Feeling committed to a robot: Why, what, when, and how?

Henry Powell
University of Glasgow

&

John Michael
(Warwick University and Central European University)

Forthcoming in Philosophical Transactions B
DOI: 10.1098/rstb.2018.0039

Abstract: The paper spells out the rationale for developing means of manipulating and of measuring people's sense of commitment to robot interaction partners. A sense of commitment may lead people to be patient when a robot is not working smoothly, to remain vigilant when a robot is working so smoothly that a task becomes boring, and to increase their willingness to invest effort in teaching a robot. We identify a range of contexts in which a sense of commitment to robot interaction partners may be particularly important.

Keywords: social robotics, commitment, joint action, coordination

1.*Why?*

As robots become increasingly prevalent in many domains of everyday life, from disaster relief to health care, education, and manufacturing [1,2,3,4,5], we can expect more and more research to be devoted to the optimization of human-robot-interaction. In this regard, it will be important not only to continuously improve safety and efficiency, but also to develop ways of making people feel comfortable and motivated to interact with robots.

While the primary focus of research in this context has been on identifying features of robots' appearance and behaviour which enhance their *likeability* [6,7,8,9,10,11,12], researchers have also explored other means of sustaining people's willingness to interact with robots. In particular, the concept of trust has received considerable attention in the field of social robotics in recent years [13,14,15]. It is easy to see why: trust is an important stabilizing force in relationships and in joint actions [16, 17]). And indeed, in some of the contexts in which people will be interacting with robots in the coming years, it could be greatly important that trust can be established and maintained. In healthcare, for example, it would be advantageous for people to be willing to rely on information or advice provided by a robot (such as doctor or a nurse), and to be comfortable divulging personal information to a robot without fearing that it will be handled without the appropriate care [18].

But, while trust is clearly a useful concept for social roboticists, it also has an important limitation. Specifically, trust helps only *indirectly* to sustain agents' motivation to contribute to joint action -- i.e., it stabilizes one agent's expectation that her partner will continue contributing to the joint action in a cooperative manner, and thereby reduces a source of uncertainty which could undermine the first agent's motivation to contribute. But it does not *directly* explain why that first agent would then herself contribute to the joint action in a cooperative manner. Indeed, she might be distracted or tempted to disengage irrespective of her level of trust in her partner.

In this brief opinion piece, we introduce a concept which is complementary to that of trust, and which overcomes the aforementioned limitation of trust: *the sense of commitment*. Our aim is to make the case that the concept of a sense of commitment may provide roboticists with a useful guide to developing effective, low-cost tools for designing robots that elicit patience and persistence and sustained motivation on the part of human interactants. We begin by briefly introducing the concept of a sense of commitment, distinguishing it from the concept of trust and explaining how it can be operationalized (Section 2: What?). We then identify a range of contexts in which the sense of commitment may be particularly important in human-robot interaction (Section 3: When?). Finally, we discuss possibilities for designing social robots which elicit a sense of commitment on the part of human users (Section 4: How?).

2. *What?*

The concept of a sense of commitment is closely related to the concept of trust, insofar as both refer to psychological mechanisms that help to sustain agents' motivation to contribute to joint action [19]. Unlike trust, however, the sense of commitment is linked *directly* to motivation. [20] hypothesize the sense of commitment as a mechanism which stabilizes agents' motivation to contribute to joint actions (and more generally to others' goals), leading them to persist and to remain focused in the face of tempting alternative options and distractions.

To illustrate the phenomenon we are focusing on here, consider the following example: Giuliana is an engineer working for a company that produces widgets. As it happens, she is also a highly skilled programmer, and sometimes helps colleagues out with IT issues, although this is not strictly speaking part of her job. On one occasion, just as she is about to head home for the day, she notices that her colleague Sam is struggling to complete his project on time because the computers at the office are running an outdated software program. She is the only person around with the requisite competence to resolve the issue, but it may take her considerable time to do so. Giuliana has no explicit commitment to help, but she may very well feel a sense of commitment to helping her colleague finish his task on time. And this sense of commitment may be enhanced by various situational factors: e.g. if Sam has already put a great deal of effort into the task, if she has helped him in the past with similar tasks and thereby created an expectation that she can be relied upon, etc. Such cases are highly common in everyday life, and it is important to be able to identify, prioritize, keep track of, and respond appropriately to our own and others' commitments in cases like that of Giuliana.

It is important to emphasize that there are a number of crucial differences between this *psychological* concept (the sense of commitment) and the *normative* concept of commitment. The normative concept refers to a relation among two agents and an action X, such that one agent has an obligation to some other agent to do X because she has intentionally expressed her willingness to do X under conditions of common knowledge, and this has been acknowledged [21,22,23,24,25]). The example of Giuliana helps to illustrate these differences. First, the sense of commitment can come in varying degrees -- i.e., it is a graded, not a binary, phenomenon. This is in contrast to the way in which theorists typically think about the normative concept of commitment -- i.e. as a relation which either obtains or does not. Second, whereas normative commitments are taken on intentionally (i.e. one deliberately makes a promise or gives an assurance because one is willing to take on the obligations which this implies), one may come to feel a sense of commitment to perform or continue performing a joint action without ever having intended to take on an obligation to do so. Third, whereas the normative concept of commitment requires that one actually believe that one has taken on certain obligations, the sense of commitment can be decoupled from beliefs about obligations -- i.e. one may feel a sense of commitment to performing a joint action even if one does not consciously believe that one has a commitment to doing so. In the example above, Giuliana may not believe that she is under any obligation to help, but she may nevertheless feel a sense of commitment which in fact leads her to do so. Indeed, she may feel committed even if she did not believe that she had any obligations at all with respect to Sam, e.g. because Sam were a robot.

These three features make it possible to operationalize the sense of commitment in terms of agents' motivation to perform actions. They also highlight the possibility that this motivation can be modulated by cues that one's partner values the joint action and may be relying on one to make one's contribution [20]. On this basis, recent research has begun to investigate the cues and situational factors that may trigger/modulate such a sense of commitment. One recent study [26], for example, revealed evidence that a high degree of *spatiotemporal coordination* within joint action (i.e. the two agents form a chain in cleaning up a pile of sand, with one agent scooping and the other pouring) can engender a greater sense of commitment than would be present if there were only a low degree of spatiotemporal coordination (i.e. the same two agents do not form a chain but, rather, work in parallel). This leads observers of the joint action to expect both agents to remain engaged in the joint action for a longer time and to be more likely to persist until the goal is achieved. Similarly, it has been [19] demonstrated that *coordinated decision-making* is sufficient to elicit a sense of commitment, leading agents to resist tempting alternatives and thereby contribute to sustaining cooperation through fluctuations in individuals' interests (Cf. [27], [28]). In a related strand of research, one recent study [29] found evidence in support of the hypothesis that the *perception of a partner's effort* elicits a sense of commitment to joint action, leading to increased persistence in the face of a temptation to disengage (see also [30]).

Of course many other factors may also give rise to or enhance a sense of commitment: in general, we may expect that one's sense of commitment can be modulated by any cue that another agent expects one to perform a particular action, by the knowledge that one has led them to form that expectation, by the knowledge that they are relying on that expectation, and/or by the expectation of reciprocity from the other party [20].

Building upon this recent research investigating situational cues that can elicit a sense of commitment in human-human joint action, we propose that some of the very same situational cues could be implemented in HRI. In the next section, we survey a broad range of contexts in which this could be particularly useful.

3. *When*?

In what contexts might a sense of commitment towards performing a task with a robot be particularly beneficial? We suggest that this would be particularly useful in any contexts in which there is a risk that a human interactant may partially or fully *disengage* from a task that s/he is performing together with a robot partner, and when this is likely to have undesirable consequences. In this section, we will survey a range of situations in which the risk of disengagement is particularly high, and discuss the kinds of undesirable consequences which may be expected in such situations.

In general, human disengagement may be an acute risk either because a robot is performing poorly (i.e. because the robot makes errors, performs more slowly than a human expects or is accustomed to, needs to stop and recalibrate, etc.) or indeed because it is performing so well that the human ceases to pay close attention to the task. Of course, there are a great many further factors which may lead people to become distracted or disengage from a task; in the following, we will focus on robot underperformance and robot overperformance because we believe that these two factors have been neglected in the literature. Let us first consider the former type of case: human disengagement in response to robot *underperformance*. We will then turn our attention to the threat of disengagement arising in response to *overperformance*.

Disengagement arising in response to underperformance is familiar to anyone who has struggled to remain patient as her laptop loaded slowly, or had to exercise self-control to refrain from smashing a phone that deleted an important picture or email. There are at least four reasons why roboticists may be particularly concerned about people's reactions to robot underperformance. First, robot underperformance is likely to be frequent, at least in the medium term. Dynamic, real-world scenarios will put strains on robots' capabilities, especially in more complex joint actions that unfold over longer time periods and involve shifting goals and strategies. Secondly, people may in general have unrealistically high expectations about the capabilities of robots (especially those with a more human-like appearance), and therefore perceive robots to be underperforming even when this is not the case. In one study [31] for instance, it was shown that people judge a robot to be more capable when that robot is able to perform conversational speech. This implies that there is more room for disappointment when a robot appears human-like and thereby raises the expectation of a smooth and effective interaction. Interestingly, [31] also showed that people felt robots with conversational speech were less capable after a failure than robots who did not have such conversational functionality. Thirdly, there is a further reason why people may incorrectly perceive robots to be underperforming -- i.e. people are typically unfamiliar with the workings of robots, and may sometimes misjudge that a robot has broken down when in fact it is taking longer to perform an action than expected simply because it is recalibrating or exploring the learning space [32]. Fourth, there may be important negative consequences of human disengagement in response to (perceived) robot underperformance.

The negative consequences of human disengagement may range from opportunity costs (e.g. the human decides to go to a different shop or service provider where s/he can interact with a reason person), to subpar productivity in industry settings involving human-robot interaction, to health and safety risks if the human interactant ceases to pay attention to the task in a potentially dangerous environment.

Of course, being able to create robots to whom people feel committed will be useful in the same situations in which having a sense of commitment to a human would be useful. This is especially true in situations where the interaction has a high-reward on success/ high-damage on failure payoff structure. Consider the example of a doctor-patient relationship. If I break my leg and am told by a doctor that I need to follow a course of physiotherapy, a sense of commitment towards that doctor (or towards my teammates if I am an athlete) might lead me to persist on course with this process even when it became difficult or tedious. If I had no sense of commitment towards the doctor (or to my teammates), then I might skip physiotherapy sessions or perform careless actions that risk further injury [33]. Given the increase in the use of robots in hospitals and care homes [34, 35, 36], being able to create robots that are able to emulate and elicit this kind of commitment will play a significant part in the successful integration of robotics platforms into the healthcare services.

Finally, in the service industry, having customers that feel a sense of commitment to a service robot will also be of considerable benefit. This is illustrated by the following example. Jack goes into a clothes shop and is greeted by a service robot who asks him what he is looking for. Once Jack has told the robot that he is looking for a new sweater, the robot leads Jack to the appropriate section and reaches for the desired item on the top shelf (beyond Jack's reach) where it knows these sweaters are kept. However, the sweaters have since been moved across the shelf by a member of staff. The robot had not registered this and has to reassess the environment and search for the new location of the sweaters. This leads to a delay. Since Jack can see where the sweaters have been moved to, he may become impatient and frustrated, and consider switching to a different shop with human assistants. Of course, this situation is potentially damaging for the reputation of the robot and for the reputation of the shop. If the robot had performed its action in such a way as to establish a sense of commitment between itself and Jack, however, Jack may have resisted the temptation to disengage and seek human help, thus mitigating the damage to the reputation of the store's quality of service.

It should be acknowledged that error-prone robots may also be endearing and evoke empathy in virtue of their errors, leading to *increased* patience on the part of human interactants. In fact, there is evidence pointing in this direction from one study [10] showing that participants actually favoured robots who made mistakes, as this made them more human-like. However, the point stands that many customers in a retail scenario may be more impatient and more demanding when they are shopping for something than when they are interacting with a robot in an experimental scenario.

Let us now consider a different type of case: human disengagement in response to robot *overperformance*. While we are not aware of any research directly investigating this possibility in the context of human-robot interaction, there has been relevant research in the context of human reliance on automated systems [37]. For example, it has been shown that pilots' excessive trust in or overreliance on autopilot systems can lead to dangerous monitoring failures [38]. Similarly, many road vehicles already contain computational systems that take control away from the driver in order to limit the inconveniences of driving (cruise control, automatic parking and so on). However, many of these systems still require that the human driver take back control in situations where the system is incapable of making safe decisions by itself. Take the case of cruise control in the following example. Claire is driving her car on a long trip across the country. Whilst on a long motorway stretch she switches on advanced cruise control and takes her hands off the steering wheel and her foot off the accelerator. She knows from experience that the car is able to slow down when appropriate and to modulate its speed effectively. The robotic system is performing its action so well that Claire loses her focus and neglects her own part of the joint action: namely, staying alert to any sudden changes in the environment (a child walking out into the road for example) that would require her to take control of the wheel and to brake or turn to avoid disaster. In this situation, had Claire been committed to the task with her robotic partner, she may have resisted the temptation to relax and to daydream, and remained engaged in the task at hand. This example generalises to other contexts where one's losing focus on a joint action due to overestimation of a robot's capabilities may entail a high risk. A human agent's lack of commitment to a joint task in a factory with a robot partner might lead that human to become uninterested or to lose focus [39]. In certain industrial contexts, this holds risks relating to potential injury to the person or damage to the robot, especially since workers are increasingly working in closer contact with their robot partners [40,41]. If, on the other hand, the human feels a sense of commitment towards the task it is performing with a robot partner, she may remain engaged in that task, thereby limiting the extent to which such injuries or damage might come about.

We would speculate that monitoring failures arising in response to robot overperformance may be particularly likely to occur in human-robot interaction settings in which a human's task is monotonous and/or where the human must be vigilant to pick up on infrequent events. It is known, for example, that the more infrequent an event is to which a human must react, the more likely it is that mistakes will occur [42,43,44]. This may be a particularly likely and dangerous possibility in factory contexts. If it is possible for humans to sustain their attention to such tasks by eliciting a sense of commitment, this could have important positive consequences.

Might there also be risks in designing robots to elicit a sense of commitment on the part of human interactants? Some recent research motivates the conjecture that a sense of commitment in joint action can lead agents to comply more readily with their partner's antisocial requests, e.g. to put sow bugs into the coffee grinder, or to produce annoying sounds to distract a competitor. Similar phenomena have also been observed in human-robot interaction: one study [45] indicated that human participants were more inclined to comply with a robot's request to throw books into a rubbish bin when the robot was physically present than when the robot was merely telepresent. Another recent study [46] found that participants followed a robot during a fire evacuation procedure even when they had observed that same robot making errors in a navigation task just before the evacuation happened. These findings motivate the conjecture that increasing a human's sense of commitment to an interaction with a robot could under some circumstances lead the human to be willing to perform actions that are not in her/his best interests. This risk could be especially serious in military or rescue contexts if humans were led to endanger themselves or others out of a sense of commitment to a robot interactant.

In this section, we have specified a range of contexts in which it would be useful – or dangerous – to be able to elicit a human's sense of commitment to performing a task together with a robot partner (See **Table 1**). In the next section, we turn to the question of how this may be achieved.

[Table 1 here]

4. *How?*

It is obviously important to understand *how* a sense of commitment might be triggered and modulated in human-robot interactions. In this section, we outline a number of ways in which this may be possible. Here we draw both on data that we have already collected, and also point to potential answers to this question from other strands of research from neuroscience, computing science, and robotics.

Of course, the most straightforward possibility is to use explicit verbal communication. For example, the robot could indicate what it expects a human to do and/or make explicit that it is relying on the human to perform a particular contribution. More indirectly, it could express its gratitude at the human's contribution and thereby reinforce the human's commitment, or indicate its conditional willingness to reciprocate in the future. Turning to the other side of the interaction, a robot could elicit an assurance from a human: just having made a promise or an agreement with a robot could lead a human to feel more committed to performing her contribution.

While we often make commitments explicitly, through the use of contracts or spoken promises for instance, it is likely that one way in which a *sense* of commitment might be communicated is through implicit forms of communication, often referred to as social signals. Social signals constitute a range of behaviours: from changes in posture [47, 48, 49], and gaze [50, 51] to changes in the non-verbal elements of an interactor's speech [52, 53] and are known to communicate a large amount of information about the agents involved in an interaction including their emotions, social status, and likely future behaviors (see [54, 55] for review). But what kinds of implicit communicative behaviours or social signals might be used to communicate that a person or a robot is committed to a joint task?

One answer to this question is provided by [29, 56]. In one recent study [29], participants played a joint version of the classic snake game where over 20 rounds they were to collect as many apples as possible with a partner. Within each round the apples appeared at an increasingly slow rate diminishing the value and increasing the tedium over time. Before each round participants observed what they believed to be a human partner (but was actually a computer algorithm) complete either a short, easy captcha or a longer, more difficult one in order to unlock the round that would follow. The data showed that participants spent significantly longer playing the rounds following the longer captchas. A recent study [56] extends this research to the realm of human-robot interaction, showing the same effect when participants were led to believe that they were playing the same game with the humanoid robot iCub [57]. This finding indicates that the perception of a robot partner's apparent investment of cognitive effort might be one important social signal that could be used to trigger commitment to a given task.

But what other ways might a sense of commitment be communicated? An obvious alternative to cognitive effort is physical effort, where the latter is understood to be the expending of physical energy through movement. Perhaps then commitment could be communicated through the way we move. This suggestion is motivated by research on intention communication in human-human interaction. In the last 30 years human neuroscience research has shown that our intentions have noticeable effects on our movements [58, 59, 60, 61]. In one study [60], it was shown that the velocity and orientation of our hands during a reach to grasp movement towards a bottle of water will vary as a function of whether we intend to lift it or to pour it. Other recent studies [62, 63, 64] have extended this research to show that these subtle changes in the kinematics of our movements extend to reflect social intentions such as when we pass an object in a competitive or cooperative scenario. Significantly, one study [64] showed that our perceptual systems are acute enough to pick up on these differences when the actions in question are performed by a social partner. Participants were able to distinguish between the goals of two visually truncated movements that only differed with respect to their kinematics.

One might contest here that this evidence does not directly show that humans are sensitive in the same ways to the movements of robots. However, a number of studies [65, 66, 67] show that adults and children are able to ascribe emotional states to robots that match the intended emotional state programmed into the robots by the experimenters. Further, it has also been shown [68, 69, 70] that we are able to predict which part of a scene a robot is attending to, and also [71, 72] that we are able to accurately assess a robot's workload (i.e. how heavy an item that it is lifting is based on its kinematics). Taken together, this evidence gives us good reason to expect that commitment to a task might be communicated in a similar fashion. For example, it might be clear to a person that a robot is committed to a task when that robot trades effort (as energy expenditure) for gains in movement legibility or predictability. This could be done by choosing an end-effector trajectory that makes it clear to the human what the goal location of that movement will be as opposed to one that is more efficient but less communicative (a rectilinear as opposed to a curvilinear path, for example) (for a computational model of such communicative movements see [73]). Indeed the modulation of commitment via increased spatio-temporal coordination [26] could be explained by each person's perceiving that their partner was investing increased physical effort into the task at hand. Thus, we propose that a worthy avenue for future research would be to investigate how changes in movement styles may be sufficient to communicate that one is committed to the task at hand, and also to trigger and modulate a sense of commitment in one's partner.

There are a number of potential problems with wanting to implement commitment functionality into a robot from the perspective of robotics and computing science given that an important goal in these fields (at least in social signal processing) is to utilize research in the cognitive neurosciences and apply it to the development of robotic platforms that are designed to act alongside humans. For cognitive effort, for instance, one might feasibly take some measure of an interactor's cognitive load over the course of a task. This is because an increase in one's cognitive effort should presuppose an increase in cognitive load. This presents a complication because one of the principal ways in which cognitive load is measured is through pupillometry [74, 75, 76, 77], which is a difficult variable to measure in conditions that are not very tightly controlled. This means that in everyday situations where robots will be interacting with humans in changing lighting conditions and with a large numbers of distractors it will be not be clear that the changes in pupil diameter that might be observed by a robot would reliably indicate a change in cognitive load.

Investigating how a sense of commitment can be communicated through *low-level features of our movements*, on the other hand, has the advantage of being more plausibly implemented. Detecting commitment as a function of physical effort would rely on the extraction of kinematic features of a partner's movement, such as their pose, effector trajectories, effector velocities and so on. There already exists a wealth of computational methods for extracting this kind of data using video analysis techniques [78, 79] (for review see [80]) with recent work using deep neural networks even going so far as to be able to accurately estimate the pose and movements of humans in busy and dynamic natural environments [81].

5. Conclusions and Further Directions

We have attempted to make the case that the concept of a sense of commitment may provide roboticists with a useful guide to designing and developing effective robots that elicit patience and persistence and sustained motivation on the part of human interactants. We have argued that such a sense of commitment may be particularly useful in contexts in which a human interactant is likely to disengage -- because her robot partner makes mistakes, because it performs slowly (while learning or recalibrating), or because it is performing so well that the human loses focus. Finally, we have identified a set of tools that may be used to implement design features for robots in different contexts to enable them to elicit a sense of commitment on the part of a human partner.

If we are right in thinking that there is great potential in the development of design features that serve to maintain a human's sense of commitment to an interaction with a robot, then it would also be very useful to design robots that are able to assess a human's sense of commitment. This would enable them to increase commitment-generating signals in order to boost the human's sense of commitment when necessary. Such an integrated approach should address the three modules that need to exist in a robot in order to succeed in dynamic interaction: perception, reasoning and action [82]. One starting point may be to measure pupil dilation as a means of assessing the human's engagement in the task [83], although it is not yet clear how pupil dilation could be measured in such a way as to provide a sufficiently reliable measure in noisy real-world environments. In some contexts, it may be possible to measure the human's gaze direction to assess her attention to task-relevant aspects of the situation [84]. A further possibility may be to harness software enabling face-based or voice-based [85] emotion recognition: if the human's emotion responses no longer track the progress of the task or reflect task-relevant events, this may be taken to indicate disengagement from the task [86].

References

1. Breazeal C, Brooks A, Gray J, Hoffman G, Kidd C, & Lee H. 2004. Humanoid robots as cooperative partners for people. *Journal of Humanoid Robots*;34.

2. Lenz C, Nair S, Rickert M, Knoll A, Rosel W, & Gast J. Joint-action for humans and industrial robots for assembly tasks. *In IEEE; 2008*. p. 130–5. Available from: http://ieeexplore.ieee.org/document/4600655/

3. Clodic A, Cao H, Alili S, Montreuil V, Alami R, & Chatila R. SHARY: A Supervision System Adapted to Human-Robot Interaction. In: Khatib O, Kumar V, Pappas GJ, editors. *Experimental Robotics*. Springer Berlin Heidelberg; 2009. p. 229–38.

4. Sciutti A, Bisio A, Nori F, Metta G, Fadiga L, & Pozzo T. Measuring Human-Robot Interaction Through Motor Resonance. *International Journal of Social Robotics*. 2012 Aug;4(3):223–34.

5. Grigore EC, Eder K, Pipe AG, Melhuish C, & Leonards U. Joint action understanding improves robot-to-human object handover. In: *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*. 2013. p. 4622–9.

6. Wainer J, Feil-seifer D, Shell D, & Mataric M. The role of physical embodiment in human-robot interaction. *In IEEE; 2006*. p. 117–22. Available from: http://ieeexplore.ieee.org/document/4107795/

7. Bainbridge WA, Hart J, Kim ES, & Scassellati B. The effect of presence on human-robot interaction. *In IEEE; 2008*. p. 701–6. Available from: http://ieeexplore.ieee.org/document/4600749/

8. Lee MK, Kiesler S, & Forlizzi J. 2010. Receptionist or Information Kiosk: How Do People Talk With a Robot? In: *Proceedings of the 2010 ACM conference on Computer supported cooperative work*. p. 31-40.

9. Brooks DJ, Begum M, & Yanco HA. Analysis of reactions towards failures and recovery strategies for autonomous robots. In: *2016 25th IEEE International Symposium on Robot and Human Interactive Communication* (RO-MAN). 2016. p. 487–92.

10. Mirnig N, Stollnberger G, Miksch M, Stadler S, Giuliani M, & Tscheligi M. 2017. To Err Is Robot: How Humans Assess and Act toward an Erroneous Social Robot. *Frontiers in Robotics and AI*. May ;4. Available from: http://journal.frontiersin.org/article/10.3389/frobt.2017.00021/full

11. Hamacher A, Bianchi-Berthouze N, Pipe AG, & Eder K. 2016. Believing in BERT: Using expressive communication to enhance trust and counteract operational error in physical Human-Robot Interaction. In: *Robot and Human Interactive Communication (RO-MAN), 2016 25th IEEE International Symposium on* (pp. 493-500). IEEE.

12. Bing, R, & Michael, J. Overcoming The Uncanny Valley Through Shared Stressful Experience with a Humanoid Robot. *Journal of Emerging Investigators*. (n.d.).

from https://www.emerginginvestigators.org/articles/overcoming-the-uncanny-valley-through-shared-stressful-experience-with-a-humanoid-robot

13. DeSteno D, Breazeal C, Frank RH, Pizarro D, Baumann J, Dickens L, & Lee JJ. 2012. Detecting the trustworthiness of novel partners in economic exchange. *Psychological science*, *23*(12), 1549-1556.

14. Lee JJ, Knox & Breazeal C. 2013. Modeling the Dynamics of Nonverbal Behavior on Interpersonal Trust for Human-Robot Interactions. In: *Trust and Autonomous Systems: Papers from the 2013 AAAI Spring Symposium*.

15. Lee JJ, Knox B, Baumann J, Breazeal C, & DeSteno D. 2013. Computationally modeling interpersonal trust. *Frontiers in psychology*, *4*, 893.

16. Berg J, Dickhaut J, & McCabe K. 1995. Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122-142.

17. Christ MH, Sedatole KL, Towry KL, & Thomas MA. 2008. When formal controls undermine trust and cooperation. *Strategic finance*. 89(7), 39.

18. Wilkes DM, Franklin S, Erdemir E, Gordon S, Strain S, Miller K, & Kawamura, K. 2010. Heterogeneous artificial agents for Triage nurse assistance. In *Humanoid Robots (Humanoids), 2010 10th IEEE-RAS International Conference on* (pp. 130-137). IEEE.

19. Chennells M, Letesson C, Székely M, Wozniak M, Lindeløv J, Butterfill S, & Michael J (under review). The Chains of habit: Coordinated decision-making boosts cooperation by eliciting a sense of commitment

20. Michael J, Sebanz N, & Knoblich G. 2016. The Sense of Commitment: A Minimal Approach. *Frontiers in Psychology*. 6. 1968.

21. Austin JL. 1962. *How to do things with words: The William James lectures.* Cambridge, MA.

22. Searle J. 1969. *Speech Acts: An Essay In The Philosophy of Language*. Cambridge: Cambridge University Press.

23. Gilbert M. 1989. *On Social Facts*. Princeton: Princeton University Press.

24. Bratman M. 1992. Shared cooperative activity. *Philosophical Review*. 101. 327–341.

25. Scanlon T. 1998. *What we owe to each other*. Harvard University Press.

26. Michael J, Sebanz N, & Knoblich G. 2016. Observing Joint Action: Coordination Creates Commitment. Cognition. Dec;157:106–13.

27. Guala F, & Mittone L. 2010. How history and convention create norms: An experimental study. *Journal of Economic Psychology*, *31*(4), 749-756.

28. Rusch H, & Luetge C. 2016. Spillovers from coordination to cooperation: Evidence for the interdependence hypothesis?. *Evolutionary Behavioral Sciences*, *10*(4), 284.

29. Székely M & Michael J. 2018. Investing in commitment: Persistence in a joint action is enhanced by the perception of a partner's effort. *Cognition*;174:37–42.

30. Chennells M. and Michael J. (In press, *Nature: Scientific Reports*). Effort and performance in a joint action are boosted by the perception of a partner's effort.

31. Cha E, Dragan A, & Srinivasa S. 2015 Perceived robot capability. In: *24th IEEE International Symposium on Robot and Human Interactive Communication*, Kobe, Japan, Aug. 1.1.

32. Kim T, & Hinds P. 2006. Who should I blame? Effects of autonomy and transparency on attributions in human-robot interaction. In: *Proceedings of the 15th IEEE International Symposium on Robot and Human Interactive Communication*.

33. Michael J, Salice A. 2017. The Sense of Commitment in Human–Robot Interaction. *International Journal of Social Robotics*. Nov;9(5):755–63.

34. Özkill A, Fan Z, Dawids S, Aanæs H, Kristensen JK, & Christensen KH. 2009. Service Robots for Hospitals: A Case Study of Transportation Tasks in a Hospital. In: *Proceedings of the IEEE International Conference on Automation and Logistics*. Aug.

35. Broadbent E, Stafford R, & MacDonald B. 2009. Acceptance of Healthcare Robots for the Older Population: Review and Future Directions. *International Journal of Social Robotics*. Nov;1(4):319–30.

36. Broekens J, Heerink M, & Rosendal H. 2009. Assistive social robots in elderly care: a review. *Gerontechnology*. Apr 1;8(2). Available from: http://gerontechnology.info/index.php/journal/article/view/1011

37. Parasuraman R, & Riley V. 1997. Humans and Automation: Use, Misuse, Disuse, Abuse. Human Factors: *The Journal of the Human Factors and Ergonomics Society*. 39(2), 230–253. https://doi.org/10.1518/001872097778543886

38. Mosier K, Skitka LJ, & Korte KJ. 1994. Cognitive and social psychological issues in flight crew/automation interaction. In Mouloua M, Parasuraman R. *Human performance in automated systems: Current Research and Trends*. Hilsdale.

39. Székely M. Michael J. Perception of a partner's effort boosts attentional control in a joint action. (in preparation).

40. Matthias B, Kock S, Jerregard H, Kallman M, & Lundberg I. 2011. Safety of collaborative industrial robots: Certification possibilities for a collaborative assembly robot concept. In *IEEE; 2011*. p. 1–6. Available from: http://ieeexplore.ieee.org/document/5942307/

41. Fryman J, & Matthias B. 2012. Safety of Industrial Robots: From Conventional to Collaborative Applications. *Robotik*:5.

42. Smallwood J, Davies JB, Heim D, Finnigan F, Sudberry M, O'Connor R, & Obonsawin M. 2004. Subjective Experience and the Attentional Lapse: Task Engagement and Disengagement During Sustained Attention. *Consciousness and Cognition*. 13(4), 657-690.

43. Manly T, Lewis GH, Robertson IH, Watson PC, & Datta AK. 2002. Coffee in the Cornflakes: Time-of-Day as a Modulator of Executive Response Control. *Neuropsychologia*. 40(1), 1-6.

44. Manly T, Robertson IH, Galloway M, & Hawkins K. 1999. The Absent Mind: Further Investigations of Sustained Attention to Response. *Neuropsychologia*. 37(6), 661-670.

45. Bainbridge WA, Hart J, Kim ES, & Scassellati B. 2008. The effect of presence on human-robot interaction. *In IEEE; 2008*. p. 701–6. Available from: http://ieeexplore.ieee.org/document/4600749/

46. Robinette P, Li W, Allen R, Howard AM, & Wagner AR. 2016. Overtrust of robots in emergency evacuation scenarios. *In IEEE; 2016*. p. 101–8. Available from: http://ieeexplore.ieee.org/document/7451740/

47. Coulson M. 2004. Attributing emotion to static body postures: Recognition accuracy, confusions, and viewpoint dependence. *Journal of nonverbal behavior*. .28(2):117–139.

48. Van den Stock J, Righart R, & de Gelder B. 2007. Body expressions influence recognition of emotions in the face and voice. *Emotion*. 7(3):487–94.

49. Scheflen AE. 1964. The Significance of Posture in Communication Systems. *Psychiatry*. Nov 1;27(4):316–31.

50. Poggi I, & Vincze L. 2008. Persuasive gaze in political discourse. In: *AISB 2008 Convention Communication, Interaction and Social Intelligence*. p. 55.

51. Vertegaal R, Slagter R, Van der Veer G, & Nijholt A. 2001. Eye gaze patterns in conversations: there is more to conversational agents than meets the eyes. In: *Proceedings of the SIGCHI conference on Human factors in computing systems.* ACM; p. 301–308.

52. Scherer KR. 1979. Personality markers in speech. In K.R. Scherer and H. Giles, editors, *Social markers in speec*h. Cambridge University Press.

53. Scherer KR., & Scherer U. 1981. Speech behavior and personality. In J. Darby, editor, *Speech evaluation in psychiatry*. Grune & Stratton, Incorporated.

54. Vinciarelli A, Pantic M, Heylen D, Pelachaud C, Poggi I, & D'Errico F. 2012. Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transactions on Affective Computing*. Jan;3(1):69–87.

55. Vinciarelli A, Pantic M, & Bourlard H. 2009 Social signal processing: Survey of an emerging domain. *Image and Vision Computing*. Nov;27(12):1743–59.

56. Szekely M, Powell H, Vannucci F, Rea F, Sciutti A, & Michael J. The Perception of a Robot Partner's Effort Elicits a Sense of Commitment to Human-Robot Interaction. *Interaction Studies*. (Under review).

57. Metta G, Sandini G, Vernon D, Natale L, & Nori F. 2008. The iCub humanoid robot: an open platform for research in embodied cognition. *In ACM Press; 2008*. p. 50. Available from:http://portal.acm.org/citation.cfm?doid=1774674.1774683

58. Marteniuk RG, Mackenzie CL, Jeannerod M, Athenes S, & Dugas C. 1987. Constraints on human arm movement trajectories. *Canadian Journal of Psychology/Revue canadienne de psychologie*.41(3):365–78.

59. Ansuini C, Giosa L, Turella L, Altoè G, & Castiello U. 2008. An object for an action, the same object for other actions: effects on hand shaping. *Experimental Brain Research*. Feb;185(1):111–9.

60. Ansuini C, Santello M, Massaccesi S, & Castiello U. 2006. Effects of End-Goal on Hand Shaping. *Journal of Neurophysiology*. Apr;95(4):2456–65.

61. Armbruster C, & Spijkers W. 2006. Movement Planning in Prehension: Do Intended Actions Influence the Initial Reach and Grasp Movement? *Motor Control*.(10). 311-329..

62. Becchio C, Sartori L, Bulgheroni M, & Castiello U. 2008. Both your intention and mine are reflected in the kinematics of my reach-to-grasp movement. *Cognition*. Feb;106(2):894–912.

63. Becchio C, Sartori L, Bulgheroni M, & Castiello U. 2008.The case of Dr. Jekyll and Mr. Hyde: A kinematic study on social intention. *Consciousness and Cognition*. Sep;17(3):557–64.

64. Sartori L, Becchio C, & Castiello U. 2011. Cues to intention: The role of movement information. *Cognition*. May;119(2):242–52.

65. Barakova EI, & Lourens T. 2010. Expressing and interpreting emotional movements in social games with robots. *Personal and Ubiquitous Computing*. Jul;14(5):457–67.

66. Beck A, Cañamero L, Hiolle A, Damiano L, Cosi P, & Tesser F. 2013. Interpretation of Emotional Body Language Displayed by a Humanoid Robot: A Case Study with Children. *International Journal of Social Robotics*. Aug;5(3):325.

67. Karg M, Samadani A-A, Gorbet R, hnlenz KK, Hoey J, & Kulic D. Body Movements for Affective Expression: A Survey of Automatic Recognition and Generation. In: IEEE Transactions on Affective Computing. 4(4):341-359.

68. Laeng B, Sirois S, & Gredebäck G. 2012. Pupillometry: A Window to the Preconscious? *Perspectives on Psychological Science*.Jan;7(1):18–27.

69. Mehlmann G, Häring M, Janowski K, Baur T, Gebhard P, & André E. 2014. Exploring a Model of Gaze for Grounding in Multimodal HRI. In *ACM Press*. p. 247–54. Available from:http://dl.acm.org/citation.cfm?doid=2663204.2663275

70. Kendon A. 1967. Some functions of gaze-direction in social interaction. *Acta Psychologica* 26:22-63 :1.

71. Palinko O, Sciutti A, Patane L, Rea F, Nori F, & Sandini G. 2014. Communicative lifting actions in human-humanoid interaction. *In IEEE; 2014*. p. 1116–21. Available from:http://ieeexplore.ieee.org/document/7041508/

72. Sciutti A, Patane L, Nori F, & Sandini G. 2014. Understanding Object Weight from Human and Humanoid Lifting Actions. *IEEE Transactions on Autonomous Mental Development*. Jun;6(2):80–92.

73. Dragan AD, Lee KCT, & Srinivasa SS. 2013. Legibility and predictability of robot motion. *In IEEE; 2013*. p. 301–8. Available from: http://ieeexplore.ieee.org/document/6483603/

74. Marshall SP. 2002. The Index of Cognitive Activity: Measuring Cognitive Workload. Proceedings of the 2002 IEEE Conference on Human Factors and Power

Plants: *IEEE 7th Human Factor Meeting*. Sep;19:7-9. Available from: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=1042860&isnumber=22350

75. Ahern S, & Beatty J. 1979. Pupillary Responses During Information Processing Vary with Scholastic Aptitude Test Scores. *Science*. (75).1289-1292.

76. Granholm E., Asarnow R. F., Sarkin A. J., & Dykes K. L. 1996. Pupillary Responses Index Cognitive Resource Limitations. *Psychophysiology*. (33). 457-461.

77. Just MA, & Carpenter PA. 1992. The Intensity Dimension of Thought: Pupillometry Indices of Sentence Processing. *Canadian Journal of Experimental Psychology*. (33). 1289-1292.

78. Akgun B, & Tunao D, & Sahin E. 2010. Action Recognition Through an Action Generation Mechanism. In: *International Conference on Epigenetic Robotics* (EPIROB).

79. Breazeal C, Edsinger A, Fitzpatrick P, & Scassellati B. 2010. Active Vision Through Social Robotics. I*EEE Transactions on Man, Cybernetics and Systems*. (20).

80. Metaxas D, & Zhang S. 2013. A review of motion analysis methods for human Nonverbal Communication Computing. *Image and Vision Computing*. Jun;31(6–7):421–33.

81. Toshev A, & Szegedy C. 2014. DeepPose: Human Pose Estimation via Deep Neural Networks. *In IEEE; 2014*. 1653–60. Available from: http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6909610

82. Yan H, Marcelo H, Ang Jr, Poo AN. 2014. A survey on perception methods for human–robot interaction in social robots. *International Journal of Social Robotic*s 6 (1). 85–119.

83. Gilzenrat MS, Nieuwenhuis S, Jepma M, & Cohen JD. 2010. Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cognitive, Affective, & Behavioral Neuroscience*. Jun;10(2):252–69.

84. Richardson DC, Dale R, & Tomlinson JM. 2009. Conversation, gaze coordination, and beliefs about visual context. *Cognitive Science*. 33(8), 1468-1482.

85. Petrushin VA. 2000. Emotion recognition in speech signal: experimental study, development, and application. In: *Sixth International Conference on Spoken Language Processing*.

86. Michael, J. 2011. Shared Emotions and Joint Action. *Review of Philosophy and Psychology*. 2(2), 355-373.