

A mind in a disk: the attribution of mental states to technological systems

Oronzo Parlange^{*}, Tommaso Chiantini and Stefano Guidi

Communication Sciences Department, University of Siena, P.zzo San Niccolò 56, 53100, Siena, Italy

Abstract. This paper reports a study about the role of different variables in the process of attributing mental states to technological systems, variables such as the number of figural elements displayed in the system and the personality traits of the subjects interacting with the systems. In an experiment, participants were interacting with a computer on whose screen several disks of various sizes and colours were blinking at different rates. Each time a disk reappeared on the screen its position was randomly varied. As in a videogame, participants had to click on the disks to increase their score. The results showed that, even in the case of such a simple system, subjects believed that the figural elements they were interacting with had some form of mental states, although their confidence in these beliefs varied in the different experimental conditions. The confidence level of the attributions, in fact, was not the same for all the different mental states considered, and it varied also both with the number of elements being displayed as well as with some personality traits of the subjects.

Keywords: mental states attribution, theory of mind, design of technological systems

1. Introduction

Seeing someone swearing at his or her computer is quite frequent. And we all have sometimes considered our computer with a suspicious mind, thinking of it as if it were aware of our intentions, and it willing to either meet or thwart them. Probably, in those moments, we have smiled or looked unkindly at the screen in front of us.

It is clear that all these behaviors cannot be considered properly rational. Human beings commonly interact with inanimate systems making use of an implicit knowledge of proper physical laws, and, in some cases that have been extensively investigated, through the consideration of rules that may be qualified as part of a naïve physics [3, 10].

The interaction with many mechanical systems, however, and particularly with information and communication technology, often seems to be based on other interpretative rules. It is well known that people tend to consider as human agents those systems that move and/or show some changes in even simple characteristics, such as shape, color, and size. This bias, that is the liability to consider human-made systems as if they were human beings, seems to depend on conceptualizing these systems as if they

were gifted with some self-generated and self-controlled cognitive ability.

It seems quite clear that this phenomenon involves one of the human tendencies that is probably amongst the most surprising and advantageous from an evolutionary point of view, namely the bias that brings us to attribute mental states, to elaborate a theory of mind [14, 4], to and for nearly all the entities with which we engage in some kind of interaction.

In the last years, the human tendency to anthropomorphize - in this context it could be said “mental-ize” - nearly everything, has gained increasing attention. For what concerns the hypotheses put forward to account for the way in which a theory of mind is developed since our birth, it is possible to identify two opposite positions, the one seeing this tendency as innate [1, 13, 12] and the one framing it as a competence that is structured mainly through actual experiences [9, 16].

In a comprehensive review of several studies, Kelemen and Carey [8] have highlighted that with respect to this tendency it is possible to trace some developmental stages starting from the early infancy and bringing the child, in the second year of his life, to attribute to the objects just one possible use, i.e.

^{*}Corresponding author. E-mail: oroparla@gmail.com

the one demonstrated by an adult. The development of this kind of competence goes on until the child is six-years old, a period in which children show themselves able to attribute to objects the intentions and the ideas of their designer [8]. This ability, however, does not seem able to completely overcome the bias to consider those systems that have a certain degree of complexity, and that are characterized by some particular features, as gifted with their own mind. Dennett [4] has put forward a theoretical frame in which three kinds of stances are identified to describe the way in which the interaction with any complex system may take place. The first is essentially based on physical laws, the second on the conceptualization of the system through the intentions of its designer, and the third is mediated by the attribution of intentional mental states.

There are several cues that can foster the attribution of mental states to non-human entities. Among these, the evident capacity of self-propulsion [2], the perception of a determined trajectory that, speaking of moving object, is the more effective the more it seems direct and pointing to a given target [5].

Morewedge et al. [11], in addition, showed that in order to see objects and animals as human beings, gifted with some mental states, these entities must preferably move at a speed that is similar to that generally exhibited by humans. Epley et al. [6] have also maintained that there are psychological determinants to the occurrence of anthropomorphism, i.e. the accessibility and applicability of anthropocentric knowledge, the motivation to understand the behavior of other agents, and the desire for social contact.

In the field of human-computer interaction, however, this issue has never received much attention. This in spite of the fact that understanding the way in which users elaborate a theory of mind for what concerns computer behaviors could be clearly very useful to design and implement more user-friendly technological systems.

Some studies have been conducted in order to investigate which determinants can induce the adoption of a theory of mind in relation to the behavior of some robots, technological systems that often, even in their appearance, can closely resemble human beings. In these cases too, the studies have generally supported the hypothesis that considers human beings more prone to the attribution of mental states if the interactive systems exhibit actions that are reactive to user behavior, and if their affordances can be more easily detected [17, 15].

Overall, however, there is still a surprising lack of knowledge about the phenomenon of mental states

attribution to artificial complex systems. So far, for instance, we do not know whether the attribution of mental states is an all-or-nothing process, or whether different mental states, such as intentionality and awareness, are seen linked together in the process of attribution of a mental entity. It is also actually unclear if some contextual variables, that are neither inherent to the user nor to the system, may affect the occurrence of such a phenomenon.

The study reported here is a first attempt aimed at finding some plausible answers to questions so relevant and complex like those ones.

2. The study

2.1. Method

2.1.1. The sample

An experiment has been conducted in which 136 subjects played a videogame specifically designed. They were 60 female and 76 male (mean age= 22,7 (sd=3,2)). About 16% of participants had a middle school diploma, 60.3% a high-school diploma, 17.6 % had a Bachelor degree and 5.9% a MS or MA degree. Most of the participants were currently students (70.6%), or worker-students (4.4%). About 2% (2.2%) were unemployed and the remaining 22.8% had a job.

Participants were all unaware of the real aims of the study.

2.1.2. Procedure and materials

Subjects were first asked to fill in a questionnaire in order to gather information on their sex, age, education, and employment. In addition, in this pre-test phase they were also asked to fill in a short questionnaire in order to assess their scores on five personality traits: agreeableness, openness to experience, extraversion, conscientiousness and emotional stability. We used an Italian translation¹ of the TIPI questionnaire - Ten Item Personality Inventory [7] -, which is a short, 10-item tool developed to measure the Big Five dimensions.

In the game specifically designed for this experiment some little disks of different colors and sizes were displayed on the computer screen. Disks remained visible for more or less short intervals: the smaller one for the shorter time and the larger one for the longer time, gradually.

¹ <http://www.psico.unipi.it/~psico/psico/psico.htm>

The position of the disks on the screen was randomly determined each time they were displayed. Subjects had to click on the disks in order to gain scores. Higher scores were associated to clicks on smaller, and faster, disks.

Subjects participated in the experiment individually, and they were randomly allotted to one of four different conditions (34 subjects per condition). In the first condition only two disks - that is the smallest one and the largest one - were, at a different pace, blinking on the screen. Increasing in steps of two, eight disks were present in the fourth condition (Figure 1).

In each condition the game lasted 80 sec.

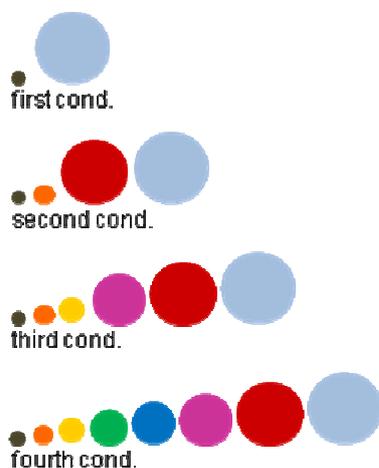


Figure 1
The four experimental conditions

In the first condition only two disks - that is the smallest one (2.5 mm) and the largest one (20.0 mm) - were, at a different pace blinking on the screen. The smallest was displayed for 0.8 sec, the largest for 2.2 sec. Increasing in steps of two, eight disks were present in the fourth condition. The size of the disks was increased in steps of 2.5 mm, and the time of permanence on the screen was respectively increasing of 0.2 sec.

The game started with a training phase during which subjects were told that their task was to click on the figures that were displayed on the screen. In the course of this training, subjects were in front of a white screen on which a black square was displayed for a lapse of time that was not longer than 3 sec. As a consequence of the click of the subject on the square, this figure was displayed in a different position.

This training phase lasted until the subject was fully acquainted with this simple task. Subjects were then informed about the actual rules of the game - i.e., the number of disks they would have seen on the

screen, the differences in the scores gaining, the game duration, and so on - and then the game started.

When the game was over, subjects had to fill in a questionnaire in which they had to report, on a seven-point scale, if they had thought the disks “had their own strategy”, “were aware of what was happening”, “had their own intentions”, and “had their own mind”. In addition, they were asked to express their degree of agreement (again on a seven-point scale) with the following control statements: “the movements of the disks were random”, “the game was regular”, and “some disks had more intentions than others”.

2.2. Results

In Figure 2 and 3 are displayed the mean ratings for each of the variables considered, averaged across experimental conditions.

As it can be seen, results show that, generally, even very simple objects like disks blinking on a screen, just changing their position, are seen as gifted with some mental states. In fact, the average ratings for all the variables relative to mental states were significantly higher than 1, the bottom value of the scale, corresponding to judgments of complete lack of mental states in the disks. However, it is also clear that different kinds of mental states were judged to differently belong to the disks. These were believed to have a strategy more than to have awareness or intentions, and the lowest ratings were relative to the disks having a mind.

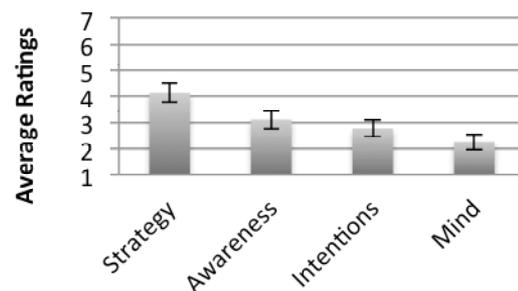


Figure 2
Average ratings for the different kind of mental states

We then looked at group differences in the ratings, conducting an ANOVA for each of the different dependent variables, using the experimental condition (number of disks: 2, 4, 6, or 8) as factor. The analyses found significant differences between the group means for two kind of mental states judgments: “Awareness” ($F(3,132)= 3,065; p<.05$), “Intentional-

ty” ($F(3,132)= 2,871$; $p<.05$). Significant differences between the group means were found also for the judgments relative to the disks having a “Random movement” ($F(3,132)= 3,129$; $p<.05$). In Figures 4, 5, and 6 are plotted the average ratings for these variables as function of the number of disks on the screen. As it can be seen, the condition with 6 disks was always the one with lowest ratings for both intentionality and awareness. Pairwise comparisons revealed that the judgments of awareness in this intermediate condition were significantly lower than in the condition with only 4 disks (average difference= -1.35; $p<.05$), and that judgments of intentionality in the 6 disks condition were significantly lower than in the condition with 8 disks (average difference= -1.26; $p<.05$). As for the judgments of “Random movement”, the average ratings in the 6 disks condition were instead significantly higher than in the 8 disks condition (average difference= 1.38; $p<.05$).

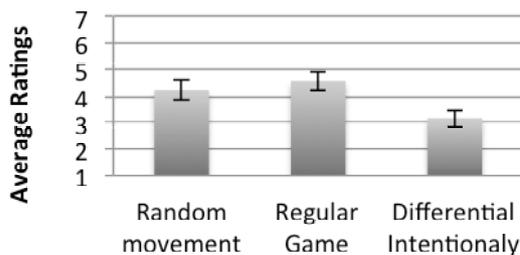


Figure 3
Average ratings for control variables.

We also found that for judgments of both “Awareness” and “Random Movement” of the disks, there was a significant interaction between the experimental condition and the sex of the participants (respectively $F(3,128)= 2,958$; $p<.05$ and $F(3,128)= 2,813$; $p<.05$). Follow-up tests of the simple effects of the number of disks revealed that for females, judgments of awareness in the 6 disks condition were significantly lower than in both the 2 disks condition (average difference= -2.1; $p <.01$) and in the 8 disks condition (average difference= -1.3; $p <.05$). On the contrary, for males, the judgments of “Awareness” of the disks in the 4 disks conditions were significantly higher than both in the 2 disks (average difference= 1.6; $p<.01$) and in the 6 disks condition (average difference= 1.3; $p <.05$). As for the judgments of disk moving randomly, we found that for female, the judgments for the 6 disks condition were significantly higher ($p<.01$) than in all the other conditions,

while for males we found significant differences only between the 2 disks and the 8 disks conditions, with participants giving higher rating of random movement to the 2 disks than to the 8 disks condition (average difference= 1.7; $p<.05$).

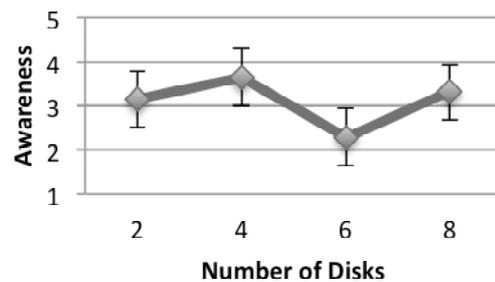


Figure 4
Average ratings of Awareness as function of Number of Disks blinking on the screen Error Bars are 95% CI.

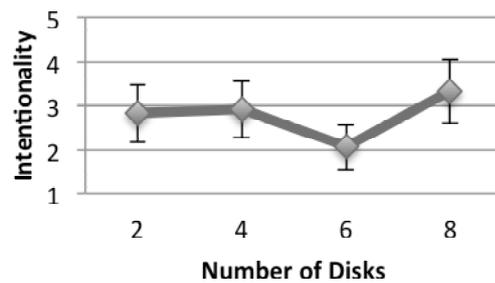


Figure 5
Average ratings of Intentionality as function of Number of Disks blinking on the screen. Error Bars are 95% CI.

Follow-up tests of the simple effects of the number of disks revealed that in the 2 disks condition, judgments of awareness of the disks were higher for females than for males (average difference= 2.3; $t(32)= 4.297$; $p<.0001$), while in the other conditions their judgments were not different. Consistently with this finding, in the 2 disks condition the judgments about Random Movement were instead higher for males than for females (average difference= 1.6; $t(32)= 2.426$; $p<.05$). Judgments of random movement, moreover, in the 6 disks condition were higher for females than for males (average difference= 1.3; $t(32)= 2.097$; $p<.05$).

Finally, we found that overall females tended to give higher ratings for “Mindfulness” of the disks than males (average difference= .6; $F(1,128)= 4,005$; $p<.05$).

We also looked at the correlations between the different mental state judgments, and found that all the bivariate correlations were highly significant, with the value of the Pearson's coefficients ranging from .37 to .54. Significant negative correlations were found between each of the judgments relative to the different kind of mental states and the judgments relative to the random movement of the disks (with coefficients ranging from -.26 to -.29).

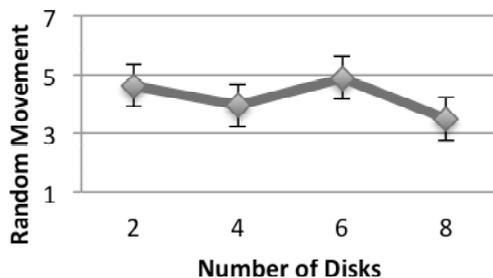


Figure 6

Average ratings of Random Movement as function of Number of Disks blinking on the screen. Error Bars are 95% CI.

Finally, we analyzed the correlations between the scores of the subject on the five personality traits (TIPI) and their judgments relative to the mental states of the disks. We found that the judgments relative to the disks having a mind tended to be lower the more the subjects were "Extravert" ($r = -.17$; $p < .05$), and higher the more they were "Agreeable" ($r = .18$; $p < .05$).

We also found a significant negative correlation between "Agreeableness" and the ratings relative to the disk moving randomly ($r = -.21$; $p < .05$). We also checked whether females and males differed in the scores of the five personality traits. Significant differences were found only for "Emotional stability", which was higher for males than for females (average difference = .99; $t(134) = 4.166$; $p < .0001$).

3. Discussion

The attribution of mental states does not seem to be an all-or-nothing judgment. Participants in our experiment showed themselves to be prone to suspect that the disks blinking on the screen had some kind of mind, with a degree of confidence that varied between the different kinds of mental states. And, on average, they never completely denied the existence of these mental states in the disks. But if participants

were quite uncertain of whether the disks had strategies (perhaps inherited from the intentions of the designer of the system), they were even less certain that disks had some kind of awareness or intentionality of their own. The strongest doubts were expressed about the disks having a full and proper mind, a complex and articulated entity comprising different mental states with a coordinated expression. The intentional stance and the design stance [4], thus, seem to be able to fade into each other: it looks like it is possible to believe that figural elements showing an autonomous behavior are more likely to move according to what their designer established, leaving at the same time open the possibility that the disks can have their own intentions and awareness. The strength of the doubts of the subjects seemed to vary with the kind of mental state considered, being milder for the ones that can be considered more simple, such as awareness, increasing in strength for mental states that require more articulated functions and processes, and being strongest of all about the existence of such a complex entity as a full-fledged mind.

Both personal level and contextual level variables seem to play a role in the attributions process, and perhaps also the interaction of these variables. This, for instance, is suggested by the findings relative to the third condition, the one in which the disks on the screen had intermediate speed and size, as often reported by the subjects themselves. In these cases, in fact, participants tended to judge with less confidence the disks as having awareness and intentionality. Judgments about awareness, then, seemed also to depend on personality traits such as extraversion. At the same time, the number of the disks present on the screen seems to affect females and males in a different way, especially in the first condition, the one with only two disks. In these cases, in fact, females seemed to judge the disks as having awareness to a greater degree than males.

At the light of our data, it seems quite hard to understand and disentangle the complex interactions between these variables. But even from the preliminary results here reported it seems possible to maintain that several different factors influence the process of attributing mental states, even in this case in which the target of the attributions are extremely simple elements such as colored disks blinking on a screen.

In certain cases, it is likely that those variables are interacting to induce the subjects to formulate an articulated theory of mind with respect to system he or she is interacting with. In other cases, the opposite seems to happen. But all of the variables here consi-

dered, i.e. the number of the elements displayed (and thus the complexity of the system), the personality traits of the observers, and the physical features of the elements, seem to influence the tendency to attribute mental states to the system. This tendency, as this study seems to confirm, seems to be a ubiquitous and unavoidable feature of human beings, and possibly one even more complex than what it is generally believed to be, which it is differently expressed for different mental states, and with respect to systems that does not have anything in common with humans.

References

- [1] S. Baron-Cohen, *Mindblindness: An Essay on Autism and Theory of Mind*, The MIT Press, Cambridge, Mas, 1995.
- [2] S. Baron-Cohen, The eye direction detector (edd) and the shared attention mechanism (sam): Two cases for evolutionary psychology,” in *Joint Attention: Its Origins and Role in Development*, C. Moore and P.J. Dunham (Eds.), Lawrence Erlbaum Associates, 1995, 3, pp. 41–59.
- [3] P. Bozzi, *Fisica ingenua*, Garzanti, Milan, 1991.
- [4] D.C. Dennett, *The Intentional Stance*, Bradford Books/MIT Press, Cambridge, Mass, 1987.
- [5] W.H. Dittrich and S.E.G. Lea, Visual perception of intentional motion, *Perception*, 23, 3, (1994), pp. 253–268.
- [6] N. Epley, A. Waytz and J.T. Cacioppo, On seeing human: A three-factor theory of anthropomorphism, *Psychological Review*, 114, 4, (2007), pp. 864–886.
- [7] S.D. Gosling, P.J. Rentfrow and W.B. Swann, Jr, A very brief measure of the Big-Five personality domains, *Journal of Research in Personality*, 37, 6, (2003), pp. 504–528.
- [8] D. Kelemen and S. Carey, The essence of artifacts: Developing the design stance, in S. Lawrence and E. Margolis (Eds.), *Creations of the Mind: Artifacts and their representation*, Oxford University Press, Oxford, 2007.
- [9] A.N. Meltzoff, Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children, *Developmental Psychology*, 31, 5, (1995), pp. 838–50.
- [10] M. Molina, G.A. Van de Walle, K. Condry and E.S. Spelke, The animate–inanimate distinction in infancy: Developing sensitivity to constraints on human actions, *Journal of Cognition and Development*, 5, (2004), pp. 399–426.
- [11] C.K. Morewedge, J. Preston and D.M. Wegner, Timescale bias in the attribution of mind, *Journal of Personality and Social Psychology*, 93, 1, (2007), 1–11.
- [12] D.I. Perrett and N.J. Emery, Understanding the intentions of others from visual signals: Neurophysiological evidence, *Current Psychology of Cognition*, 13, (1994) pp. 683–694.
- [13] D. Premack, The infant’s theory of self-propelled objects, *Cognition*, 36, 1, (1990), pp. 1–16.
- [14] D. Premack and G. Woodruff, Does the chimpanzee have a theory of mind?, *The Behavioral and Brain Sciences*, 4, (1978), pp. 515–526.
- [15] K. Terada, Y. Shamoto, H. Mei and A. Ito, Reactive Movements of non-humanoid robots cause intention attribution in humans. *Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems San Diego, CA, USA, Oct 29 - Nov 2, 2007*.
- [16] M. Tomasello, *The cultural origins of human cognition*, Harvard University Press, 1999.
- [17] E. Wang, C. Lignos, A. Vatsal and B. Scassellati, Effects of head movement on perceptions of humanoid robot behavior, *Proceeding of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, (2006), pp. 180–185.