, with further development of this taxonomy, we will be able to establish a clearer understanding of how users react to systems designed to elicit social anthropomorphization. The goal of this research direction is not only to inspire designers, but also to apply this understanding to improving usability, increasing user engagement and even exploring novel socially-charged applications for interactive media.

REFERENCES

1. Barrington, L., Lyons, M.J., Diegmann, D., and Abe, S. Ambient Display using Musical Effects. *11th Conf. on Intelligent user interfaces*, ACM (2006), 372-374.
2. C. Breazeal. Emotion and sociable humanoid robots. *Int. J. of Human-Comp. Studies*, 59 (2003), 119-155.
3. Carbonell, J.G., Siekmann, J., Kowalczyk, R., Müller, J.P., Tianfield, H., and Unland, R., eds. *Agent Technologies, Infrastructures, Tools, and Applications for E-Services*. Springer 2003.
4. Coradeschi, S., Ishiguro, H., Asada, M., et al. Human-inspired robots. *IEEE Intelligent Systems* (2006), 74-85
5. Dourish, P. *Where the action is : the foundations of embodied interaction*. MIT Press, 2001.
6. Dreyfus, H.L. *What Computers Still Can't Do: A Critique of Artificial Reason.* MIT Press, 1992.
7. R. Gockley, et al. Designing robots for long-term social interaction. *Proceedings of IROS 2005.*
8. Green, M. *Self-expression.* Clarendon Press; Oxford University Press, Oxford; New York, 2007.
9. Huebscher, M.C. and McCann, J.A. A survey of autonomic computing; degrees, models, and applications. *ACM Comput. Surv. 40*, 3 (2008), 1-28.
10. M. Mori, "Bukimi no tani {The Uncanny Valley}," (in Japanese) *Energy* 7, 4 (1970) 33--35.
11. Peter, C. and Beale, R. *Affect and Emotion in Human-Computer Interaction: From Theory to Applications.* Springer, 2008.
12. Picard, R.W. Affective computing: challenges. *Int. J. of Human-Comp. Studies 59*, 1-2 (2003), 55–64.
13. Picard, R.W. *Affective computing*. MIT Press, 1997.
14. Torre, I. Adaptive systems in the era of the semantic and social web, a survey. *User Modeling and User-Adapted Interaction 19*, 5 (2009), 433-486.
15. Turing, A. *The essential Turing: seminal writings in computing, logic, philosophy, artificial intelligence, and artificial life, plus the secrets of Enigma.* Clarendon/Oxford University Press; 2004.
16. Volosencu, C. Identification of distributed parameter systems, based on sensor networks and artificial intelligence. *WTOS 7*, 6 (2008), 785-801.
17. Xbox 360 Kinect Videokinect Demo at E3 http://www.youtube.com/watch?v=QBsTimrYTIQ
18. Yohanan, S. and MacLean, K.E. A tool to study affective touch. *27th int. conference on Human factors in computing systems*, ACM (2009), 4153-4158.