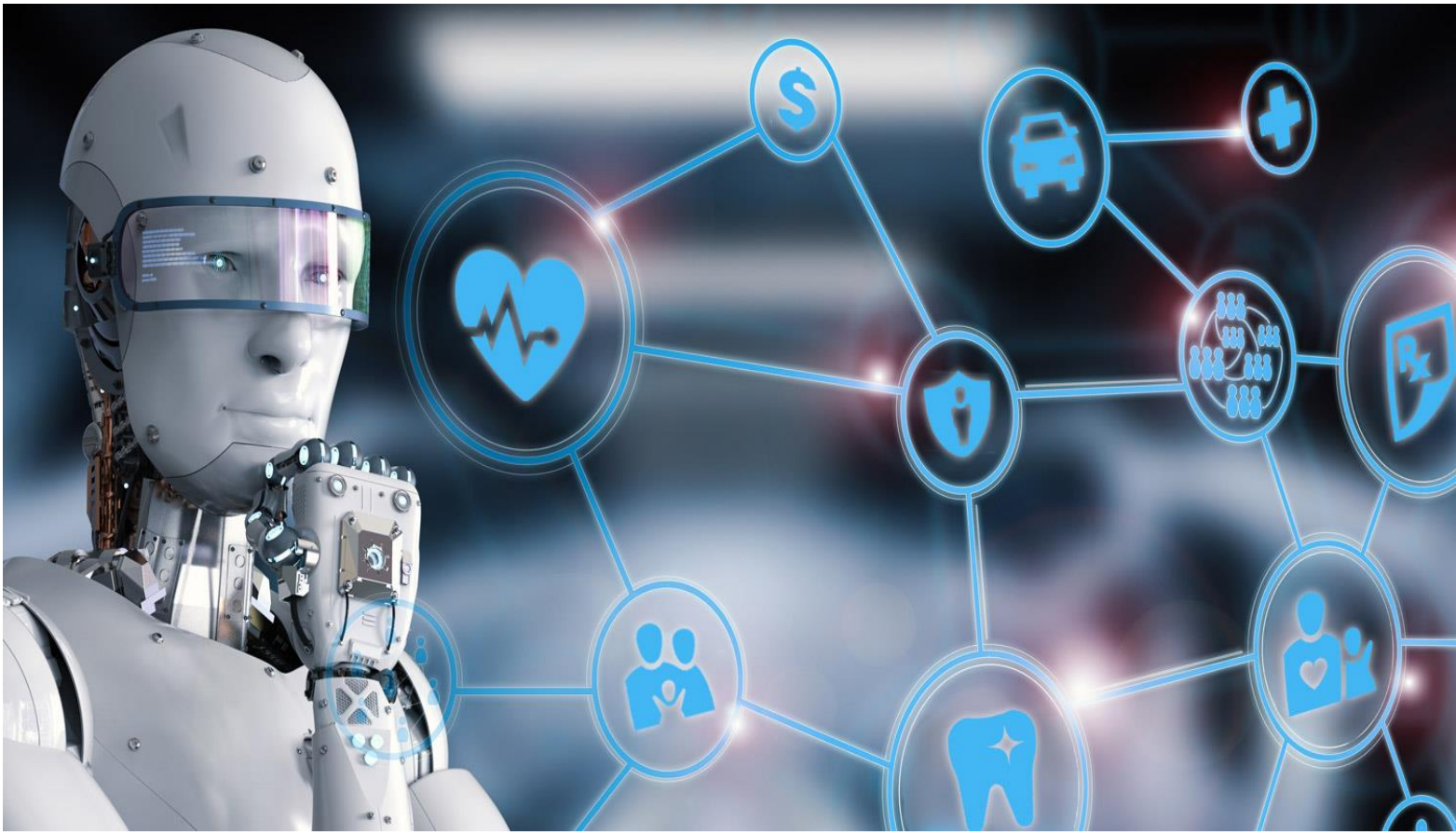
Research Seminar essay



Laura Verhoeven 5605105
Research Seminar in Social Neuroscience (200900353) Ruud Custers
12-04-2021
4578 words

Part 1: Social Motivation

In the first workshop of this course, we discussed the topic of social motivation. In the first place, it needs to be said that motivation is not an easy concept to define (Custers, 2021)[1]. In an article by Braver and colleagues (2014), they highlighted the different definitions motivation has in different areas of research. In order to work together in different disciplines, it is critical that the same definition of motivation is followed. In my study of Liberal Arts and Sciences, I have come across a lot of different study fields and disciplines where most of these areas do have their own thoughts about certain concepts or theories. They have their own way of looking at something. However, I have come to learn that with a lot of phenomena it is useful to look at it from multiple points of view and thus using other disciplines. In the case of motivation, the definitions and concepts used differ but can be merged into an interdisciplinary definition of motivation. In the lecture, both intentions and values have an influence on motivation, which in turn has an influence on effort recruitment and exertion which has an influence on performance (Custers, 2021). This can be seen in the figure below.

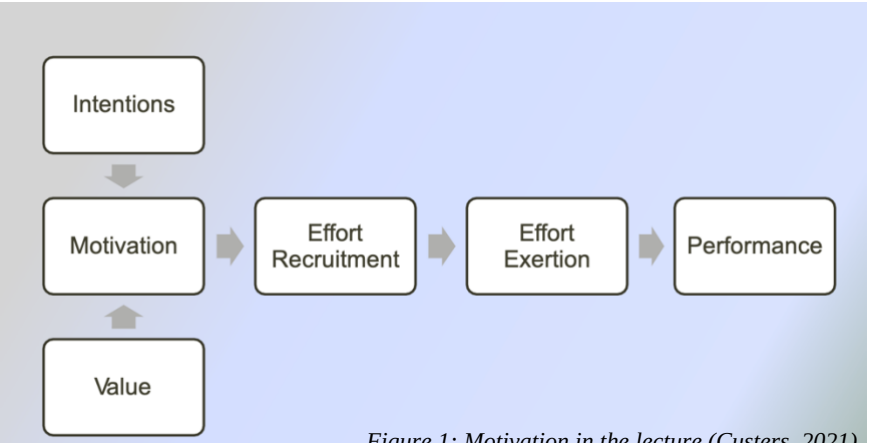


Figure 1: Motivation in the lecture (Custers, 2021)

In the social, affective and personality psychology, these intentions and values might be explained a bit differently. Intentions can be defined as feasibility, which are the "expectations of the probability of attaining the desired future outcome, on the basis of experiences in the past" (Braver et al., 2014, p.446). Values can be defined as desirability, which can be subdivided into motive strength and incentive value and entail the "estimated value of a specific future outcome" (446). In cognitive neuroscience, this 'future outcome'

[1] Source through MS Teams (not publicly available)

can be linked to decisions regarding effort recruitment and effort exertion (447). They all kind of say the same things, influences on motivation might be divided into two separate concepts. Either intentions and values, or feasibility and desirability respectively. The broader explanation of motivation that is shown in the figure from the lecture could be said to be derived from multiple disciplines thus creating a whole picture out of smaller parts. I have created a comparable figure to the figure from the lecture, in which I have added a few parts and changed a few names as to fit the multiple disciplines.

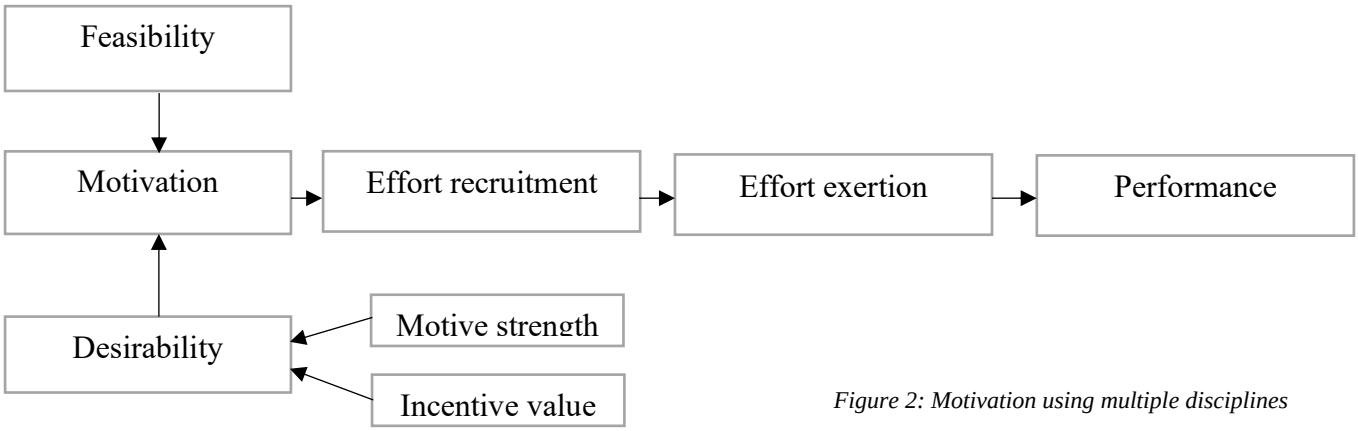


Figure 2: Motivation using multiple disciplines

However, we have not touched upon the topic of unconscious motivation which was a big part of the lecture. Conscious will or motivation is typically seen as the starting point of performance, however unconscious or subliminal reward cues do also motivate behavior. The concepts of feasibility and desirability are thought to need consciousness, however people can unconsciously detect the desirability and feasibility (Custers & Aarts, 2010, 48). For example, it has been found that even if people cannot see the reward cues, there is more and faster activation to higher rewards (Custers, 2021). Personally, this was not big news for me since I already knew some things about the unconscious reward processing from other courses. This was not necessarily the same as the knowledge I have received here, but I did find a relation between the two. Being intrinsically motivated, for me explains that we can be unconsciously motivated by high rewards for example. It is certainly not the same thing, but in my opinion can be linked together since it is perhaps not always clear why we are motivated to do something or if we are even motivated to do something. One possible explanation for unconscious reward processing is explained in the article by Custers and Aarts, by the ideomotor principle (49). The ideomotor principle entails that simply thinking about an action or outcome, might activate the body in such a way that it moves towards that action without a conscious decision (49). Another related explanation for this concept might be the Pavlovian-to-Instrumental-Transfer effect (PIT-effect), which says that the stimulus

itself does not merely indicate what people earn on a certain trial but the rewards themselves have a value attached to them that triggers motivational processes in the brain (Cartoni et al., 2016). The nucleus accumbens and ventral striatum, which process reward cues, not only play a role in the determination of rewarding values but are also connected to the frontal cortex which codes for goal pursuit (Custers & Aarts, 2010, 49). The PIT paradigm entails three stages, a Pavlovian training, an instrumental training and a transfer test. The idea is that the auditory cues presented in the Pavlovian training and the instrumental cues in the instrumental training are linked in the transfer test without the rewarding outcome, in the hopes that a transfer effect will take place in which only hearing the Pavlovian cue will trigger the instrumental cue (831). This can be linked to the ideomotor principle. The PIT paradigm indirectly links two stimuli to a rewarding outcome and to each other, hoping that these two stimuli will create an unconscious motivation to get the rewarding outcome. I did have some difficulty in this area of the workshop, since it took quite some time to really understand the workings and the usefulness of the Pavlovian to Instrumental Transfer effect. It still is a bit confusing how this PIT-effect could be useful in research or, most importantly, how it links to everyday life. For now, I think I have found a reason why it is useful to study. The PIT-effect might explain how we can be unconsciously motivated to do something, it might explain how we link certain cues with an outcome without it being logical to actually see this outcome as related to the certain cue(s).

In addition to this literature, I have looked up three researches that might be well linked to the topic of motivation. A first article covers the human PIT effect in which Talmi and colleagues show that the PIT effect does also exist in humans and shows that the nucleus accumbens and the amygdala are activated in this PIT effect (Talmi et al., 2008). For me, the relevance of the PIT effect became clearer when reading this article. The fact that this PIT effect does exist for humans, and not just for nonhuman animals like mice or rats as the other article proved (Cartoni et al., 2016), provides furthermore more proof that unconscious reward processing in humans does exist.

The other two articles contain more information on unconscious reward processing in relation to the speed-accuracy tradeoff (Bijleveld et al., 2010) and performance (Bijleveld et al., 2011). Bijleveld and colleagues (2011) showed that concentrating on task stimuli when consciously perceiving rewards prevents improvement in task performance. Participants only performed better when unconscious high rewards were used (868). However, the speed-accuracy tradeoff also needs to be taken into account. In another experiment by Bijleveld and

colleagues (2010) where participants had to solve an arithmetic problem, participants who consciously saw high rewards started to respond more slowly but more accurately. If participants did not consciously see them, they did respond faster but it did not affect accuracy (333). This finding can be connected to the task performance improvement. If people are not aware that a high reward is at stake, they perform better (Bijleveld et al., 2011), and they respond faster without it affecting their accuracy. Whether high rewards are consciously or unconsciously perceived, this does effect performance of participants in a certain task. It might even mean that when people are consciously aware there is a lot at stake, they become more cautious and thus perform worse or slower. In my opinion, this relates to the importance of unconscious reward processing. If people do not feel the pressure to perform, because they are not aware of how much is at stake, they might perform better which might be related to everyday life as well.

Part 2: Social Interaction

In the second workshop of this course, we touched upon the social interaction between robot and human. In the lecture we mainly talked about two big themes, the attribution of socialness to an artificial agent and the social interactions with these agents (Hortensius, 2021)[2]. Attributing socialness to an artificial agent is a difficult topic, since it is not entirely clear what exactly it is that makes a robot social. In the lecture a social robot is defined as a "physically embodied agent with some/full autonomy that interacts with humans by means of communication, cooperation and decisions" (Hortensius, 2012). In attributing socialness, we would not only have to take the visual features of the robot into account but also the human knowledge. This is also clearly stated in the article by Hortensius and Cross (2018) in which the impact of the visual features and the impact of the knowledge cues in humans are researched. In this article, as well as in the lecture, it is stated that some of the same areas that are activated in human-human interaction are active in human-robot interaction as well (2018). There are three main brain networks that can be distinguished: the person-perception network, the action observation network and the theory-of-mind network (95). These networks are all in some way activated in human-robot interaction. For example, the person perception network is activated in movement in artificial agents and both this network and the action observation network are activated in mutual gaze (96-99). Perhaps the knowledge cues have a bigger role than the stimulus cues do, since these knowledge cues consistently impact engagement of behavioral and brain mechanisms supporting the attribution of socialness to artificial agents.

Returning to the lecture, the attribution of socialness is also related to perceived agency and experience of the artificial agent (Hortensius, 2021). An example of agency might be memory and an example of experience, rage. Artificial agents are not perceived as being able to feel or sense something (i.e., experience) and are not (or less) perceived as having memory. In my opinion, this might be logical since most of the artificial agents that are available nowadays do not look human or living which makes it more difficult to actually perceive them as having a mind. Plus, if an artificial agent does look human its actions often do not correspond with the humanness of its features which might make it uncomfortable for the human. This is

[2] Source through MS Teams (not publicly available)

called the Uncanny Valley (see figure 4). The question on my part is, however, if it is necessary to actually have a robot that looks and acts like a human. Would it not be more useful to study how we can incorporate artificial agents in day-to-day life as animals or objects as opposed to humans? This might be something to study further.

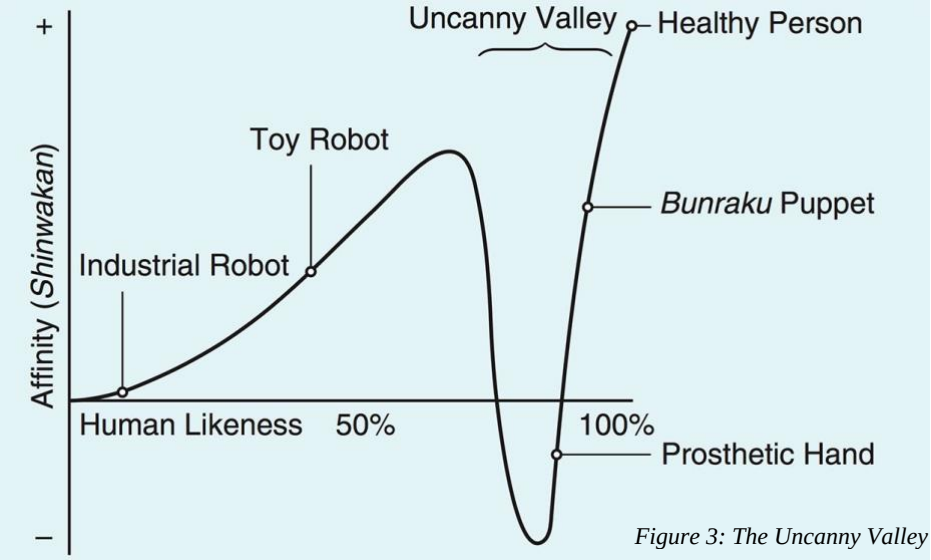


Figure 3: The Uncanny Valley

In the second part of the lecture, we talked about social interaction with artificial agents. Three different topics were touched upon: automatic imitation, empathy and social interaction. Namely empathy is a difficult topic, since there is no overlap in neural mechanisms of empathy when considering artificial agents. However, if the robot is part of the ingroup, people tend to be more empathetic (Hortensius, 2021). In my opinion, this might also mean that we do consider artificial intelligent beings as humans or at least human-like. Why else would we consider them ingroup, and a more important question perhaps is if we also consider animals sometimes as our ingroup? If this is not the case, there might be something special about the robot and there might be a bright future in the realization of human characteristics and human liking of the robots. This might be able to tell us a lot about the way humans interact with each other as well. This can be linked to the two other articles by Wykowska and colleagues (2016), and Henschel and colleagues (2020). In these articles, the focus is on human-robot interaction in order to be able to say something about human-human interaction. In the article by Wykowska and colleagues (2016), they also consider the socialness of artificial agents. According to them, robots are social when they activate the same mechanisms to the same extent as humans do during interaction (7). In certain ways artificial agents do indeed activate the same mechanisms, however action and perception need to be matched. The same amount of human likeness needs to be achieved in both visual features and actions (2016). This can be linked back to the Uncanny Valley in which this

coupling is not achieved. In the article by Henschel and colleagues (2020), they propose using a certain technique to be able to get a clearer image of the changes in interaction with a robot. The technique they propose is the functional near infrared spectroscopy (fNIRS), which is a mobile neurocognitive method. This could be a very interesting technique to study the interaction, since we might be able to see which brain areas are activated in the interaction at different timepoints (380-81). We might be able to see the live changes in interactions, instead of waiting until later to see if the brain areas activated in robot interaction changed after longer interactions with the robot.

I have also found an article that might be of interest in terms of this new method, fNIRS. One line of use in this method is with providing evidence for the Uncanny Valley. This article by Strait and Scheutz (2014) states that the human likeness of an artificial intelligent being might induce a change from liking to disliking in humans. This change is termed the Uncanny Valley. They found that this Uncanny Valley, in addition to subjectively disliking the robot, also changes the activation in the anterior prefrontal cortex which is involved in emotion regulation (1132). Striat and Scheutz found that there was a greater hemodynamic change in the anterior prefrontal cortex when viewing humanlike robots and they found that this change might be related to aversion or disliking (1132). By using this new method, they were able to show the difference in exchange between a human, a human-like robot and a nonhuman-like robot in order to see how the Uncanny Valley affects the brain areas associated with emotion. A second article proposes a new framework for studying human-machine interactions (Cross and Ramsey, 2020). The most interesting part Cross and Ramsey propose, which might be of interest in further study, was the fact that perhaps robots should not be seen as animals or humans but as objects (206). In fact, robots are objects that are created with humanlike characteristics. In the article they state that perhaps we should use the information that is available about object perception to incorporate this in the human-machine interaction framework (206). A last article also proposes a framework, but then a cognitive empathy framework for social robots (Bagheri et al., 2020). They propose that we need to establish emotive and empathic behavior in social robots, since this makes them perceived as being more friendly and more positive (1). It is quite difficult for the robot to learn this kind of behavior; thus, they need to learn this behavior through reinforcement. For this to work, humans and robots need to interact on a daily basis where the robot can learn to give appropriate empathic responses to human facial expressions (3-4). I think it might be very useful to incorporate this as well in the human-machine interaction. In my opinion, we need

to broaden the human-machine interaction framework proposed in the article by Cross and Ramsey to both capture the object perception and capture the learning abilities of the robots. In this way, robots might learn from the human interaction and we as humans do feel more comfortable with the human likeness of the robot if the robot also actually has human characteristics outside of its visual features.

Part 3: Social Identity

In the third and last workshop, Felice van Nunspeet told us about social identity and morality (2021)[3]. Social identity differs in each context, according to the group someone comes in contact with. A distinction can be made between an ingroup and an outgroup. An ingroup is the group you feel you belong to and an outgroup is the group you do not feel you belong to (Van Nunspeet, 2021). There has been a lot of research concerning in- and outgroups, in which it becomes clear that there is a certain positive bias for ingroups and negative bias for outgroups. This is also shown in a study by van Bavel and Pereira (2018) in which they proposed an identity-based model of political belief. They showed that someone's political belief can shape someone's identity, and can even shape the way someone sees the world around them (214). The reason why this happens is because people want to belong to a certain group. I think the fact that people are motivated to change their own beliefs to fit in with a certain group, is amazing. I did know about an in- and outgroup before this lecture. However, I did not give the in- and outgroups that big a role within social identity. I did know about the fact that we empathize less with the outgroup and that there are certain biases involved for example when associating positive or negative words with in- or outgroups, but I did not think or realize the in- and outgroups could actually change the way we see the world when it comes to changing our beliefs. Even if beliefs in that social group are in contrast with the truth or are even disproven, people are motivated to reduce or even suppress this (215).

This might also be explained by the social identity theory, which is explained in an article by Scheepers and Derks (2016). The social identity theory, the fact that people define a part of the self through group membership, can be explained by a will to survive since it serves basic human needs (74). There is even neuroscientific evidence for this social identity, for example that the prefrontal cortex is not only activated in self-referential processing but also in ingroup representation (75). Political partisanship (as in other social groups) thus influences someone's identity and beliefs, because they conform to the norm of the political party. This social conformity is seen in every social group, even if this ingroup is just created for research purposes. For example, the line judgment task used in a study by Chen and colleagues (2012) found that people conform to the norm of the group they are in. Participants would judge the length of the line in comparison with other lines, and they found

[3] Source through MS Teams (not publicly available)

that participants would change their answer to fit the answers of the rest of the group (2012). As was said earlier in this essay, an ingroup bias can also be seen in social groups. For example, people are more likely to rate their own ingroup as being more likely to be fastest in a game (Van Nunspeet, 2021). This ingroup bias is also seen when observing someone's suffering, they show similar brain responses when they themselves are sad as when ingroup members are sad but not with outgroup members (Scheepers, 75).

This is then a nice bridge to morality. People are readily influenced by the perceived moral norm of the group when making their own moral choice in a group context (Van Nunspeet, 2021). And they are also influenced by an emphasis on either morality or competence when making judgments. Van Nunspeet and colleagues (2015) performed a study in which participants had to do an implicit association task, where either an ingroup or an outgroup member was shown at the end of each trial. These in- or outgroup members showed whether they did the trial right or wrong. They found that if participants saw an ingroup member at the end of the trial, they showed a greater ERP for incorrect trials when focusing on morality than with an outgroup member (2015). People are sensitive to the judgment of other ingroup members, which might cause an alteration in their own judgment (Ellemers & van Nunspeet, 2020). It is even found that people actually see answer options differently when having heard other people's answers and thus consider these options differently (515). A fear of social exclusion and moral criticism of other people, might make it logical for someone to depend on other people to make their own moral choice. This is also something I find fascinating; we actually see the world differently according to the group we associate ourselves with. The group's opinion is worth so much that we might actually change our opinion on what we feel is the right thing to do.

In another article by van Nunspeet and colleagues (2014) which precedes the article of 2015 mentioned above, they also focused on this morality. In this experiment, participants performed an implicit association test (IAT) in which the focus was either on morality or on competence (142). It was shown that people have a smaller IAT-effect when focusing on morality than on competence, which means that people reduce their bias towards Muslim women in this case when focusing on morality (145). It does not in fact take the group's opinion into accordance, as the article of 2015 did, but goes back to the basis in which we, as individuals, make moral choices. It shows us that we have an implicit bias, but when focusing on morality, this might be reduced. Of course, as they show in the article in 2015, this effect

of a reduced implicit bias because of focusing on morality is reduced because of the introduction of an in- or outgroup member. An article by Delgado and colleagues (2005) also adds a social factor into their experiment, now focusing on the moral character of people. Participants had to take the moral character of their partners into account when playing a trust game. It was shown that participants made the most risky choices with the morally good partner and thus trusted them the most (1611). These risky choices were also correlated with the caudate nucleus associated with feedback learning (1615). This might be a strange finding, since the feedback did not seem to influence the risky choices the participants made. It seems that the moral character of their partner was the most influential factor in deciding whether to trust them or not (1615). To switch from morality back to social conformity, an article by van de Waal and colleagues (2013) is of interest. They showed that social learning and social conformity is also established in the wild. By having the monkeys avoid a certain bitter-tasting food, colored either blue or pink, in a first phase of the experiment, they tried to establish a new group norm of avoiding this type of food (484). After this, a couple of months later they saw that young monkeys did indeed show a certain social learning in which they avoided this type of food (484). The most interesting thing they showed was that monkeys who switched between two groups in which different norms were held, also switched between the group norms (484).

Part 4: Integration

For this last, integrative part, I have chosen to focus mainly on the social interaction with robots and how this relates to the other two themes. I have made this choice because my personal interest lies within the field of social. This fourth part of the essay is divided into three different parts: ingroups, reinforcement learning and motivation.

The first part concerns ingroups. In the workshop on social identity, this topic was mainly discussed. It was stated that there is a certain ingroup bias when seeing someone suffer for example (Scheepers and Derks, 2016, 75). Similar brain responses are seen when they are sad and when their ingroup members are sad. This same ingroup bias might be related to social interaction with robots. As was stated in the second workshop, if the robot is part of the ingroup people are more empathetic (Hortensius, 2021). I related this to the ingroup bias as it might mean that when a robot is part of the ingroup, people might also activate the same brain areas when the robot is sad. Not only this, but it might also be related to morality. It was said that ingroup members can influence certain moral choices people make (Ellemers & Van Nunspeet, 2020). If robots are perceived as ingroup members, could this also mean that they are able to change the opinion people have on what is the right thing to do morally. In my opinion, this would be a very interesting line of study. It might broaden our understanding on how much the ingroup has an effect on this morality. However, as the prefrontal cortex is activated in ingroup representation (Scheepers and Derks, 75), it was also found that in the Uncanny Valley the anterior prefrontal cortex is more activated for more humanlike robots (Striat and Scheutz, 2014). This was related to the disliking or aversion. It could be that these humanlike robots activate this region more, because they are not perceived as an ingroup member or that it is related to this area because people are deceived. The action-perception coupling is not achieved, thus the robot might not even be called social. In fact, according to Wykowska and colleagues (2016), social robots are only social when they activate the same mechanisms in human-robot interaction as is achieved in human-human interaction. Thus, before changing moral choices or being part of an ingroup, a certain socialness of the robot needs to be established by coupling action and perception. Below, I have added a visualization of all the different elements that can be related to ingroups in order to make it little bit more clear.

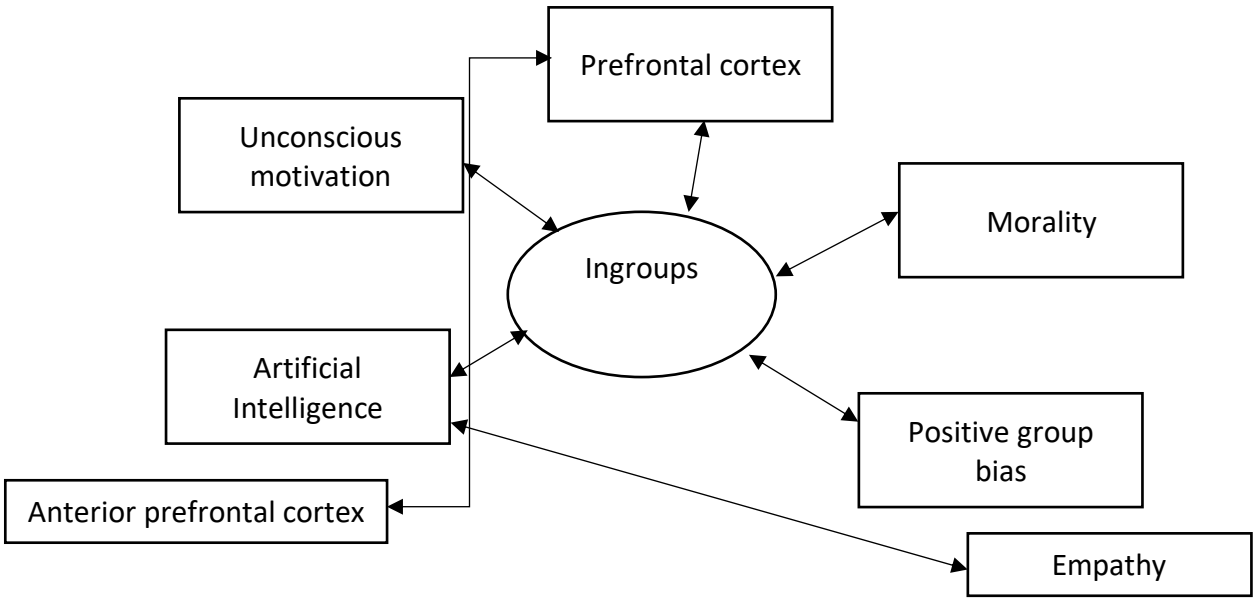


Figure 4: Ingroups as the middle point

The second part relates to reinforcement learning. In order for the robot to be perceived as more friendly, it needs to learn to be empathetic (Bagheri et al., 2020). This behavior is not easily learned and thus needs to be learned through reinforcement. A question that came to mind within this topic was: how far might we be able to induce a certain seemingly unconscious motivation to be empathetic on the part of the robot? How do we achieve a certain interaction in which the human does think the robot is actually empathetic? We might need to establish a certain reward on the part of the robot, for example a happy face when giving the right empathetic answer. When I thought about relating this to the PIT-paradigm, I did not immediately think this could be related to robots. In fact, how do we establish a reward with a robot. It can learn to associate certain facial features to certain outcomes, but I don't think it can learn to associate this happy facial expression to lead to some other outcome because what outcome would be useful for the robot?

A third part relates more broadly to motivation. It does not immediately relate to robots, but it does relate to ingroups. Perhaps an explanation to conform to the group, is an unconscious higher reward. For example, social acceptance. When asking people why they conform to fit in with the group, I don't think they have an explanation for it. The underlying reason might be because they don't want to be socially rejected, and this is a bigger punishment than changing your own opinion for example.

Bibliography

Part 1: Social Motivation

Bijleveld, E., Custers, R., & Aarts, H. (2011). Once the money is in sight: Distinctive effects of conscious and unconscious rewards on task performance. *Journal of Experimental Social Psychology, 47*(4), 865–869. doi:10.1016/j.jesp.2011.03.002

Bijleveld, E., Custers, R., & Aarts, H. (2010). Unconscious reward cues increase invested effort, but do not change speed –accuracy tradeoffs. *Cognition, 115,* 330–335. doi: 10.1016/j.cognition.2009.12.012

Braver, T. S., Krug, M. K., Chiew, K. S., Kool, W., Andrew Westbrook, J., Clement, N. J., ... Somerville, L. H. (2014). Mechanisms of motivation-cognition interaction: Challenges and opportunities. *Cognitive, Affective and Behavioral Neuroscience, 14,* 443-472. doi:10.3758/s13415-014-0300-0

Cartoni, E., Balleine, B., & Baldassarre, G. (2016). Appetitive Pavlovian-instrumental Transfer: A review. *Neuroscience and Biobehavioral Reviews, 71,* 829–848. doi: 10.1016/j.neubiorev.2016.09.020

Custers, R. (2021, 18th February). Social Motivation. Source through MS Teams.

Custers, R., & Aarts, H. (2010). The unconscious will: How the pursuit of goals operates outside of conscious awareness. *Science, 329,* 47–50. doi:10.1126/science.1188595

Talmi, D., Seymour, B., Dayan, P., & Dolan, R. J. (2008). Human Pavlovian instrumental transfer. *Journal of Neuroscience, 28,* 360–368. doi: 10.1523/JNEUROSCI.4028-07.2008

Part 2: Social Interaction

Bagheri, E., Roesler, O., Cao, HL. *et al* (2020). A Reinforcement Learning Based Cognitive Empathy Framework for Social Robots. *Int J of Soc Robotics, 2020*. doi:10.1007/s12369-020-00683-4

Cross, E. S., & Ramsey, R. (in press). Mind Meets Machine: Towards a Cognitive Science of Human–Machine Interactions. *Trends in Cognitive Sciences, 25*, (3), 200-212. doi: 10.1016/j.tics.2020.11.009

Henschel A., Hortensius R. & Cross E.S. (2020). Social Cognition in the Age of Human Robot Interaction. *Trends in Neurosciences, 43*, 373-384. doi: 10.1016/j.tins.2020.03.013

Hortensius, R. (2021, 3th March). Social Interaction. Source through MS Teams.

Hortensius, R. & Cross, E.S. (2018). From automata to animate beings: The scope and limits of attributing socialness to robots. *Annals of the New York Academy of Sciences, 426*, 93-110. doi: 10.1111/nyas.13727

Strait, M. & Scheutz, M. (2014). Measuring users' responses to humans, robots, and human like robots with functional near infrared spectroscopy. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, 1128-1133. doi: 10.1109/ROMAN.2014.6926403.

Wykowska A, Chaminade T, Cheng G (2016). Embodied artificial agents for understanding human social cognition. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*. doi: 10.1098/rstb.2015.0375

Part 3: Social Identity

Delgado, M. R., Frank, R. H., & Phelps, E. A. (2005). Perceptions of moral character modulate the neural systems of reward during the trust game. *Nature neuroscience, 8,* (11), 1611-1618. doi: 10.1038/nn1575

Ellemers, N., & Van Nunspeet, F. (2020). Neuroscience and the social origins of moral behavior: How neural underpinnings of social categorization and conformity affect everyday moral and immoral behavior. *Current Directions in Psychological Science, 29*, (5), 513-520. doi: 10.1177/0963721420951584

Scheepers, D., & Derks, B. (2016). Revisiting social identity theory from a neuroscience perspective. *Current Opinion in Psychology,11*, 74-78. doi: 10.1016/j.copsyc.2016.06.006

Van Bavel, J. J., & Pereira, A. (2018). The partisanbrain: An identity-based model of political belief. *Trends in Cognitive Sciences, 22*, (3), 213-224. doi: 10.1016/j.tics.2018.01.004

Van Nunspeet, F. (2021, 17th March). Social Identity. Source through MS Teams.

Van Nunspeet, F., Ellemers, N., Derks, B., & Nieuwenhuis, S. (2014). Moral concerns increase attention and response monitoring during IAT performance: ERP evidence. *Social, Cognitive, and Affective Neuroscience, 9*, (2), 141-149. doi: 10.1093/scan/nss118

Van de Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent Social Learning and Conformity Shape a Wild Primate's Foraging Decisions. *Science, 340*, (6131), 483-485. doi:10.1126/science.1233675