



HAL
open science

Cognitive load increases anthropomorphism of humanoid robot. The automatic path of anthropomorphism.

Thierry Chaminade, Nicolas Spatola

► To cite this version:

Thierry Chaminade, Nicolas Spatola. Cognitive load increases anthropomorphism of humanoid robot. The automatic path of anthropomorphism.. International Journal of Human-Computer Studies, 2022, 10.1016/j.ijhcs.2022.102884 . hal-03438173

HAL Id: hal-03438173

<https://hal.science/hal-03438173v1>

Submitted on 25 Nov 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Copyright

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/362975729>

Cognitive load increases anthropomorphism of humanoid robot. The automatic path of anthropomorphism

Article in *International Journal of Human-Computer Studies* · November 2022

DOI: 10.1016/j.ijhcs.2022.102884

CITATIONS

4

READS

42

2 authors:



Nicolas Spatola

Istituto Italiano di Tecnologia

53 PUBLICATIONS 309 CITATIONS

[SEE PROFILE](#)



Thierry Chaminade

French National Centre for Scientific Research

89 PUBLICATIONS 5,576 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Learning to be Human [View project](#)



Hierarchical reinforcement learning [View project](#)

Cognitive load increases anthropomorphism of humanoid robot.

The automatic path of anthropomorphism.

Nicolas Spatola¹, & Thierry Chaminade²

¹ Artimon Perspectives, Paris, France.

*² Aix-Marseille Université, Institut de Neurosciences de la Timone, UMR 7289, CNRS,
Marseille, France.*

Corresponding authors:

Nicolas Spatola

Artimon Perspectives

18 rue de la Victoire

Paris

E-mail: NSpatola@artimon.fr

Abstract. Humanoid robots might become more and more present in the most ordinary contexts of millions of people worldwide. Humans reason about these artificial agents mainly through the attribution of human characteristics, a process called anthropomorphism. However, despite number of studies, how we develop and structure the representation of non-human agents is still an open question. In the present paper, we aim at integrating the anthropomorphism into the cognitive control theory, a construct from cognitive neuroscience that refers to information processing and cognitive resources managing that varies adaptively to the situation. In three experiments we manipulated the cognitive load of participants during the observation of an active robot to investigate how the load could impact the online structuration of participants' mental representation of the robot. The two first experiment converged in arguing for a control process resource-demanding to switch from the social cognition to the physical cognition inhibiting anthropomorphic inferences. The third experiment investigated the influence of the "what" and "why" observation goals on the cognitive load effect arguing that an explicit focus on intentionality attribution bias the automatic process of anthropomorphism. The representation and perception of robots are further discussed in term of cognitive control theory and social cognition.

1 Introduction

Interaction with robots is likely to become a daily activity in the years to come. Human-Robot Interaction (HRI) studies, such as referenced below, have demonstrated that humans tend to reason about robot through a phenomenon known as anthropomorphism. Anthropomorphism is defined as the attribution of human characteristics to nonhuman entities such as animals (Eddy et al., 1993; Wynne, 2006) or robots (Nyangoma et al., 2017; Złotowski et al., 2015). Robots' anthropomorphism is not anecdotal as it entails not only how humans consider robots (Darling, 2017; Eyssel & Kuchenbrandt, 2012; Kuchenbrandt et al., 2013; Spatola et al., 2019a) but also how robots impact humans own cognition and behaviours (Riether et al., 2012; Spatola et al., 2018a, 2019a). In this paper, we propose to investigate how cognitive control mechanisms influence robots' anthropomorphism. In particular, cognitive load has been shown modulate how people explain others' actions and make inferences about them in terms of physical or mental properties (Rosset, 2008). For example, people tend to attribute more divine intentions to natural events and produce more teleological interpretation when they experience a shortage of cognitive resources (Kelemen & Rosset, 2009). We further develop the theory of physical/mental social cognitions dichotomy and the modulating role of cognitive resources.

1.1 Physical and social processing

Humans interpret changes in the world as either physical (e.g., the processing of objects and motions) or mental processes (e.g., the processing of agents and intentions). Philosopher Daniel Dennett (D. Dennett, 2009) proposed that individuals adopt a specific stance depending on the phenomenon under scrutiny. Whenever they try to predict physical phenomenon such as the trajectory of a kicked ball, they take a *physical stance*, using empirical rules. However, more complex phenomenon such as human behaviour cannot be efficiently described using these physical rules. The *intentional stance* describes another mode of representation that relies on mentalizing, and in particular the attribution of mental states such as intentions and would be

more social in nature as it supposes to consider the other as a (social) agent compared to an object.

This representation of the cognition system echoes the model of two core cognitive processes (Darlow & Sloman, 2010; Evans, 2008; Evans & Stanovich, 2013). According to this model, cognition may be delineated in a Type 1 automatic and Type 2 controlled processes. The two cognitions type differ in their characteristics in terms of resource demands (influenced by the speed of processing, the amount of information processed, etc.). The Type 1 requires a lower activation of working memory through the use of stored representation and is more rigid than the second, making it a fast processing and would be, for that reason, the default mode of cognition (Goodson, 2005; Hansen et al., 1976; Mars et al., 2012). The Type 2 requires cognitive decoupling and higher amount of cognitive resources, resulting in a slower but controlled processing. Some authors posit that the social cognition would be based on Type 1 cognition (Jacoboni et al., 2004; Jack, Dawson, & Norr, 2013; Saris et al., 2020; Tavares et al., 2008). The reason would be that social functioning is necessary for human survival (Eisenberger & Cole, 2012; Humphrey, 1976; Valtorta et al., 2018). In contrast, explicit cognitive control mechanisms, including inductive reasoning and causal reasoning about the characteristics of the physical system, are demanding in cognitive resources, slower than implicit processes and limited by working memory capacity and would rely on Type 2 cognition (Evans, 2008). Importantly, as Type 1 and Type 2 cognition are exclusive socio-cognitive and physical system would be, as a consequence, mutually exclusive so that when the social cognition is active the physical system is inactive (Jack, Dawson, Begany, et al., 2013). Therefore, anthropomorphism of robots could depend on this dual process in which the observer could accumulate information based on the social (Type 1) or physical (Type 2) cognition system. We further consider this association between the type of cognition (i.e. Type 1 and 2)

and the nature of representation (i.e. Social and Physical) as *social cognition* and *physical cognition*.

1.2 Physical and social processing with robots

According to Epley and colleagues, observing a non-human agent, using human representation (anthropocentric schemas), would be an easy and fast strategy to explain, understand and communicate about its behaviour (Epley et al., 2007). This view was supported by further studies (Meyer et al., 2012; Spunt & Lieberman, 2013). This view is in line with the Computer As Social Actor theory (Nass & Moon, 2000) that states that people apply rules and social expectations to computers mindlessly, especially when these agents violate their expectations (Epley et al., 2007, 2008; Marsh et al., 2014). However, these attributions do not imply that people believe that computers may have intention or will. This phenomenon would be more likely to occur when context (i.e. the particular setting in which an event occurs) triggers an individual to use stored schemas that require no conscious awareness or rapid processing (Langer, 1992), for instance when facing irrational or threatening behaviours (Marsh et al., 2014). With robots, the distinction between mindless and mindfulness attributions is more complex than with computers because of their “social” purpose and interactivity or their anthropomorphic design (at least for the humanoids) is more likely to trigger some social cognition favouring intentional and phenomenal experience attribution, which is less likely the case for computers or other non-anthropomorphic designed tools on which the Computer as Social Actor (CASA) theory apply (Coeckelbergh, 2012; Gunkel, 2018, 2019; Levy, 2009; Neely, 2014). From a cognitive view, anthropomorphism could be conceptualized as a process in which the evaluation and the representation of the incoming sensory information about the agent are continuously generated according to the social or physical cognition system. However, as the representation is hypothesized as forming during the observation, the process has to be iterative. This iterative characteristic supposes two axioms: First, if the process is

iterative, the default anthropomorphic reasoning (social cognition) can be bypassed by non-anthropomorphic information (physical cognition) and embedded in the representation when it is still malleable. Second, as the anthropomorphism is the default reasoning path, it has to be actively counterbalanced which requires resources. Therefore, this controllability should be proportional to the amount of cognitive resources available to initiate and carry out the process.

1.3 The role of cognitive control and cognitive load

As many dual-processing accounts of reasoning, judgment, and social cognition (Evans, 2008), we propose that anthropomorphism should be considered in a cognitive control framework. Cognitive control entails the processing of relevant information given a specific context (Engle, 2002). It is an active system describing the ability to select or inhibit perceptual or mental elements. Unlike automatic processes which act at an uncontrolled level, cognitive control acts, by definition, at a controllable level and, as a consequence, is slower and requires more cognitive resources (Ridderinkhof, Ullsperger, et al., 2004; Ridderinkhof, Van Den Wildenberg, et al., 2004). Under certain circumstances, the cognitive control system may, for example, inhibit automatic processes (Augustinova et al., 2018; Sharma et al., 2010). However, this inhibition requires cognitive resources that are limited. The cognitive load theory formalizes how the system allocates this limited amount of cognitive resources. It describes the storage and process of information in working memory and the integration of new information (Leppink et al., 2015; Paas et al., 1994). For example, more resources are available when individuals perform a single task compared to two tasks simultaneously (Moscovitch, 1994), or when competing for interfering cues are present (Augustinova et al., 2018; MacLeod, 1992; Stroop, 1935). In tasks in which individuals have to inhibit an automatic process to select a valid answer, the cognitive control would inhibit the competing cue, a process that could apply to anthropomorphism (Urquiza-Haas & Kotrschal, 2015).

Some authors hypothesized that the tendency to anthropomorphize human agents should increase in a situation of high cognitive load (Waytz et al., 2010), but this has never been demonstrated. Urquiza-Haas and Kotrschal have proposed that a high cognitive load could interfere with the suppression of the default social cognition when it is triggered by purely perceptual information (Urquiza-Haas & Kotrschal, 2015). We agree with that default system theory and propose that if one considers that social cognition is the automatic process, and the physical cognition is at rest by default, the perception of robots should rely, at first, on that social cognition (Type 1) favouring anthropomorphic inferences. However, because robots are not human, the cognitive control system could inhibit this automatic path and activate the physical cognition (Type 2) (Darlow & Sloman, 2010; Evans & Stanovich, 2013). This physical cognition being active, anthropomorphic inferences would be inhibited and robots considered as objects. This inhibition would require cognitive resources and, as such, would depend on the amount of cognitive resources available during the observation of the agent (figure 1).

The model we propose may be considered as a loop in which the representation of the agent is built through an iterative process. For each iteration, information embedded in the representation of the agent would depend either on the social or physical cognition, considered as mutually exclusive (Schilbach et al., 2008). However, in a second model, the parallel-competitive form, both processes are activated in parallel as information treatment pipelines that weight automatic and controlled information to provide a common output (Smith & DeCoster, 2000). Our current hypotheses would apply in both framework and we do not aim at investigating this epistemological question in the present study.

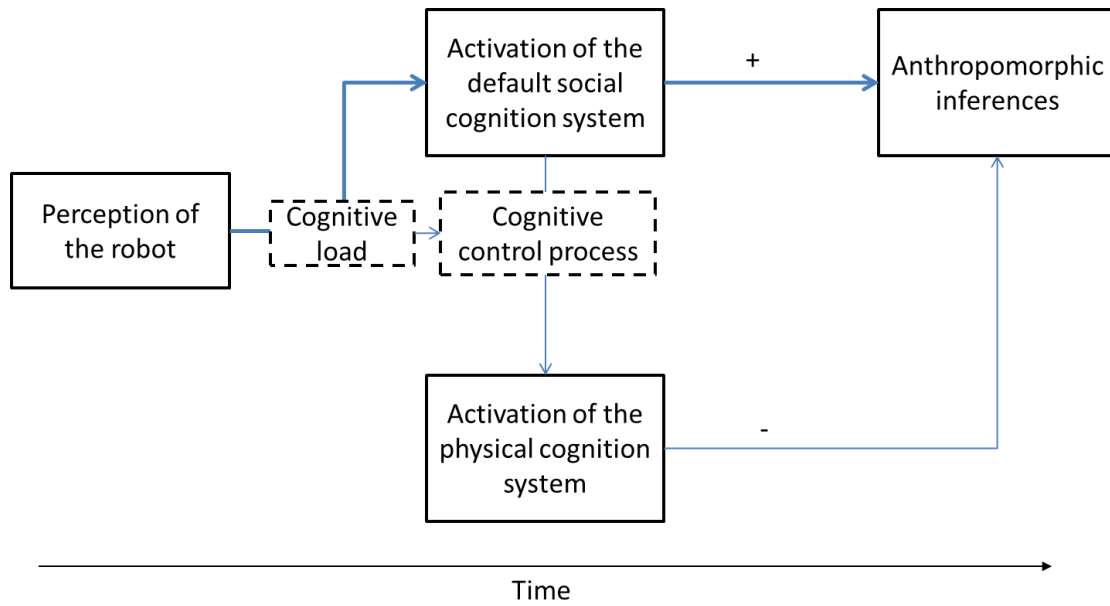


Figure 1. From the social and physical cognition to the anthropomorphic process. In the present study the context refers to the modification of participants cognitive load but we assume that other variables (ex. type of observed action, engagement in the situation, appraisal, arousal, etc.) could impact the general process.

As such, we developed our first and second hypothesis:

{H1}: Higher cognitive load increases people's tendency to anthropomorphize robots.

{H2}: Under high cognitive load, the higher the level of cognitive load felt by participants, the higher the anthropomorphism

1.4 The goal of observation

Caporael (Caporael, 1986; Hansen et al., 1976) argued that, observing and forming a representation of an agent, the goal of the observation has to be taken into account. For instance, Spunt and Lieberman (2013) showed that attributing mental properties to an agent was modulated by cognitive load only when participants were prompted to attributed a motive to observed action (“Try to understand why the agent is doing this action”) but not when the

observer adopts a more descriptive perspective (“Try to understand how the agent is doing this action”). In other words, individuals tend to anthropomorphize more when they are explicitly asked to (Meyer et al., 2012; Spunt & Lieberman, 2013). However, contrary to our hypothesis, Spunt and Lieberman (2013) demonstrated that, when individuals were instructed to focus on the intention of the observed agent, mental properties attribution to a target decreased as the function of the cognitive load. This inadequacy between our hypothesis and Spunt and Lieberman results may be disentangle by de Lange and colleagues study (de Lange et al., 2008). The authors showed that asking participants to reflect deliberately on goals and intention could bias how the mirror neurons and mentalizing areas interconnect (de Lange et al., 2008) and could impact the synergy between the two systems (Sperduti et al., 2014; Van Overwalle & Baetens, 2009). In other words, explicit reflection on goal and intention could bypass the automatic process of mentalization that is, according to Spunt and colleagues further study, automatically primed by the default mode network (Spunt et al., 2015). As such, we developed our third hypotheses:

{H3}: Explicit instruction to reflect on intentional goal (*why*) of the robot’s action decreases anthropomorphic in high cognitive load context while the opposite should be true—in line with H1, H2 and H3—in the absence of instruction or a “non-intention” focused instruction (*what*).

Indeed, we may assume that primed goal could produce different results comparing high vs. low cognitive load situation. Cognitive load would impair social cognition process when participants are explicitly requested to reason about the intention of a target agent and therefore decrease anthropomorphism (Lin et al., 2010; Spunt & Lieberman, 2013). Conversely, without specific or “non-intentionality” related instructions, people would tend to attribute more anthropomorphic characteristics under cognitive load (Kelemen & Rosset, 2009). The reason is that when individuals are explicitly required to identify the intention of an agent, they turn

an part-unconscious process in a controlled, conscious process (Moors & De Houwer, 2006; Wegner & Bargh, 1998). When explicit, the intention identification task would require higher amount of cognitive resources (compared to a non-explicit situation), especially because non-usual, and, as such, would be more likely to be impaired by a lack of cognitive resources.

1.5 The present study

To test whether robots' anthropomorphism is a default process that is actively inhibited we conducted three experiments (experiment 1 and 2). In each experiment, we manipulated the level of cognitive load while participants watched a video presenting a robot. Participants were either in a low cognitive load condition or a high cognitive load condition. We subsequently measured anthropomorphic inferences. In the last experiment (experiment 3), we investigate the moderating role of the goal in the model hypothesizing that when the goal explicitly requires to explain or predict the behaviour of an entity is important, anthropomorphic attributions should increase. We further test the complete model in the form of a decision tree that explain the anthropomorphic attributions as the function of the cognitive load and the goal of evaluation.

2 Experiment 1

2.1 Method

Participants were 83 males, 42 females and 4 others ($M_{age} = 20.6$, $SD = 4.8$) recruited online. The experiment was processed with Qualtrics. Participants took part in this experiment on a voluntary basis. We estimated the requested sample size with G*Power, with a medium effect size, a 0.05 α , power set at 0.90 and a 1/1 between-group ratio (Faul et al., 2007). The requested sample size was 108.

The experimenter informed participants that the experiment was a memory test. Instructions provided the schedule of the experiment: 1) a questionnaire, 2) a sequence of numbers to memorize, 3) a video to watch on which, subsequently, questions would be asked. First, they had to complete the Negative Attitudes towards Robots Scale (NARS) (Syrdal et al., 2009). This scale was used to assess that both experimental groups were comparable considering attitudes toward robots.

Second, they were randomly assigned either to the high cognitive load (n=66) or the low cognitive load (n=63) group. In both groups they were informed that a sequence of numbers would appear for 30 seconds, their task being to memorize it (Gilbert & Hixon, 1991; Sherman et al., 1998; Sherman & Frost, 2000; Wehling et al., 2008). In the high cognitive load experimental condition, the sequence was composed of 9 digits randomly selected (i.e. “214087539”, “257142809”, “952176034”). In the low cognitive load experimental condition, the sequence was composed of 4 digits also randomly selected (i.e. “2148”, “2571”, “9521”). The 9 and 4 digits sequences are based on the limits of human capacity to process information proposed by Miller with the magical number 7 ± 2 (Miller, 1956).

Third, a screen instructed participants that they would see the video and had to keep the digits in memory. The video presented the NAO robots interacting with a human, an object and another NAO for 1.36 minutes. The video came from an Aldebaran Nao presentation video (from <https://www.youtube.com/watch?v=rSKRgasUEko>). We chose a video format as it has been shown that perceiving a moving stimulus with animacy characteristics increased the social cognition activation to prime social information processing (Wheatley et al., 2007), a position that was already supported by other seminal studies on anthropomorphism (Berry et al., 1992; Heider & Simmel, 1944). For instance, Berry and colleagues (1992), using the well-known Heider and Simmel (1944) animated film demonstrated that disruption of shapes' animation

reduced anthropomorphism while strobe-like stuttering disruption of patterns eliminated it. The authors argued that dynamic transformation was used as evidence of intentionality. Besides, the NAO robot presents the more important facial features that are eyes and mouth to energize anthropomorphic perception (DiSalvo et al., 2002). To control from external priming effect, the video was cut to not display any logo and sound. After the video participants had to write the sequence of digits and judge their agreement about two models to evaluate the (subjective) cognitive effort induced by the experimental condition (i.e. “Regarding the memory task you performed, the task was very complex”, “Regarding the memory task you performed, you provided a very high mental effort to solve it”, $\alpha = .85$) (Brünken et al., 2004; Leppink et al., 2015; Paas et al., 1994).

Finally, they completed a questionnaire adapted from Haslam dehumanization taxonomy (Haslam, 2006; Spatola et al., 2018a, 2019a) and the Robotic Social Attribute Scale (RoSAS) (Carpinella et al., 2017). We used these questionnaires as a measure of the conceptual distance between participants and robots and anthropomorphic attributions respectively.

Attitudes toward robots. The NARS scale is designed to explain participants' evaluations of robot behaviour styles in a Human-Robot Interaction (HRI) study. The NARS scale is made of 3 subscales: 1) Relational attitudes ($\alpha = .73$) (e.g. I would feel very nervous just standing in front of a robot), 2) Future/social influence ($\alpha = .70$) (e.g. I am concerned that robots would be a bad influence on children), 3) Actual interactions and situations ($\alpha = .75$) (e.g. I would feel relaxed talking with robots). For each dimension, participants rated whether they agreed or disagreed (from 1 to 7).

Anthropomorphism measure 1. Participants also filled out the humanness scale based on Haslam's dehumanization taxonomy with human uniqueness (e.g., moral sensibility) and human nature (e.g. interpersonal warmth) as the humanization dimension ($\alpha = .80$), and the

animalistic dehumanization (e.g., irrationality) and mechanistic dehumanization (e.g., inertness) as the dehumanization dimension ($\alpha = .68$). Again, for each dimension, participants rated whether they agreed or disagreed (from 1 to 7) to characteristics attributed to the robot presented on the video.

Anthropomorphism measure 2. Finally, participants filled out the RoSAS with warmth (e.g., happy) ($\alpha = .80$), competence (e.g. capable) ($\alpha = .70$), and discomfort (e.g., scary) ($\alpha = .83$) dimensions. For each dimension, participants rated whether they agreed or disagreed (from 1 to 7) to characteristics attributed to the robot.

2.2 Results

2.2.1 Control analyses

On prior attitudes of participants toward robots, a multivariate ANCOVA showed that there was no differences between participants assigned to high vs. low cognitive load experimental conditions on any of the three dimensions: Relational attitudes, $F(1,128) = .12$, $p = .735$, $\eta^2_p < .01$, 95%CI [-.29, .42], Future/social influence, $F(1,128) = 1.23$, $p = .270$, $\eta^2_p = .01$, 95%CI [-.58, .16], and Actual interactions and situations, $F(1,128) = .26$, $p = .613$, $\eta^2_p < .01$, 95%CI [-.34, .57]. As expected, results showed that participants in the high load cognitive load condition (compared to low cognitive load condition) declared higher cognitive effort during the task, $F(1,128) = 37.11$, $p < .001$, $\eta^2_p = .23$, 95%CI [1.16, 2.28].

2.2.2 Cognitive load effect on anthropomorphic attributions

To test the effect of the cognitive load manipulation (high vs. low) on humanness scores (humanization, de-humanization) and RoSAS scores (warmth, competence and discomfort), we conducted a multivariate ANCOVA including the cognitive effort measure as a covariate. Results showed no significant results, all $ps > .05$.

To test whether the cognitive effort declared by participants was related to NAO anthropomorphic and human traits attributions we conducted regression analyses with the level of cognitive effort as predictor of attributions scores. Results showed that the higher the declared cognitive effort the higher the warmth, $b = .29$, $t(128) = 3.41$, $p = .001$, $R^2 = .08$, 95%CI [.08, .30], competence, $b = .28$, $p = .007$, $R^2 = .06$, 95%CI [.04, .23], and human traits attributions, $b = .34$, $t(128) = 4.06$, $p < .001$, $R^2 = .12$, 95%CI [.10, .29]; also the lower the discomfort, $b = -.22$, $t(128) = -2.59$, $p = .011$, $R^2 = .05$, 95%CI [-.27, -.04], and dehumanization attributions, $b = -.30$, $t(128) = -3.50$, $p = .001$, $R^2 = .09$, 95%CI [-.24, -.07] (Figure 2).

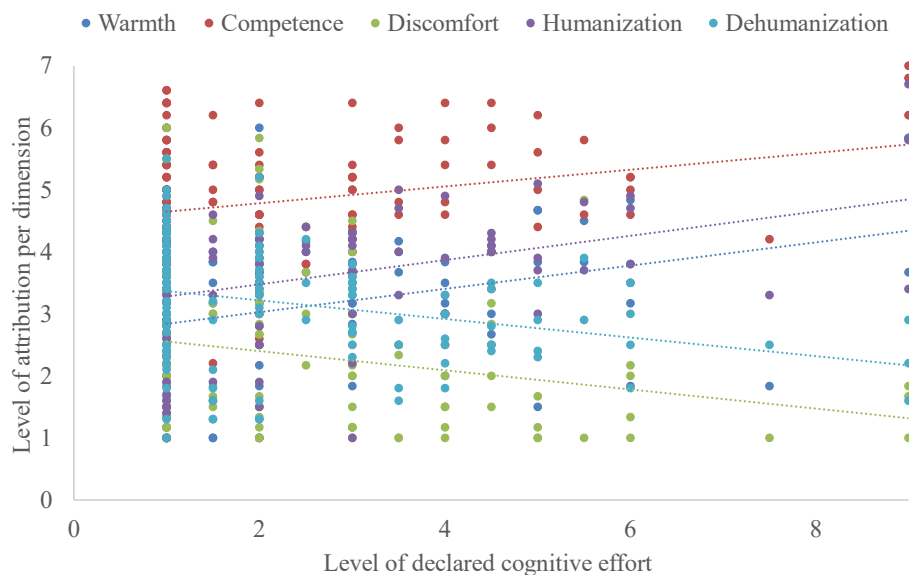


Figure 2. Level of warmth, competence, discomfort, humanization and dehumanization attribution as a function of declared level of cognitive effort

2.3 Discussion experiment 1

The present study aimed at investigating whether, in situations of high cognitive load, participants would anthropomorphize the robot being presented to a larger extent than in low cognitive load situations, expected if they fail to inhibit the default social process as per our

hypotheses. Results showed no evidence for differences in the evaluation of the robot being presented on the video according to the objective level of cognitive load. However, we found on subjective cognitive effort a positive relation to anthropomorphic and human traits attributions. The present results argue for an influence of subjective cognitive effort on anthropomorphism toward robots. However, as the cognitive load manipulation did not modulate attributions, this assumption has to be taken with caution. Also, the effect of the subjective cognitive effort showed low R^2 . An hypothesis to explain this absence of effect could be that the cognitive load manipulation was too low to produce any effect, as such, despite the significant results, the cognitive effort was relatively low. The consequence would be that only participants who experienced difficulty with the task showed the hypothesized effect. The second experiment tested whether this result could be strengthened with a more active interfering task.

3 Experiment 2

We replicated the study testing our cognitive load hypothesis by choosing a different interfering task to maximize the cognitive load effect. We replaced the memory task by a mental calculation task presented during the video (Ayres, 2001; Lamberts et al., 2000). The main difference between the memory task and the mental calculation task is the presentation of new stimuli during the video which should increase the level of cognitive load compared to the first experiment and strengthened the results.

Again, we hypothesized that in a situation of high cognitive load, participants should not be able to inhibit the default social process and should anthropomorphize the robot displayed on the video to a larger extent than participants in a low cognitive load situation. Also, the higher the level of cognitive effort declared by participants, the higher the anthropomorphic attributions.

3.1 Method

Participants were 75 males, 31 females and 2 others ($M_{age} = 19.2$, $SD = 2.5$) recruited online. The experiment was processed with Qualtrics in groups. Participants took part in this experiment on a voluntary basis. We estimated the requested sample size with G*Power, with a medium effect size, a 0.05 α , power set at 0.90 and a 1/1 between-group ratio (Faul et al., 2007). The requested sample size was 108.

The experimenter informed participants that they will take part in a cognitive test. Instructions provided the schedule of the experiment: 1) a questionnaire (NARS scale), 2) a video to watch on which, subsequently, 3) questions would be asked (i.e. the same video as in experiment 1). In contrast to the first experiment, one-digit numbers (e.g., 5) were displayed during the video. Participants were randomly assigned to one of our two conditions. Half the participants were instructed to sum all numbers (i.e. low cognitive load experimental condition), while the other half the instruction was to multiply them (i.e. high cognitive load experimental condition). We choose these two arithmetic tasks because they both rely on the same memory network (Geary et al., 1986), however, multiplication requires more cognitive resources than addition (Sweller, 1994). The location and timing of digits were identical across conditions.

After the video participants had to write the results of their calculation and rate the same two models as in experiment 1 to evaluate the (subjective) cognitive effort induced by the experimental condition (i.e. “Regarding the memory task you performed, the task was very complex”, “Regarding the memory task you performed, you provided a very high mental effort to solve it”, $\alpha = .88$) (Brünken et al., 2004; Leppink et al., 2015; Paas et al., 1994). Finally, they completed the same questionnaires as experiment 1.

Attitudes toward Robots. Again, we used the NARS scale made of 3 subscales: 1) Relational attitudes ($\alpha = .71$), 2) Future/social influence ($\alpha = .69$), 3) Actual interactions and situations ($\alpha = .67$).

Anthropomorphism measure 1. After watching the video, participants filled out the humanness scale based on Haslam's dehumanization taxonomy with humanization ($\alpha = .76$), and dehumanization ($\alpha = .70$) dimension.

Anthropomorphism measure 2. Finally, participants filled out the RoSAS with warmth ($\alpha = .77$), competence ($\alpha = .50$, non-reliable), and discomfort ($\alpha = .78$) dimensions.

3.2 Results

3.2.1 Control analyses

A multivariate ANOVA showed that there was no significant prior differences between participants in high cognitive load vs. low cognitive load experimental conditions on the three dimensions of the NARS filled prior to video exposure: Relational attitudes, $F(1,107) = .01$, $p = .927$, $\eta^2_p < .01$, 95%CI [-.38, .42], Future/social influence, $F(1,107) = .64$, $p = .424$, $\eta^2_p = .01$, 95%CI [-.22, .51], and Actual interactions and situations, $F(1,107) = .38$, $p = .539$, $\eta^2_p < .01$, 95%CI [-.33, .62]. It should be mentioned that the Cronbach's alpha reliability estimate was questionable for the last two dimensions (Brown, 2002; Cortina, 1993; Cronbach, 1951). As expected, results showed that participants in the high load cognitive load condition (compared to low cognitive load condition) declared higher cognitive effort during the task, $F(1,107) = 95.60$, $p < .001$, $\eta^2_p = .47$, 95%CI [2.47, 3.72].

3.2.2 Cognitive load effect on anthropomorphic attributions

To test the effect of the cognitive load manipulation (high vs. low) on humanness scores (humanization, de-humanization) and RoSAS scores (warmth, competence and discomfort), we conducted a multivariate ANCOVA including the cognitive effort measure as a covariate.

Results showed that participants in the high cognitive load condition attributed more human traits, $F(1,107) = 48.91, p < .001, \eta^2_p = .32, 95\%CI [.69, 1.24]$, and more warmth, $F(1,107) = 40.58, p < .001, \eta^2_p = .28, 95\%CI [.82, 1.55]$, and more competence, $F(1,107) = 7.73, p = .006, \eta^2_p = .07, 95\%CI [.12, .73]$, to the robot compared to participants in the low cognitive load condition. All the other $ps > .05$.

To test whether the level of cognitive effort declared by participants could explain the effect of the experimental condition on anthropomorphic attributions we conducted mediation analyses using PROCESS plugin in SPSS (Bolin, 2014; Fritz et al., 2012; Hayes & Rockwood, 2017; Hayes & Scharkow, 2013; Preacher & Hayes, 2004; Shrout & Bolger, 2002; Snyder, 1961; Yzerbyt et al., 2018) for on humanness and warmth scores, with experimental conditions (high vs. low cognitive load) as a predictive factor (X) and the level of declared cognitive effort as mediating factor (M, bootstrap = 10000) (MacKinnon et al., 2004) (Figure 3).

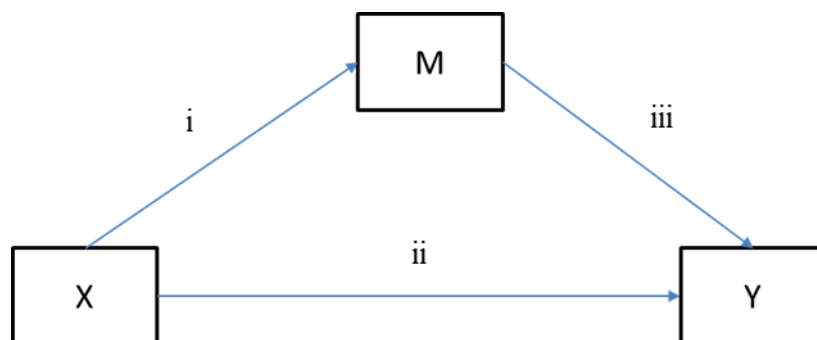


Figure 3. Mediation model

Humanization. In the mediation model (corresponding to (ii) in Figure 2) the direct path between experimental conditions (X) and human traits attributions remained significant, $b = .24, t(106) = 2.36, p = .020, 95\%CI [.07, .84]$ (Figure 3). Results also showed that the higher the declared level of cognitive effort (M), the more participants attributed humanness traits to the robot presented in the video (Y), $b = .44, t(106) = 4.76, p < .001, 95\%CI [.12, .29]$ (corresponding to (iii) in Figure 2). The mediation path $X \rightarrow M \rightarrow Y$ was also significant $b =$

.42, 95%CI [.23, .65]. In other words, dissimilar to experiment 1, participants in the high cognitive load condition attributed higher human traits to the robots compared to participants in the low cognitive load condition and this effect was explained, at least in part, by the higher level of participants declared cognitive efforts (Figure 4).

Warmth. In the mediation model including the warmth scores as DVs, the effect of the experimental condition became non-significant, $b = .06$, $t(106) = .21$, $p = .833$, 95%CI [-.55, .69]. We found that the higher the level of subjective cognitive effort (M), the more participants attributed warmth characteristics (Y) to the robot presented in the video, $b = .45$, $t(106) = 3.22$, $p = .002$, 95%CI [.09, .36]. Our analyses argue for a mediation of the effect of the experimental condition on warmth attribution by the declared level of cognitive effort, $b = .70$, 95%CI [.28, 1.15] (Figure 4).

Competence. The Cronbach alpha was not reliable for this dimension, therefore the following results should be taken with caution. In the mediation model, the effect of the experimental condition on competence attributions became non-significant, $b = -.51$, $t(106) = -1.74$, $p = .084$, 95%CI [-.82, .05]. Also, we found that the higher the declared level of cognitive effort (M), the more participants attributed competence characteristics (Y) to the robot presented in the video, $b = .52$, $t(106) = 3.53$, $p < .001$, 95%CI [.08, .28]. Our analyses showed a significant mediation of the effect of the experimental condition on competence attributions by the declared level of cognitive effort, $b = .61$, 95%CI [.30; .93] (Figure 4).

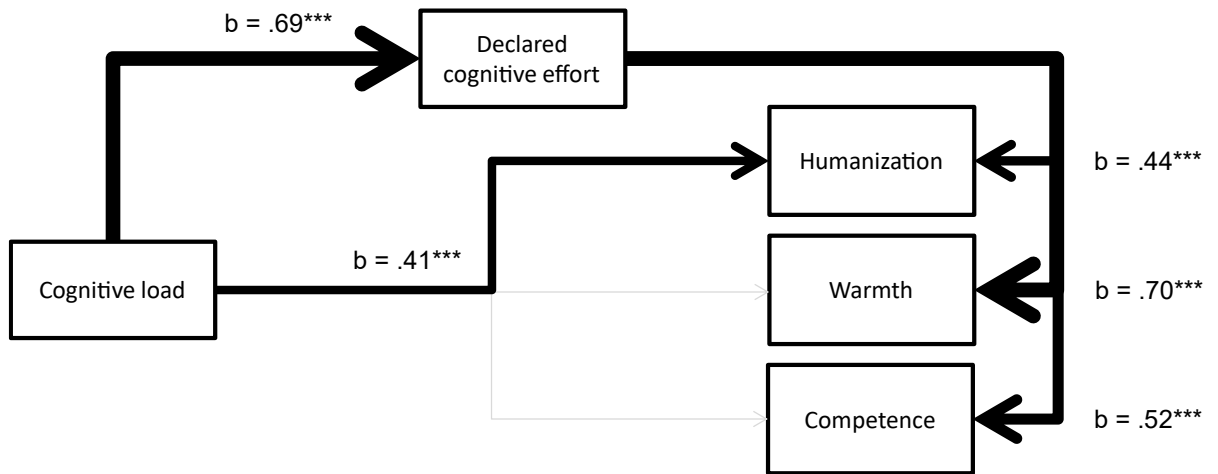


Figure 4. Mediation models Experimental conditions → Cognitive effort → humanization, warmth and competence.

3.2.3 Exploratory analyses

We conducted an ANOVA including the cognitive load experimental condition (high vs. low) and the experiment (experiment 1 vs. experiment 2) as IVs and the declared cognitive effort as DVs to test whether the change of interfering task increased the cognitive effort declared by participants. We found an experiment by cognitive load experimental condition interaction, $F(1,236) = 9.21, p = .003, \eta^2_p = .04$. Contrast with Bonferroni correction showed that while there was no difference between study on the low cognitive load conditions, $F(1,233) = 2.01, p = .157, \eta^2_p = .01, 95\%CI [-.16, 1.01]$, participants in experiment 2 high load condition declared higher cognitive effort (compared to their experiment 1 counterparts), $F(1,233) = 33.03, p < .001, \eta^2_p = .12, 95\%CI [1.11, 2.72]$.

3.3 Discussion experiment 2

The second study aimed at replicating and strengthen the results of study 1 investigating whether, in situations of high cognitive load, participants would fail to inhibit the default social process and therefore anthropomorphize the robot being presented to a larger extent than in low cognitive load situations. Congruently with our hypotheses, we found significant differences in

the evaluation of the robot according to the objective and subjective level of cognitive load. In addition, the increase of positive anthropomorphic inferences (warmth, competence) and human traits attribution (humanization) was significantly mediated by the reported level of cognitive effort (all CIs excluded 0).

In line with our hypothesis experiment, 2 results argue for an important influence of cognitive resources available when making an anthropomorphic judgement about a humanoid robot. The fewer resources are available, the higher are ratings in positive scales and lower on negative scales. These results argue for an automatic process of robots using the social cognitive system reflected in higher anthropomorphism when fewer control resources are available. And corollary, reduced positive anthropomorphism observed in the low cognitive load condition would reflect, within the present dual-route framework, increased reliance on controlled cognition, namely, adopting a physical stance to make judgements about the robots being shown on the video.

Using a more active task than memory maintenance (online calculation in experiment 1), the second study makes it possible to conclude for a role of the amount of available resources in the structuration of participants' mental representation of robots and, as a consequence, anthropomorphism processes that result from this representation.

4 Experiment 3

Experiment 1 and 2 aimed at investigating the cognitive control process (relying on available cognitive resources) in anthropomorphism. Taken together the results argue that people fail to inhibit the anthropomorphic inferences only when the depletion of the cognitive resource is sufficient. Otherwise, the control process allows a switch from the social to the physical cognition resulting in less anthropomorphic attributions. However, as mentioned in the introduction, some authors argued that the goal was of prime importance and then

anthropomorphism was engaged when explaining or predicting the behaviour of the target entity was important (Caporael, 1986; Hansen et al., 1976; Meyer et al., 2012; Spunt & Lieberman, 2013).

In experiment 3, we aimed at disentangling how focusing on the intention (Spunt & Lieberman, 2013) or the goal of actions (Gazzola et al., 2007) may result in different influence of cognitive load, considering the intentional stance as the default mode (Dennett, 1971; Spunt et al., 2015). We asked participants to watch a video presenting a robot with either the objective of answering the question “Why the robot is doing that?” (“why” goal condition focusing on the intention behind the action) or the question “What is the robot doing” (“what” goal condition focusing on the goal of the action). With this experimental manipulation, we assumed de Lange and colleagues limit showing that deliberate reflection on goals and intention could bias mentalization process (de Lange et al., 2008). We also introduced a control condition of passive observation (observation goal condition) in which participants were informed that they will have to watch a video but there was no mention about further questions about it. The purpose was to avoid participants to focus on the robot’s actions. Also, we manipulated the cognitive load of participants (high vs low cognitive load conditions).

In line with our previous hypotheses, participants in the high cognitive load condition should anthropomorphize more the robot than participants in the low load condition. We also hypothesized that participants in the “why” goal condition should anthropomorphize the robot more than the “what” condition as the focus would be made on intentional causal attribution in the former and on a more pragmatic level of analysis in the later (Dennett, 1971; Heil & Heil, 2019; Meyer et al., 2012; Spunt & Lieberman, 2013). In addition, both should result in higher anthropomorphic attributions compared to the (passive) observation goal condition because the video should be considered as an interfering element in the task resulting in a low level of attention on the robot. However, in line with Spunt and Lieberman (2013), under high cognitive,

when participants are primed by the “why” goal, an impairment of anthropomorphic attribution should occur because to reflect deliberately on intention could impair mentalization process (de Lange et al., 2008). Conversely, under the “what” goal, because people should not reflect on intention, participants should attribute more anthropomorphic traits to the robot in the high cognitive load condition (compare to the low load condition). Therefore, we hypothesized that the goal should condition the cognitive load effect.

4.1 Method

Participants were 93 males, 206 females and 53 others or who preferred not to answer (Mage = 22.2, SD = 8.7) recruited online. The experiment was processed with Qualtrics. Participants took part in this experiment on a voluntary basis. We estimated the requested sample size with G*Power, with a medium effect size, a 0.05 α , power set at 0.90 and a 1/1 between-group ratio (Faul et al., 2007). The requested sample size was 350.

First, building upon Spunt and Lieberman (Spunt & Lieberman, 2013), the experimenter informed participants that the experiment consisted in a memory test. Instructions provided the schedule of the experiment: a video to watch on which, subsequently, questions would be asked. Participants were then randomly assigned to one of the two experimental conditions. Half of the participants were instructed to remember an easy numbers sequence (e.g., 555-5555; further mentioned as the *low load condition*) while the other half had to remember a difficult numbers sequence (e.g., 813-5647; further mentioned as the *high load condition*).

Second, we manipulated the goal factor by instructing participants to adopt one of three goals while watching the video: i) to actively recognize the action (“You will have to describe WHAT the robot is doing”; the *what goal* condition), ii) to understand the motive behind the action (“You will have to describe WHY the robot is doing these actions”; the *why goal* condition), or iii) to passively observe the video (the *observation goal* condition). The same video as in

experiment 1 and 2 was used to ensure the comparison between the experiments. Participants were randomly assigned to a specific goal.

Third, after the video participants had to recall the numbers sequence they had to memorized and to evaluate the robot on the Human-Robot Interaction Evaluation Scale (HRIES) (Spatola, Kühnlenz, et al., 2020) which contains four dimensions of robot evaluation including Sociability (4 items, e.g., Warm, $\alpha = .90$), Agency (4 items, e.g., Self-reliant, $\alpha = .73$), Animacy (4 items, e.g., Alive, $\alpha = .63$), and Disturbing (4 items, e.g., Creepy, $\alpha = .83$). This scale makes it possible to evaluate static, in motion or interactive robots on a broad spectrum of anthropomorphic attributions. For each item, participants rated whether they agreed or disagreed (on a scale from 1 to 7) on the attribute related characteristics to the iCub robot (presented on a picture above the scale). (i.e., “For each trait, you will have to evaluate whether, according to you, it corresponds or not to the robot that is presented to you.”). For each trait, a 7-points slider scale was presented from 1 “not at all” to 7 “totally”. We chose the HRIES in the third experiment because the scale has been developed as an aggregated and updated version of the two anthropomorphic measure previously used during the current study project. It also makes it possible to test the hypothesis with a slightly different but comparable measure.

4.2 Results

We conducted a MANOVA including the type of goal (“*what*” vs “*why*” vs *observation*) and the cognitive load condition (*high load* vs *low load*) as independent variables and the HRIES dimensions scores as dependent variables. Results showed a main effect of the goal on Agency, $F(2,349) = 55.11, p < .001, \eta^2_p = .21$, Sociability, $F(2,349) = 34.49, p < .001, \eta^2_p = .08$, and Animacy, $F(2,349) = 7.22, p = .001, \eta^2_p = .04$, dimensions but not the Disturbing one, $F(2,349) = 1.81, p = .165, \eta^2_p = .01$. Pairwise contrasts with Bonferroni correction showed that all comparison were significant except for the “*what*” and “*why*” goal conditions on animacy dimension (see table 1). Participants attributed anthropomorphic characteristics to a higher

extent in the “*what*” and “*why*” goal conditions compared to the *observation* one and in the “*why*” goal condition compared to the “*what*” one.

	Goal		t statistics	p.value	95%CI	
Disturbing	Observation	What	$F(1,232) = .008$.999	-.41	.39
	Observation	Why	$F(1,231) = 2.38$.332	-.13	.67
	What	Why	$F(1,235) = 3.09$.276	-.39	.41
Agency	Observation	What	$F(1,232) = 40.19$	<.001	-1.27	-.58
	Observation	Why	$F(1,231) = 68.50$	<.001	-1.69	-1.00
	What	Why	$F(1,235) = 7.71$.010	-0.76	-.08
Sociability	Observation	What	$F(1,232) = 8.91$.008	-1.07	-.12
	Observation	Why	$F(1,231) = 27.63$	<.001	-1.56	-.61
	What	Why	$F(1,235) = 5.91$.036	-.965	-.02
Animacy	Observation	What	$F(1,232) = 6.08$.033	-.77	-.02
	Observation	Why	$F(1,231) = 13.08$.001	-.95	-.20
	What	Why	$F(1,235) = 1.32$.715	-.55	.19

Table 1. Contrasts’ statistics with Bonferroni correction for each anthropomorphic dimensions.

We also found significant interactions between the goal and the cognitive load manipulation on Agency, $F(2,349) = 23.12$, $p < .001$, $\eta^2_p = .12$, Sociability, $F(2,349) = 8.34$, $p < .001$, $\eta^2_p = .05$, and Animacy, , $F(2,349) = 6.79$, $p = .001$, $\eta^2_p = .04$. The interactions are described at continuation.

To investigate these interactions we used a tree-based classification model to predict the values of the dependent variables (Disturbing, Agency, Sociability, Animacy) with a Chi-squared Automatic Interaction Detection (CHAID) algorithm. This algorithm identifies the independent variables (predictors) that have the strongest interaction with the dependent variable. Non-significant predictors are merged in one category. This process allows extracting a hierarchy from the data taking into account the different level of interaction between the predictors (i.e., goals and cognitive load). Overall results showed a primacy of the goal (*observation* vs. “*why*”

vs. “*what*”) in all anthropomorphic dimensions except disturbing (see figure 5, 7, 8). The effect of the cognitive load was therefore conditioned by the type of goal supporting our hypothesis.

Agency. Results on Agency attributions showed that the three types of goal (*Observation*, “*what*”, “*why*”) primed the effect of the cognitive load on attributions (Figure 5).

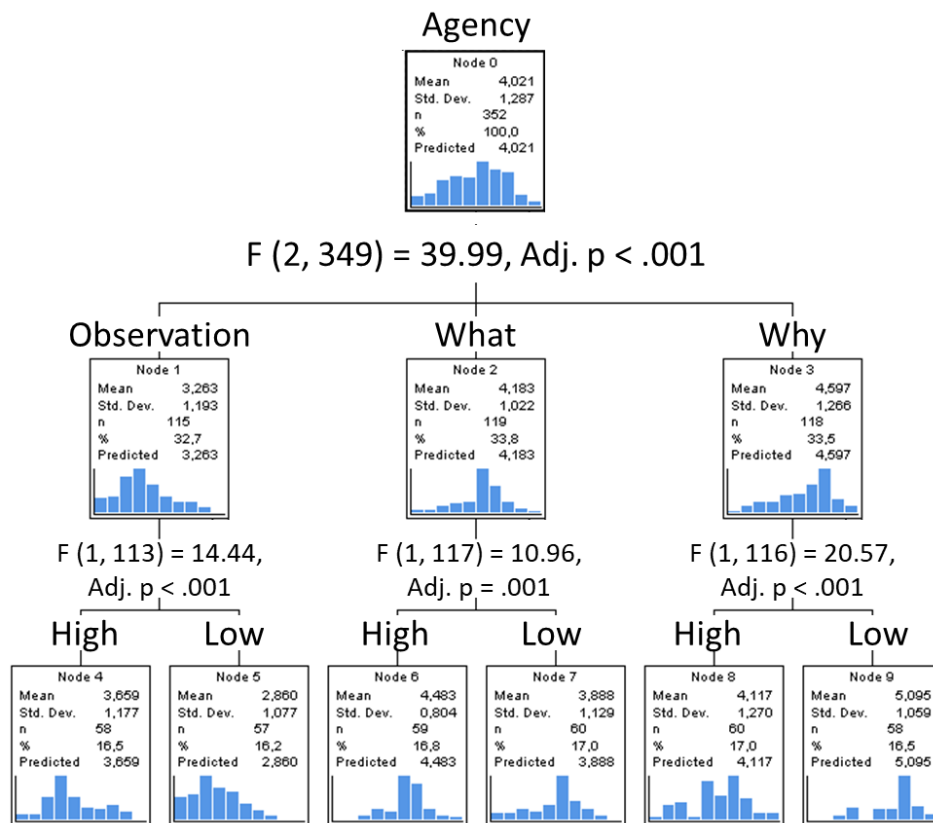


Figure 5. Decision tree to predict agency traits attribution as the function of goal and cognitive load

While participants attributed more traits to the robot in the “*observation*”, $F(1,113) = 14.44, p < .001, \eta^2_p = .11, 95\%CI [.48, 1.22]$, and “*what*”, $F(1,117) = 10.96, p = .001, \eta^2_p = .09, 95\%CI [.24, .95]$, goals conditions in the high compared to low cognitive load; in the “*why*” condition participants attributed less agency traits in *high* (compared to *low*) cognitive load condition, $F(1,116) = 20.57, p < .001, \eta^2_p = .15, 95\%CI [-1.41, -.55]$ (Figure 6).

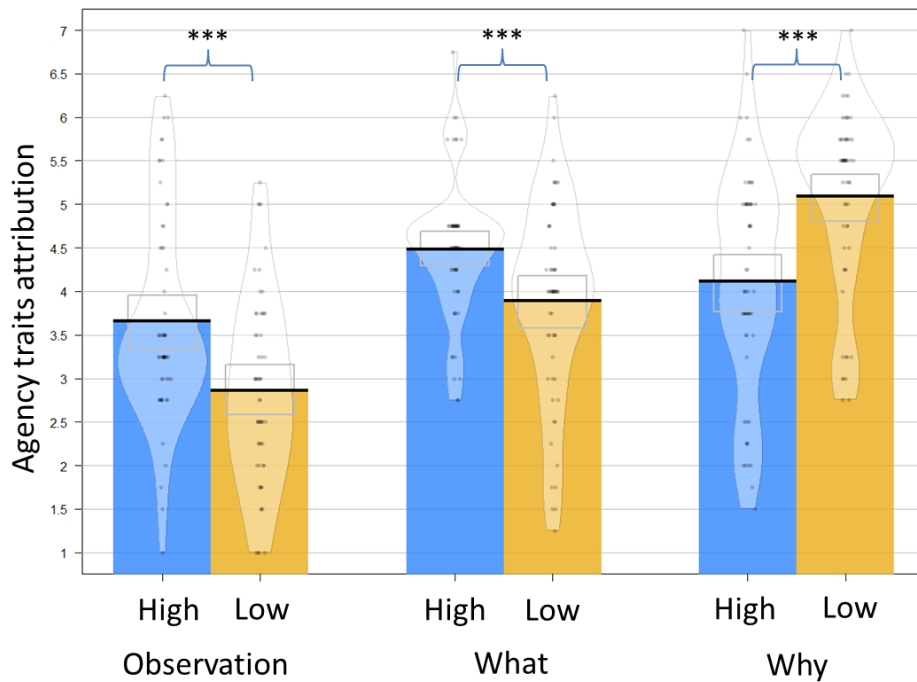


Figure 6. Agency traits attribution as the function of the goal (observation vs “what” vs “why) and cognitive load (high vs low). *** : $p < .001$.

Sociability. On the sociability dimensions, the goal primed the effect of the cognitive load in the *observation* and “*why*” but not “*what*” goals (Figure 7). While there was no effect of cognitive load in the “*what*” goal condition, $F(1,117) = 2.68, p = .104, \eta^2_p = .02, 95\%CI [-.09, .98]$, participants in the observation condition attributed more sociability traits to the robot in the *high* (compared to *low*) cognitive load condition, $F(1,113) = 9.79, p = .002, \eta^2_p = .08, 95\%CI [.31, 1.40]$, and the reverse pattern was true for the “*why*” goal condition, $F(1,116) = 5.81, p = .017, \eta^2_p = .05, 95\%CI [-1.27, -.13]$.

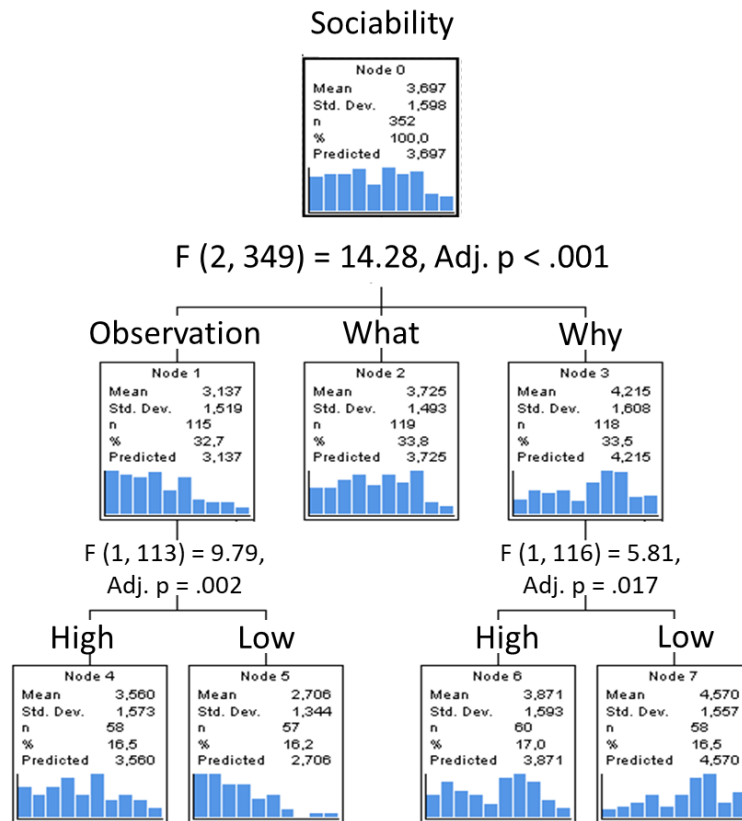


Figure 7. Decision tree to predict sociability traits attribution as the function of goal and cognitive load

Animacy. Finally, on Animacy attributions, results showed that the effect of the load was only dependent of the *observation* goal while there was no effect in the “*what*” and “*why*” conditions ($p_s > .05$) (Figure 8). Results showed higher attribution in *high* (compared to *low*) load in the *observation* condition, $F(1,113) = 6.08, p = .015, \eta^2_p = .05, 95\%CI [.12, .99]$. The Cronbach alpha of the Animacy dimension was not reliable (Cronbach, 1951), therefore results on this dimension should be taken with caution.

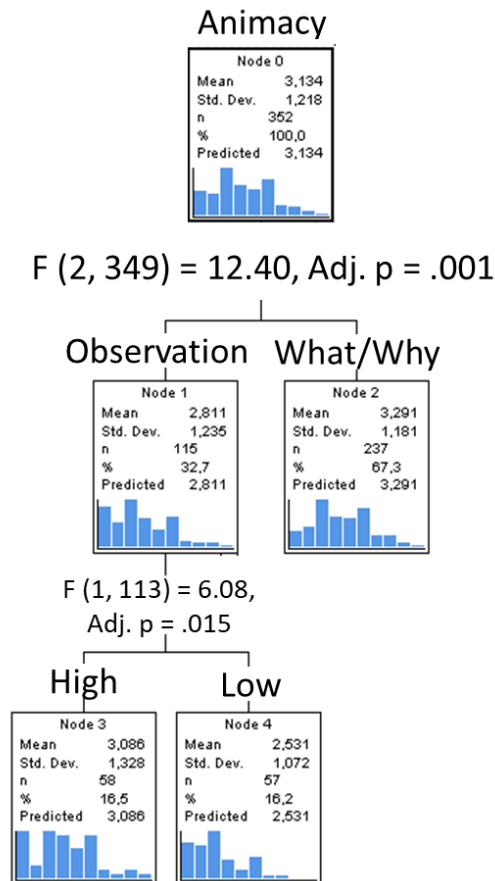


Figure 8. Decision tree to predict animacy traits attribution as the function of goal and cognitive load

4.3 Discussion Experiment 3

Experiment 3 aimed at evaluating how could the type of interpretative goal modulate the influence of cognitive load on anthropomorphic attributions. Besides, we aimed at investigating whether the goal was the modulator or was modulate by the cognitive load.

First, In line with previous studies (Meyer et al., 2012; Spunt & Lieberman, 2013), we found that individuals attributed more anthropomorphic traits when they were explicitly asked to (“*why*” goal condition). Participants anthropomorphized the robot more when they had to reflect on the intention of the robot (“*why*” goal condition) compared to the condition priming

an emphasis on the action (“*what*” goal condition), the both resulting in higher attribution compared to the mere observation goal condition.

Second, we found a goal by cognitive load interaction. Participants in the *observation* and “*what*” goals attributed more anthropomorphic characteristics to the robot in *high* compared to *low* load condition while it was the opposite in the “*why*” condition. This last result is in line with Spunt and Lieberman findings argues that when people are primed to reason about the intention of an agent, the increase of cognitive load impairs anthropomorphism (Spunt & Lieberman, 2013). However, when the emphasis is made on another characteristic of the agent behaviour such as “what the agent is doing” or mere observation, high cognitive load (compared to low load) result in higher anthropomorphic attributions (similar to the results in experiment 2).

The main differences between the “*why*” and the “*what*” and mere *observation* goal is that when people deliberately reflect on intention, the synergy between mirror neurons and mentalizing systems which play a central role in social cognition (de Lange et al., 2008; Sperduti et al., 2014; Van Overwalle, 2009; Van Overwalle & Baetens, 2009). In other terms, in the “*why*” goal, because of the goal, the automatic synergy between mentalization and embodiment could be bypassed and become a controlled process (unlike when participants do not are explicitly required to deliberate on intention). This interpretation is in line with fMRI research demonstrating increased activity in mentalization areas and a loss of the synergy between the mirror and mentalizing systems when participants are explicitly instructed to understand the intention of an agent (de Lange et al., 2008; Spunt & Lieberman, 2013). Spunt and colleagues, in a further study, showed that mentalization is primed by the default mode of human brain function, and could be highly sensitive to the type of task demand (Spunt et al., 2015). Therefore, if the attribution of intention become a controlled process (requiring resources), cognitive resources depletion could impair the process of intention attribution

(Engström et al., 2017) or, as named by Dennett, the adoption of the intentional stance (Dennett, 1971).

5 General discussion

Across 3 experiments, we aimed at investigating the process of anthropomorphism regarding the cognitive control theory. More specifically, we grounded our hypotheses in a framework that posits two cognition systems in human perception. First, a social cognition as an automatic path based on the social perception of the environment and the activation of social representation (social schema, social scripts). Second, a physical cognition as a non-automatic path focused on the concreteness of stimuli and empirical rules (Jack, Dawson, Begany, et al., 2013; Mars et al., 2012; Tavares et al., 2008). With respect to this framework, we proposed that human perception facing human-like agents' (such as humanoid robots) behaviours, would be first social resulting in anthropomorphic attributions but could be actively controlled to switch to physical reasoning about the target (Heider & Simmel, 1950; Jack, Dawson, Begany, et al., 2013; Tavares et al., 2008) and more target-specific (compared to anthropomorphic) representation. Finally, we aimed at investigating the interaction between cognitive load and goal of evaluation, especially when individual are required to deliberate on the intentional motivation of an observed agent. In line with previous studies (de Lange et al., 2008; Spunt & Lieberman, 2013), we hypothesized that when individuals observe a robotic agent, a mental representation of this artificial agent is built based on information from the social cognition that could be actively bypassed by the physical cognition if sufficient cognitive resources are available. (Experiments 1, 2). However, this social vs physical inferences could be modulated by the goal of observation as explicitly focused on mental states or not switching an automatic process in a controlled process and countervailing with its standard *modus operandi* (Experiment 3).

In line with previous research on mental representation (Huang & Awh, 2018), our results (experiment 1 and 2) demonstrate that judgments about robots are based on a pre-structured mental representation rather than a reactivation and reevaluation of the percept (i.e., the sensitive content of the video). This view of anthropomorphism is in accordance with Epley and colleagues model who posit that the core of anthropomorphism is the representation of the object rather than the object (Epley et al., 2007; Waytz et al., 2010). However, the structuration of the representation may be conceptualized as an iterative reprocessing model (Cunningham & Zelazo, 2007). Indeed, if the process was only one-way, we could not observe a switch from a representation mainly based on social information (anthropomorphic) to a representation mainly based on physical information (non-anthropomorphic). In other word, the representation of the agent remains malleable as long as the observation goes. Therefore, we could assume that the longer the observation, the more structured and the less malleable the representation (Carlston & Smith, 2007). The reason is that the accumulation of information runs against the volatility of the representation. The higher the amount of information, the higher the inertia in which the weight of an information is pondered by its coherence with the social vs. physical dominance of the representation (e.g. confirmation bias). Therefore, in HRI, the first representation of robots remain crucial to define attitudes. Because the access to direct interaction and information about these new artificial agents potentialities and limits are limited, representation are often biased by the pop-culture and unrealistic positive or negative expectation. This phenomenon has been described as the “Hollywood robot syndrome” (Sundar et al., 2016) and could explain the actual high level of inter-individual differences in attitudes (e.g. interest, fear) and willingness to interact and accept robots. Also, because our results argue for the importance of the first representation (even biased), assessment method of initial representations and its malleability are crucial to develop. Here we used explicit measures but we could posit that implicit measures could be better suited. Explicit measures operate on a

conscious level and generally use self-reports (e.g. questionnaires), while implicit measures rely on unconscious and automatic processes, and typically use reaction time paradigms or implicit association test (De Houwer et al., 2009). Research suggests that implicit measures might constitute better predictors of future intentions and behaviors (Kurdi et al., 2019), and thus be more representative of real representations than explicit declarations. Implicit measures have also proved to be well equipped to predict the behavioral consequences of individuals' implicit representations either with humans (Friese et al., 2008; Kurdi et al., 2019) or robots (Spatola & Wudarczyk, 2020, 2021).

As posited by the cognitive control framework –in which we anchored our hypotheses–, the amount of cognitive resources available during this iterative process of representation structuration is crucial to take into account (Lavie, 2010). As hypothesized, we found that cognitive load might modulate the amount of anthropomorphic attribution associated with the robot. The lower the amount of cognitive resources available, the higher the anthropomorphic attributions. These results are in line with decades of social psychology literature, for example pertaining to stereotypes, demonstrating that people lacking of cognitive resources tend to reason about their environment by reducing the complexity of the environment in using automatic representation (Sherman et al., 1998). Taken together, results from studies 1 and 2 support the hypothesis of anthropomorphism as a default mechanism that might be controlled when sufficient resources are available. In experiment 1, results showed that the higher the amount of cognitive resources hold by the task, the higher the anthropomorphic attributions.

Anthropomorphism makes it easier to analyze robot interactions because it is based on the best-known model of the human observer: the human (Epley et al., 2007), despite a weaker perceptual input than when observing a real human being (Chaminade et al., 2005). As Dennett pointed out (Dennett, 1971), human observers can often misinterpret information that have been easily extracted and integrated into a representation in a specific situation (e.g., to use human

characteristics easily accessible to structure the mental representation of a robot in case of high cognitive load). When activated, during judgment, for example, they will consider these information recovered from mental representations as accurate regardless of the initial anthropomorphic bias (O'Rourke, 1993). The attribution of human-like characteristic to artificial human-like entities such as robots is therefore bias by these representations. To illustrate the process, we may consider the case of peripheral vision as a metaphor for the mediating effect of the cognitive load on the anthropomorphism process. While observers do not have "a clear image" of the full picture because of the peripheral objects represented in their visual system, they subjectively perceive these peripheral objects as clear because they can structure "clear images" of these objects by completing their representation with usually situationally correlated elements stored in memory, even if the actual sensory information is not complete or accurate. Completing a representation of a robot by using appropriate accessible elements that can easily be retrieved from memory (i.e., human characteristics could be interpreted as a completion and stabilization of the representation strategy) especially if the robot is human-like. This would be even truer when the observer lack of cognitive resources, "blurring the image" or resulting in an incomplete representation. However, when individuals have sufficient cognitive resources to deeply reason about the robot ("to clean the image"), individuals would refer to a more accurate and thus less anthropomorphic representation of the robot, relying mainly upon what Dennett called "the physical/design stance". The results would be a switch from the social cognitive model to the physical cognitive model, due to the rationalization of the robot as an artefact. Saving resources using human schemas (i.e., social cognition) would therefore no longer be necessary and detailed analysis (i.e., physical cognition) would become possible.

Also, our experiment 3 results demonstrated that the goal entailed by observers could modulate how the cognitive load biased anthropomorphic attributions. Priming participants to

adopt, on purpose, the intentional stance resulted in a reversed pattern with lower anthropomorphism in high (compared to low) cognitive load situation. This result argues to further consider how the control (in a task) of a default process may bias this process (Spunt et al., 2015). This result is in line with the CASA hypothesis (Nass & Moon, 2000). While individuals tend to use anthropomorphic attribution to think about a non-human (e.g. robotic) agent, asking them explicitly whether they actually consider them as anthropomorphic will result in a denial of such attribution, and become the dominant response. However, how this denial occurs remained poorly understood. Whether it could arise from reactance (Brehm, 1981; Steindl et al., 2015), ironic rebound (Wegner & Schneider, 2003), contrast comparison between humans and robots (R. Brown & Hewstone, 2005; Spatola & Urbanska, 2019) remained unanswered.

Finally, using different cognitive load manipulations in the 3 experiments, we may now assume that the present effects do not depend on a specific task but more generally on a common process within the cognitive control framework, putatively in relation to cognitive load. These results, again, argue for the crucial role of the context in anthropomorphism. In two different contexts, two different representations and, as a consequence, judgments of the same robot could be produced by the same individual. Taking into account this information we could define social-cognition strategies to increase or reduce the anthropomorphic attribution toward robots on purpose. For instance, it could be considered as positive to anthropomorphize a host robot in a store. Therefore programming the robot to transmit a sufficient amount of information, could reduce the amount of resources available for customers and increase the anthropomorphic attributions. This can be done by mixing the cues that the social cognition will perceive such as facial emotional cues, body gesture either to transcript emotions or orientation indices, the logorrhea and the type of semantic (e.g. pragmatic vs emotional). In this example, the increase of the number of stimuli is twofold: i) the increase of the social signals, triggering the activation

of the social cognition, and ii) the increase of the amount of signal to process, requiring more cognitive resources alleviating the counter-anthropomorphic process. Going further, because anthropomorphism is sensitive to the context, we could posit that the representation could be modulated by the type of cognitive load sources, for example, positive or negative semantic related stimuli (Hermans et al., 2003; Iida et al., 2012; Meisner, 2012) and as such modulate the experience with the robot but also the evaluation of the action of the robot.

Limits

Some limits have to be discussed. First, we cannot consider here the gap with the attribution of these anthropomorphic characteristics between a robot and a human performing the same actions. Thus, even if robots are more anthropomorphic in high cognitive load condition, the distance between them and a complete anthropomorphism (i.e., comparison to the human prototype) can't be evaluated.

Second, our study doesn't make it possible to study the effect of time and the hypothesis of a reduction in anthropomorphic attributions over time. In light of recent brain imaging studies, it could, therefore, be assumed that the difference on structures such as the parietal temporal junction (associated to the attribution of intentionality) or the hypothalamus subnuclei (associated to the social bonding) highlighted by Rauchbauer and colleagues (Rauchbauer et al., 2019) in the HHI and HRI comparisons only appear at certain moments of the interaction, for example, if no differences of activity were found at the beginning of the interaction (Spatola et al., n.d.).

Third, our studies only emphasize the "interpretative" aspect of anthropomorphism. However, Fisher identified two different ways to engage in anthropomorphic thinking: 1) interpretative, the attribution of anthropomorphic characteristics based on observation and 2) imaginative, as the representation of fictional entities detached from concrete perception.

Therefore, we could assume that the cognitive cost or motivation to switch from the social to the physical cognition could be influenced by this prior. Finally, facing robots, the design of the agent can be a huge source of variability (Duffy, 2003). It could be argued that attributions could be modulated by the anthropomorphic characteristics of the design such as the presence of a mouth or eyes (DiSalvo et al., 2002). Still, understanding the actions of robots, even non-anthropomorphic ones, take advantage of the intricate brain mechanisms that developed to understand the actions of our fellow creatures, it is to say mirror neuron and mentalization systems (Gazzola et al., 2007).

Fourth, while participants were on the usual age average in psychological study, this could be a limit to the generalization of the results. The reason is that in relation to technology, age is an important factor. Indeed, acceptance or knowledge are two dimensions correlated to age. Considering that in our proposed model, knowledge about robots is an important part of the use of an anthropomorphic representation (or not) (Epley et al., 2007, 2008; Waytz et al., 2010), age could be a moderator interesting to take into account. Also, the same comment could be done about gender.

Fifth, we agree that other factors that we did not investigate in present study, such as the human-like appearance (Nyangoma et al., 2017; Salem et al., 2013), the motor resonance process (Chaminade et al., 2010, 2012), the social context (Spatola et al., 2018a, 2019a), the observer's knowledge about robots (Epley et al., 2007, 2008; Waytz et al., 2010), could reinforce or interfere with the control of the social vs physical information weight. In other words, the more the context provides social information or triggers social processing, the more it is difficult for the cognitive control to inhibit the default social cognition path.

Finally, as we mentioned in introduction section, the nature of the relationship between social and physical cognition remains an open question. In the parallel-competitive form, both

processes are activated in parallel as information treatment pipelines that weight automatic and controlled information to provide a single output (Smith & DeCoster, 2000). In the default system theories, the social cognition produces the initial output that can be (or not) corrected at a second stage by the physical cognition, similar to evidence accumulator correction model that is computed until the production of the final output (Evans & Stanovich, 2013). The present results do not make it possible to disentangle the two models.

6 Conclusion

Anthropomorphism is fundamental to human-robot (and other non-human agents) interaction as it helps to make sense of their behaviour and is activated by default (as a dominant response) and need to be actively inhibited. To embed anthropomorphism in cognitive control theory and the dual-route (social cognition, physical cognition) framework makes it possible to predict better the situation in which one would engage in anthropomorphism, taking into account the observation goal. Indeed, observation goal, and in particular focusing on intention, modifies how people adopt an intentional stance toward non-human agents. Bringing into the field of consciousness a process activated by default, such as anthropomorphism or the attribution of intentionality, tends to inhibit this process as cognitive resources diminish, conversely to a situation in which one reflects about non-human agents without –intentional inferences– goal.

If the integration of social robots is a possible future for our societies, it seems important to understand how we view them. This understanding is a necessary condition for a positive integration in contexts that can be critical. The present experiments invite us to question the contexts of use of robots and the manipulation of anthropomorphic processes. Depending on the context of use, anthropomorphism may be more or less relevant. For example, robots tend to be used in school contexts requiring students to produce significant cognitive efforts.

Therefore, under constraint of cognitive resources, students, beyond a question of age, will tend to produce anthropomorphic attributions. However, research has shown that during cognitive tasks, such as learning, the presence of a robot characterized by anthropomorphic traits can positively or negatively modulate cognitive performance (Spatola et al., 2018b, 2019b; Spatola, Monceau, et al., 2020; Spatola & Normand, 2021). The more anthropomorphized a robot is, the more it is considered similar to a human in socio-cognitive processes. The result is the reproduction, for example, of the social presence or comparison effects studied by social psychology.

Compliance with Ethical Standards

The study was conducted in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki).

Conflict of Interest:

The authors declare that they have no conflict of interest.

Open Practices

All data are publicly available via the Open Science Framework and can be accessed at <https://osf.io/6zq8n/>

7 References

- Augustinova, M., Silvert, L., Spatola, N., & Ferrand, L. (2018). Further investigation of distinct components of Stroop interference and of their reduction by short response-stimulus intervals. *Acta Psychologica, 189*, 54–62. <https://doi.org/10.1016/j.actpsy.2017.03.009>
- Ayres, P. L. (2001). Systematic Mathematical Errors and Cognitive Load. *Contemporary Educational Psychology, 26*(2), 227–248. <https://doi.org/10.1006/ceps.2000.1051>

- Berry, D. S., Misovich, S. J., Kean, K. J., & Baron, R. M. (1992). Effects of Disruption of Structure and Motion on Perceptions of Social Causality. *Personality and Social Psychology Bulletin*, 18(2), 237–244. <https://doi.org/10.1177/0146167292182016>
- Bolin, J. H. (2014). Hayes, Andrew F. (2013). Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach. New York, NY: The Guilford Press. *Journal of Educational Measurement*, 51(3), 335–337. <https://doi.org/10.1111/jedm.12050>
- Brehm, S. S. (1981). Psychological reactance and the attractiveness of unobtainable objects: Sex differences in children's responses to an elimination of freedom. *Sex Roles*, 7(9), 937–949. <https://doi.org/10.1007/BF00290976>
- Brown, J. D. (2002). The Cronbach alpha reliability estimate The Cronbach alpha reliability estimate How should we interpret Cronbach alpha? *JALT Testing & Evaluation SIG Newsletter*, 6(1), 17–18.
- Brown, R., & Hewstone, M. (2005). An integrative theory of intergroup contact. *Advances in Experimental Social Psychology*, 37, 255–343. [https://doi.org/10.1016/S0065-2601\(05\)37005-5](https://doi.org/10.1016/S0065-2601(05)37005-5)
- Brünken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science*, 32(1–2), 115–132. <https://doi.org/10.1023/b:truc.0000021812.96911.c5>
- Caporael, L. R. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior*, 2(3), 215–234. [https://doi.org/10.1016/0747-5632\(86\)90004-X](https://doi.org/10.1016/0747-5632(86)90004-X)

- Carlston, D. E., & Smith, E. R. (2007). Principles of mental representation. *Social Psychology: Handbook of Basic Principles*.
- Carpinella, C. M., Wyman, A. B., Perez, M. A., & Stroessner, S. J. (2017). The Robotic Social Attributes Scale (RoSAS): Development and Validation. *ACM/IEEE International Conference on Human-Robot Interaction, Part F1271*, 254–262.
<https://doi.org/10.1145/2909824.3020208>
- Chaminade, T., Franklin, D. W., Oztop, E., & Cheng, G. (2005). Motor interference between humans and humanoid robots: Effect of biological and artificial motion. *Proceedings of 2005 4th IEEE International Conference on Development and Learning, 2005*, 96–101.
<https://doi.org/10.1109/DEVLRN.2005.1490951>
- Chaminade, T., Rosset, D., Da Fonseca, D., Nazarian, B., Lutchter, E., Cheng, G., & Deruelle, C. (2012). How do we think machines think? An fMRI study of alleged competition with an artificial intelligence. *Frontiers in Human Neuroscience*, 6(MAY 2012).
<https://doi.org/10.3389/fnhum.2012.00103>
- Chaminade, T., Zecca, M., Blakemore, S. J., Takanishi, A., Frith, C. D., Micera, S., Dario, P., Rizzolatti, G., Gallese, V., & Umiltà, M. A. (2010). Brain response to a humanoid robot in areas implicated in the perception of human emotional gestures. *PLoS ONE*, 5(7).
<https://doi.org/10.1371/journal.pone.0011577>
- Coeckelbergh, M. (2012). Growing moral relations: Critique of moral status ascription. In *Growing Moral Relations: Critique of Moral Status Ascription*.
<https://doi.org/10.1057/9781137025968>
- Cortina, J. M. (1993). What Is Coefficient Alpha? An Examination of Theory and Applications. *Journal of Applied Psychology*, 78(1), 98–104. <https://doi.org/10.1037/0021-9010.78.1.98>

- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. <https://doi.org/10.1007/BF02310555>
- Cunningham, W. A., & Zelazo, P. D. (2007). Attitudes and evaluations: a social cognitive neuroscience perspective. *Trends in Cognitive Sciences*, *11*(3), 97–104. <https://doi.org/10.1016/j.tics.2006.12.005>
- Darling, K. (2017). “Who’s johnny?” Anthropomorphic framing in human-robot: Interaction, integration, and policy. In *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*. <https://doi.org/10.1093/oso/9780190652951.003.0012>
- Darlow, A. L., & Sloman, S. A. (2010). Two systems of reasoning: Architecture and relation to emotion. In *Wiley Interdisciplinary Reviews: Cognitive Science* (Vol. 1, Issue 3, pp. 382–392). <https://doi.org/10.1002/wcs.34>
- De Houwer, J., Teige-Mocigemba, S., Spruyt, A., & Moors, A. (2009). Implicit Measures: A Normative Analysis and Review. *Psychological Bulletin*, *135*(3), 347–368. <https://doi.org/10.1037/a0014211>
- de Lange, F. P., Spronk, M., Willems, R. M., Toni, I., & Bekkering, H. (2008). Complementary Systems for Understanding Action Intentions. *Current Biology*, *18*(6), 454–457. <https://doi.org/10.1016/j.cub.2008.02.057>
- Dennett, D. (2009). Intentional Systems Theory. In *The Oxford Handbook of Philosophy of Mind* (Vol. 68, Issue 4, pp. 87–106). <https://doi.org/10.1093/oxfordhb/9780199262618.003.0020>
- Dennett, (1971). Intentional Systems. *Journal of Philosophy*, *68*(4), 87–106. <https://doi.org/10.2307/2025382>

- DiSalvo, C. F., Gemperle, F., Forlizzi, J., & Kiesler, S. (2002). All robots are not created equal: The design and perception of humanoid robot heads. *Proceedings of the Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques, DIS*, 321–326. <https://doi.org/10.1145/778712.778756>
- Duffy, B. R. (2003). Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3–4), 177–190. [https://doi.org/10.1016/S0921-8890\(02\)00374-3](https://doi.org/10.1016/S0921-8890(02)00374-3)
- Eddy, T. J., Gallup, G. G., & Povinelli, D. J. (1993). Attribution of Cognitive States to Animals: Anthropomorphism in Comparative Perspective. *Journal of Social Issues*, 49(1), 87–101. <https://doi.org/10.1111/j.1540-4560.1993.tb00910.x>
- Eisenberger, N. I., & Cole, S. W. (2012). Social neuroscience and health: Neurophysiological mechanisms linking social ties with physical health. In *Nature Neuroscience* (Vol. 15, Issue 5, pp. 669–674). <https://doi.org/10.1038/nn.3086>
- Engle, R. W. (2002). Working memory capacity as executive attention. *Current Directions in Psychological Science*, 11(1), 19–23. <https://doi.org/10.1111/1467-8721.00160>
- Engström, J., Markkula, G., Victor, T., & Merat, N. (2017). Effects of Cognitive Load on Driving Performance: The Cognitive Control Hypothesis. *Human Factors*, 59(5), 734–764. <https://doi.org/10.1177/0018720817690639>
- Epley, N., Waytz, A., Akalis, S., & Cacioppo, J. T. (2008). When we need a human: Motivational determinants of anthropomorphism. *Social Cognition*, 26(2), 143–155. <https://doi.org/10.1521/soco.2008.26.2.143>
- Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On Seeing Human: A Three-Factor Theory of Anthropomorphism. *Psychological Review*, 114(4), 864–886. <https://doi.org/10.1037/0033-295X.114.4.864>

- Evans, J. S. B. T. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annual Review of Psychology*, 59(1), 255–278. <https://doi.org/10.1146/annurev.psych.59.103006.093629>
- Evans, J. S. B. T., & Stanovich, K. E. (2013). Dual-Process Theories of Higher Cognition: Advancing the Debate. *Perspectives on Psychological Science*, 8(3), 223–241. <https://doi.org/10.1177/1745691612460685>
- Eyssel, F., & Kuchenbrandt, D. (2012). Social categorization of social robots: Anthropomorphism as a function of robot group membership. *British Journal of Social Psychology*, 51(4), 724–731. <https://doi.org/10.1111/j.2044-8309.2011.02082.x>
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Friese, M., Hofmann, W., & Schmitt, M. (2008). When and why do implicit measures predict behaviour? Empirical evidence for the moderating role of opportunity, motivation, and process reliance. *European Review of Social Psychology*, 19(1), 285–338. <https://doi.org/10.1080/10463280802556958>
- Fritz, M. S., Taylor, A. B., & MacKinnon, D. P. (2012). Explanation of Two Anomalous Results in Statistical Mediation Analysis. *Multivariate Behavioral Research*, 47(1), 61–87. <https://doi.org/10.1080/00273171.2012.640596>
- Gazzola, V., Rizzolatti, G., Wicker, B., & Keysers, C. (2007). The anthropomorphic brain: The mirror neuron system responds to human and robotic actions. *NeuroImage*, 35(4), 1674–1684. <https://doi.org/10.1016/j.neuroimage.2007.02.003>

- Geary, D. C., Widaman, K. F., & Little, T. D. (1986). Cognitive addition and multiplication: Evidence for a single memory network. *Memory & Cognition*, *14*(6), 478–487. <https://doi.org/10.3758/BF03202519>
- Gilbert, D. T., & Hixon, J. G. (1991). The Trouble of Thinking: Activation and Application of Stereotypic Beliefs. *Journal of Personality and Social Psychology*, *60*(4), 509–517. <https://doi.org/10.1037/0022-3514.60.4.509>
- Goodson, J. L. (2005). The vertebrate social behavior network: Evolutionary themes and variations. *Hormones and Behavior*, *48*(1 SPEC. ISS.), 11–22. <https://doi.org/10.1016/j.yhbeh.2005.02.003>
- Gunkel, D. J. (2018). The other question: can and should robots have rights? *Ethics and Information Technology*, *20*(2), 87–99. <https://doi.org/10.1007/s10676-017-9442-4>
- Gunkel, D. J. (2019). The Machine Question. In *The Machine Question*. <https://doi.org/10.7551/mitpress/8975.001.0001>
- Hansen, R. J., Hinze, H., & Holzer, H. (1976). Assay of phosphoenolpyruvate carboxykinase in crude yeast extracts. In *Analytical Biochemistry* (Vol. 74, Issue 2, pp. 576–584). [https://doi.org/10.1016/0003-2697\(76\)90240-2](https://doi.org/10.1016/0003-2697(76)90240-2)
- Haslam, N. (2006). Dehumanization: An integrative review. *Personality and Social Psychology Review*, *10*(3), 252–264. https://doi.org/10.1207/s15327957pspr1003_4
- Hayes, A. F., & Rockwood, N. J. (2017). Regression-based statistical mediation and moderation analysis in clinical research: Observations, recommendations, and implementation. *Behaviour Research and Therapy*, *98*, 39–57. <https://doi.org/10.1016/j.brat.2016.11.001>

- Hayes, A. F., & Scharkow, M. (2013). The Relative Trustworthiness of Inferential Tests of the Indirect Effect in Statistical Mediation Analysis: Does Method Really Matter? *Psychological Science*, 24(10), 1918–1927. <https://doi.org/10.1177/0956797613480187>
- Heider, F., & Simmel, M. (1944). An Experimental Study of Apparent Behavior. *The Japanese Journal of Psychology*, 57(2), 243. <https://doi.org/10.2307/1416950>
- Heider, F., & Simmel, M. (1950). An Experimental Study of Apparent Behavior. *The Japanese Journal of Psychology*, 20(2), 67–74. https://doi.org/10.4992/jjpsy.20.2_67
- Heil, J., & Heil, J. (2019). The Intentional Stance. In *Philosophy of Mind* (pp. 151–165). <https://doi.org/10.4324/9780429506994-9>
- Hermans, D., Spruyt, A., De Houwer, J., & Eelen, P. (2003). Affective priming with subliminally presented pictures. *Canadian Journal of Experimental Psychology*, 57(2), 97–114. <https://doi.org/10.1037/h0087416>
- Huang, L., & Awh, E. (2018). Chunking in working memory via content-free labels. *Scientific Reports*, 8(1). <https://doi.org/10.1038/s41598-017-18157-5>
- Humphrey, N. (1976). The social function of intellect. In *Growing Points in Ethology* (pp. 303–317). Cambridge University Press.
- Iacoboni, M., Lieberman, M. D., Knowlton, B. J., Molnar-Szakacs, I., Moritz, M., Throop, C. J., & Fiske, A. P. (2004). Watching social interactions produces dorsomedial prefrontal and medial parietal BOLD fMRI signal increases compared to a resting baseline. *NeuroImage*, 21(3), 1167–1173. <https://doi.org/10.1016/j.neuroimage.2003.11.013>

- Iida, S., Nakao, T., & Ohira, H. (2012). Prior cognitive activity implicitly modulates subsequent emotional responses to subliminally presented emotional stimuli. *Cognitive, Affective and Behavioral Neuroscience*, *12*(2), 337–345. <https://doi.org/10.3758/s13415-012-0084-z>
- Jack, A. I., Dawson, A. J., Begany, K. L., Leckie, R. L., Barry, K. P., Ciccio, A. H., & Snyder, A. Z. (2013). fMRI reveals reciprocal inhibition between social and physical cognitive domains. *NeuroImage*, *66*, 385–401. <https://doi.org/10.1016/j.neuroimage.2012.10.061>
- Jack, A. I., Dawson, A. J., & Norr, M. E. (2013). Seeing human: Distinct and overlapping neural signatures associated with two forms of dehumanization. *NeuroImage*, *79*, 313–328. <https://doi.org/10.1016/j.neuroimage.2013.04.109>
- Kelemen, D., & Rosset, E. (2009). The Human Function Compunction: Teleological explanation in adults. *Cognition*, *111*(1), 138–143. <https://doi.org/10.1016/j.cognition.2009.01.001>
- Kuchenbrandt, D., Eyssel, F., Bobinger, S., & Neufeld, M. (2013). When a Robot's Group Membership Matters: Anthropomorphization of Robots as a Function of Social Categorization. *International Journal of Social Robotics*, *5*(3), 409–417. <https://doi.org/10.1007/s12369-013-0197-8>
- Kurdi, B., Seitchik, A. E., Axt, J. R., Carroll, T. J., Karapetyan, A., Kaushik, N., Tomzsko, D., Greenwald, A. G., & Banaji, M. R. (2019). Relationship between the implicit association test and intergroup behavior: A meta-analysis. *American Psychologist*, *74*(5), 569–586. <https://doi.org/10.1037/amp0000364>
- Lamberts, J., Van Den Broek, P. L. C., Bener, L., Van Egmond, J., Dirksen, R., & Coenen, A. M. L. (2000). Correlation dimension of the human electroencephalogram corresponds with cognitive load. *Neuropsychobiology*, *41*(3), 149–153. <https://doi.org/10.1159/000026647>

- Langer, E. J. (1992). Matters of mind: Mindfulness/mindlessness in perspective. *Consciousness and Cognition*, *1*(3), 289–305. [https://doi.org/10.1016/1053-8100\(92\)90066-J](https://doi.org/10.1016/1053-8100(92)90066-J)
- Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science*, *19*(3), 143–148. <https://doi.org/10.1177/0963721410370295>
- Leppink, J., van Gog, T., Paas, F., & Sweller, J. (2015). Cognitive load theory: Researching and planning teaching to maximise learning. *Researching Medical Education*, 207–218. <https://doi.org/10.1002/9781118838983.ch18>
- Levy, D. (2009). The ethical treatment of artificially conscious robots. *International Journal of Social Robotics*, *1*(3), 209–216. <https://doi.org/10.1007/s12369-009-0022-6>
- Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, *46*(3), 551–556. <https://doi.org/10.1016/j.jesp.2009.12.019>
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, *39*(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4
- MacLeod, C. M. (1992). The Stroop task: The “gold standard” of attentional measures. *Journal of Experimental Psychology: General*, *121*(1), 12–14. <https://doi.org/10.1037/0096-3445.121.1.12>
- Mars, R. B., Neubert, F. X., Noonan, M. A. P., Sallet, J., Toni, I., & Rushworth, M. F. S. (2012). On the relationship between the “default mode network” and the “social brain.” *Frontiers in Human Neuroscience*, *6*(JUNE 2012), 1–9. <https://doi.org/10.3389/fnhum.2012.00189>

- Marsh, L. E., Mullett, T. L., Ropar, D., & Hamilton, A. F. de C. (2014). Responses to irrational actions in action observation and mentalising networks of the human brain. *NeuroImage*, *103*, 81–90. <https://doi.org/10.1016/j.neuroimage.2014.09.020>
- Meisner, B. A. (2012). A meta-analysis of positive and negative age stereotype priming effects on behavior among older adults. *Journals of Gerontology - Series B Psychological Sciences and Social Sciences*, *67 B(1)*, 13–17. <https://doi.org/10.1093/geronb/gbr062>
- Meyer, M. L., Spunt, R. P., Berkman, E. T., Taylor, S. E., & Lieberman, M. D. (2012). Evidence for social working memory from a parametric functional MRI study. *Proceedings of the National Academy of Sciences of the United States of America*, *109(6)*, 1883–1888. <https://doi.org/10.1073/pnas.1121077109>
- Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, *63(2)*, 81–97. <https://doi.org/10.1037/H0043158>
- Moors, A., & De Houwer, J. (2006). Automaticity: A theoretical and conceptual analysis. *Psychological Bulletin*. <https://doi.org/10.1037/0033-2909.132.2.297>
- Moscovitch, M. (1994). Cognitive Resources and Dual-Task Interference Effects at Retrieval in Normal People: The Role of the Frontal Lobes and Medial Temporal Cortex. *Neuropsychology*, *8(4)*, 524–534. <https://doi.org/10.1037/0894-4105.8.4.524>
- Nass, C., & Moon, Y. (2000). Machines and mindlessness: Social responses to computers. *Journal of Social Issues*, *56(1)*, 81–103. <https://doi.org/10.1111/0022-4537.00153>
- Neely, E. L. (2014). Machines and the moral community. *Philosophy and Technology*, *27(1)*, 97–111. <https://doi.org/10.1007/s13347-013-0114-y>

- Nyangoma, E. N., Olson, C. K., Painter, J. A., Posey, D. L., Stauffer, W. M., Naughton, M., Zhou, W., Kamb, M., & Benoit, S. R. (2017). Syphilis Among U.S.-Bound Refugees, 2009–2013. *Journal of Immigrant and Minority Health*, *19*(4), 835–842. <https://doi.org/10.1007/s10903-016-0397-z>
- O'Rourke, J. (1993). Consciousness explained. In *Artificial Intelligence* (Vol. 60, Issue 2). Penguin uk. [https://doi.org/10.1016/0004-3702\(93\)90006-W](https://doi.org/10.1016/0004-3702(93)90006-W)
- Paas, F. G., Van Merriënboer, J. J., & Adam, J. J. (1994). Measurement of cognitive load in instructional research. *Perceptual and Motor Skills*, *79*(1 Pt 2), 419–430. <https://doi.org/10.2466/pms.1994.79.1.419>
- Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, and Computers*, *36*(4), 717–731. <https://doi.org/10.3758/BF03206553>
- Rauchbauer, B., Nazarian, B., Bourhis, M., Ochs, M., Prévot, L., & Chaminade, T. (2019). Brain activity during reciprocal social interaction investigated using conversational robots as control condition. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1771). <https://doi.org/10.1098/rstb.2018.0033>
- Ridderinkhof, K. R., Ullsperger, M., Crone, E. A., & Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. In *Science* (Vol. 306, Issue 5695, pp. 443–447). <https://doi.org/10.1126/science.1100301>
- Ridderinkhof, K. R., Van Den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. In

Brain and Cognition (Vol. 56, Issue 2 SPEC. ISS., pp. 129–140).
<https://doi.org/10.1016/j.bandc.2004.09.016>

Riether, N., Hegel, F., Wrede, B., & Horstmann, G. (2012). Social facilitation with social robots? *HRI'12 - Proceedings of the 7th Annual ACM/IEEE International Conference on Human-Robot Interaction*, 41–47. <https://doi.org/10.1145/2157689.2157697>

Rosset, E. (2008). It's no accident: Our bias for intentional explanations. *Cognition*, *108*(3), 771–780. <https://doi.org/10.1016/j.cognition.2008.07.001>

Salem, M., Eyssel, F., Rohlfling, K., Kopp, S., & Joublin, F. (2013). To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics*, *5*(3), 313–323. <https://doi.org/10.1007/s12369-013-0196-9>

Saris, I. M. J., Penninx, B. W. J. H., Dinga, R., van Tol, M. J., Veltman, D. J., van der Wee, N. J. A., & Aghajani, M. (2020). Default Mode Network Connectivity and Social Dysfunction in Major Depressive Disorder. *Scientific Reports*, *10*(1), 1–11. <https://doi.org/10.1038/s41598-019-57033-2>

Schilbach, L., Eickhoff, S. B., Rotarska-Jagiela, A., Fink, G. R., & Vogeley, K. (2008). Minds at rest? Social cognition as the default mode of cognizing and its putative relationship to the “default system” of the brain. *Consciousness and Cognition*, *17*(2), 457–467. <https://doi.org/10.1016/j.concog.2008.03.013>

Sharma, D., Booth, R., Brown, R., & Huguet, P. (2010). Exploring the temporal dynamics of social facilitation in the stroop task. *Psychonomic Bulletin and Review*, *17*(1), 52–58. <https://doi.org/10.3758/PBR.17.1.52>

- Sherman, J. W., & Frost, L. A. (2000). On the encoding of stereotype-relevant information under cognitive load. *Personality and Social Psychology Bulletin*, 26(1), 26–34. <https://doi.org/10.1177/0146167200261003>
- Sherman, J. W., Lee, A. Y., Bessenoff, G. R., & Frost, L. A. (1998). Stereotype Efficiency Reconsidered: Encoding Flexibility under Cognitive Load. *Journal of Personality and Social Psychology*, 75(3), 589–606. <https://doi.org/10.1037/0022-3514.75.3.589>
- Shrout, P. E., & Bolger, N. (2002). Mediation in experimental and nonexperimental studies: New procedures and recommendations. *Psychological Methods*, 7(4), 422–445. <https://doi.org/10.1037/1082-989X.7.4.422>
- Smith, E. R., & DeCoster, J. (2000). Dual-process models in social and cognitive psychology: Conceptual integration and links to underlying memory systems. *Personality and Social Psychology Review*, 4(2), 108–131. https://doi.org/10.1207/S15327957PSPR0402_01
- Snyder, R. G. (1961). Vibrational spectra of crystalline n-paraffins. II. Intermolecular effects. In *Journal of Molecular Spectroscopy* (Vol. 7, Issues 1–6, pp. 116–144). [https://doi.org/10.1016/0022-2852\(61\)90347-2](https://doi.org/10.1016/0022-2852(61)90347-2)
- Spatola, N., Belletier, C., Chausse, P., Augustinova, M., Normand, A., Barra, V., Ferrand, L., & Huguet, P. (2019a). Improved Cognitive Control in Presence of Anthropomorphized Robots. *International Journal of Social Robotics*, 11(3), 463–476. <https://doi.org/10.1007/s12369-018-00511-w>
- Spatola, N., Belletier, C., Chausse, P., Augustinova, M., Normand, A., Barra, V., Ferrand, L., & Huguet, P. (2019b). Improved Cognitive Control in Presence of Anthropomorphized Robots. *International Journal of Social Robotics*, 11(3), 463–476. <https://doi.org/10.1007/S12369-018-00511-W/FIGURES/4>

- Spatola, N., Belletier, C., Normand, A., Chausse, P., Monceau, S., Augustinova, M., Barra, V., Huguet, P., & Ferrand, L. (2018a). Not as bad as it seems: When the presence of a threatening humanoid robot improves human performance. *Science Robotics*, 3(21), eaat5843. <https://doi.org/10.1126/scirobotics.aat5843>
- Spatola, N., Belletier, C., Normand, A., Chausse, P., Monceau, S., Augustinova, M., Barra, V., Huguet, P., & Ferrand, L. (2018b). Not as bad as it seems: When the presence of a threatening humanoid robot improves human performance. *Sci. Robot*, 3, 5843. <https://doi.org/10.1126/scirobotics.aat5843>
- Spatola, N., & Chaminade, T. (in press). Precuneus Brain Response Changes Differently During Human-Robot and Human-Human Dyadic Social Interaction. *Scientific Report*.
- Spatola, N., Kühnlenz, B., & Cheng, G. (2020). Perception and evaluation in human-robot interaction: The Human-Robot Interaction Evaluation Scale (HRIES) – a multicomponent approach of anthropomorphism. *International Journal of Social Robotics*. <https://doi.org/10.1007/s12369-020-00667-4>
- Spatola, N., Monceau, S., & Ferrand, L. (2020). Cognitive Impact of Social Robots: How Anthropomorphism Boosts Performances. *IEEE Robotics and Automation Magazine*, 27(3), 73–83. <https://doi.org/10.1109/MRA.2019.2928823>
- Spatola, N., & Normand, A. (2021). Human vs. machine: the psychological and behavioral consequences of being compared to an outperforming artificial agent. *Psychological Research*, 85(3), 915–925. <https://doi.org/10.1007/S00426-020-01317-0/FIGURES/3>
- Spatola, N., & Urbanska, K. (2019). God-like robots: the semantic overlap between representation of divine and artificial entities. *AI and Society*, 35, 329–341. <https://doi.org/10.1007/s00146-019-00902-1>

- Spatola, N., & Wudarczyk, O. (2020). Implicit Attitudes Towards Robots Predict Explicit Attitudes, Semantic Distance Between Robots and Humans, Anthropomorphism, and Prosocial Behavior: From Attitudes to Human–Robot Interaction. *International Journal of Social Robotics*, 1–11. <https://doi.org/10.1007/s12369-020-00701-5>
- Spatola, N., & Wudarczyk, O. A. (2021). Ascribing emotions to robots: Explicit and implicit attribution of emotions and perceived robot anthropomorphism. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2021.106934>
- Sperduti, M., Guionnet, S., Fossati, P., & Nadel, J. (2014). Mirror Neuron System and Mentalizing System connect during online social interaction. *Cognitive Processing*, 15(3), 307–316. <https://doi.org/10.1007/s10339-014-0600-x>
- Spunt, R. P., & Lieberman, M. D. (2013). The Busy Social Brain: Evidence for Automaticity and Control in the Neural Systems Supporting Social Cognition and Action Understanding. *Psychological Science*, 24(1), 80–86. <https://doi.org/10.1177/0956797612450884>
- Spunt, R. P., Meyer, M. L., & Lieberman, M. D. (2015). The default mode of human brain function primes the intentional stance. *Journal of Cognitive Neuroscience*, 27(6), 1116–1124. https://doi.org/10.1162/jocn_a_00785
- Steindl, C., Jonas, E., Sittenthaler, S., Traut-Mattausch, E., & Greenberg, J. (2015). Understanding psychological reactance: New developments and findings. In *Zeitschrift für Psychologie / Journal of Psychology* (Vol. 223, Issue 4, pp. 205–214). <https://doi.org/10.1027/2151-2604/a000222>
- Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology*, 18(6), 643–662. <https://doi.org/10.1037/h0054651>

- Sundar, S. S., Waddell, T. F., & Jung, E. H. (2016). The Hollywood robot syndrome: Media effects on older adults' attitudes toward robots and adoption intentions. *ACM/IEEE International Conference on Human-Robot Interaction, 2016-April*, 343–350. <https://doi.org/10.1109/HRI.2016.7451771>
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312. [https://doi.org/10.1016/0959-4752\(94\)90003-5](https://doi.org/10.1016/0959-4752(94)90003-5)
- Syrdal, D. S., Dautenhahn, K., Koay, K. L., & Walters, M. L. (2009). The Negative Attitudes Towards Robots Scale and reactions to robot behaviour in a live Human-Robot Interaction study. *Adaptive and Emergent Behaviour and Complex Systems - Proceedings of the 23rd Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour, AISB 2009*, 109–115. <https://doi.org/10.1.1.159.9791>
- Tavares, P., Lawrence, A. D., & Barnard, P. J. (2008). Paying attention to social meaning: An fMRI study. *Cerebral Cortex, 18*(8), 1876–1885. <https://doi.org/10.1093/cercor/bhm212>
- Urquiza-Haas, E. G., & Kotrschal, K. (2015). The mind behind anthropomorphic thinking: Attribution of mental states to other species. *Animal Behaviour, 109*, 167–176. <https://doi.org/10.1016/j.anbehav.2015.08.011>
- Valtorta, N. K., Kanaan, M., Gilbody, S., & Hanratty, B. (2018). Loneliness, social isolation and risk of cardiovascular disease in the English Longitudinal Study of Ageing. *European Journal of Preventive Cardiology, 25*(13), 1387–1396. <https://doi.org/10.1177/2047487318792696>
- Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. In *Human Brain Mapping* (Vol. 30, Issue 3, pp. 829–858). <https://doi.org/10.1002/hbm.20547>

- Van Overwalle, F., & Baetens, K. (2009). Understanding others' actions and goals by mirror and mentalizing systems: A meta-analysis. In *NeuroImage* (Vol. 48, Issue 3, pp. 564–584). <https://doi.org/10.1016/j.neuroimage.2009.06.009>
- Waytz, A., Cacioppo, J., & Epley, N. (2010). Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science*, 5(3), 219–232. <https://doi.org/10.1177/1745691610369336>
- Wegner, D. M., & Bargh, J. a. (1998). Control and automaticity in social life. In D. Gilbert, S. T. Fiske, & G. Lindzey (Eds.), *Handbook of social psychology* (pp. 446–496). McGraw-Hill.
- Wegner, D. M., & Schneider, D. J. (2003). The White Bear Story. *Psychological Inquiry*, 14(3–4), 326–329. https://doi.org/10.1207/s15327965pli1403&4_24
- Wehling, T. O., Lichtenstein, A. I., & Katsnelson, M. I. (2008). First-principles studies of water adsorption on graphene: The role of the substrate. *Applied Physics Letters*, 93(20), 196–210. <https://doi.org/10.1063/1.3033202>
- Wheatley, T., Milleville, S. C., & Martin, A. (2007). Understanding animate agents: Distinct roles for the social network and mirror system: Research report. *Psychological Science*, 18(6), 469–474. <https://doi.org/10.1111/j.1467-9280.2007.01923.x>
- Wynne, C. D. L. (2006). What are Animals? Why Anthropomorphism is Still Not a Scientific Approach to Behavior. *Comparative Cognition & Behavior Reviews*, 2. <https://doi.org/10.3819/ccbr.2008.20008>
- Yzerbyt, V., Muller, D., Batailler, C., & Judd, C. M. (2018). New recommendations for testing indirect effects in mediational models: The need to report and test component paths.

Journal of Personality and Social Psychology, 115(6), 929–943.

<https://doi.org/10.1037/pspa0000132>

Złotowski, J., Proudfoot, D., Yogeeswaran, K., & Bartneck, C. (2015). Anthropomorphism: Opportunities and Challenges in Human–Robot Interaction. *International Journal of Social Robotics*, 7(3), 347–360. <https://doi.org/10.1007/s12369-014-0267-6>