Are humanoid robots perceived as mindless mannequins?

Emmanuele Tidoni [a,b,*], Emily S. Cross [c,d,**], Richard Ramsey [e,f], Michele Scandola [g]

[a] School of Psychology, University of Leeds, Leeds, United Kingdom
[b] School of Psychology and Social Work, University of Hull, United Kingdom
[c] Institute of Neuroscience and Psychology, University of Glasgow, Glasgow, Scotland, United Kingdom
[d] Chair for Social Brain Sciences, Department of Humanities, Social and Political Sciences, ETHZ, Zürich, Switzerland
[e] Department of Health Sciences and Technology, ETHZ, Zürich, Switzerland
[f] Department of Humanities, Social and Political Sciences, ETHZ, Zürich, Switzerland
[g] NPSY.Lab.VR & BASIC_NPSY, Department of Human Sciences, University of Verona, Italy

ARTICLE INFO

ABSTRACT

The shape and texture of humans and humanoid robots provide perceptual information that help us to appropriately categorise these stimuli. However, it remains unclear which features and attributes are driving the assignment into human and non-human categories. To explore this issue, we ran a series of five preregistered experiments wherein we presented stimuli that varied in their appearance (i.e., humans, humanoid robots, non-human primates, mannequins, hammers, musical instruments) and asked participants to complete a match-to-category task (Experiments 1-2-3), a priming task (Experiment 4), or to rate each category along four dimensions (i.e., similarity, liveliness, body association, action association; Experiment 5). Results indicate that categorising human bodies and humanoid robots requires the integration of both the analyses of their physical shape and visual texture (i.e., to identify a humanoid robot we cannot only rely on its visual shape). Further, our behavioural findings suggest that human bodies may be represented as a special living category separate from non-human animal entities (i.e., primates). Moreover, results also suggest that categorising humans and humanoid robots may rely on a network of information typically associated to human being and inanimate objects respectively (e.g., humans can play musical instruments and have a mind while robots do not play musical instruments and do have not a human mind). Overall, the paradigms introduced here offer new avenues through which to study the perception of human and artificial agents, and how experiences with humanoid robots may change the perception of humanness along a robot—human continuum.

1. Introduction

The cognitive and neural processes that support the classification of a perceptual stimulus are fundamental for survival and widely studied across animal species (Grill-Spector & Weiner, 2014; Groen et al., 2022; Pitcher & Ungerleider, 2021). More recently, investigation of human's perceptual abilities to appropriately recognise and categorise humanoid robots has become critically important for the future of human—robot interactions and for understanding the mechanisms supporting the perception of living and non-living entities (c.f., Hortensius & Cross, 2018; Cross & Ramsey, 2021).

Current research agrees that processing the sensory (e.g., visual shape and visual textures) and functional features (e.g., graspability) of

a visual stimulus help to identify the superordinate category it belongs to (e.g., a human, a non-human animal, or a human-like object; Caramazza & Shelton, 1998; Proklova et al., 2016; Bracci et al., 2017; Bracci et al., 2019; Proklova & Goodale, 2022). For example, when we look at a mug with the shape of a cow but the typical visual texture of an inanimate object (i.e., ceramic, non-animal skin texture), we do not think that we are looking at a living cow possessing the functions of a mug. On the contrary, the visual textures of a cow mug suggest that it belongs to a non-living category and its functional shape (e.g., a handle and a concave bottom) indicates it may possess the qualities of mug (Bracci et al., 2019). Similarly, when observing a human body and a humanoid robot, gestalt processes responsible to gain a holistic perception (Wagemans et al., 2012a, 2012b) of their global shape allows to infer the

---

same internal structure (i.e., shape skeleton; Lowet et al., 2018), and the integration of other features like their textures or local properties (e.g., curvature, contours), may also guide accurate person or robot identification processes. Indeed, recent studies suggest that during object perception, common features that are shared between different objects (e.g., appearance, shape, function) may be processed using common cognitive, perceptual and neural systems, before being integrated into a combined representation that enables accurate classification (Ayzenberg & Behrmann, 2022; Bracci & Op de Beeck, 2023).

In this sense, Contini and colleagues (2020) have suggested that robot stimuli may be represented as hybrid entities (e.g., human-like non-living entities) through a similar process of feature integration. This may not be surprising as a non-living entity resembling the appearance of a living one (e.g., a cow mug) has been found to engage brain areas responsible for processing the corresponding living entity (e.g., a cow; Bracci et al., 2019). Likewise, the observation of tools (e.g., a tennis racket, a hammer, or a fork that can be used to reach distant objects) recruits brain regions typically active during the observation of human body parts (Bracci and Peelen, 2013; see also Pobric et al., 2010 for a role of parietal and anterior temporal brain areas for tool naming and Ralph et al. (2017) for a distributed account of semantic processing). Moreover, while nonhuman mammals presented with a bipedal presentation or nonhuman animals wearing clothes may be perceived holistically (Welsh et al. 2014, 2023), accumulating evidence also demonstrates the involvement of object-specific brain regions when people perceive a variety of models of social robots (Henschel et al., 2020). In this sense, Sacino et al. (2022) suggested that humanoid robots may be processed, at least partly, like human bodies. The authors used a body-inversion paradigm and showed a reduced ability to match the posture of inverted humanoid robots (and inverted mannequins), a result that is reliably reported for inverted human bodies (Reed et al., 2003; Zlotowski & Bartneck, 2013). This may not be surprising as the human likeness of robots may increase people's tendency to anthropomorphize them (Epley et al., 2007).

These findings suggest that while people can accurately categorise humanoid robots and distinguish them from human agents and inanimate objects, common features of an agent or object, such as their form, texture and function, are processed in a common way, in terms of cognitive and perceptual systems. These recent proposals and findings in the domain of object perception suggest that humanoid robots may be processed by neural circuits encoding both living agents, as well as non-living objects (Cross & Ramsey, 2021). As with any emerging field of research, however, many questions remain unanswered in the context of humanoid robot perception. The proposals thus far are very broad and cover a wide range of possibilities, in terms of the cognitive and brain structures that may support the perception and classification of humanoid robots. For instance, although perceiving humanoid robots likely shares mechanisms with the perception of both living agents and non-living objects, the extent to which humanoid robots are processed by mechanisms that are more or less similar to the perception of more specific entities is unclear. For example, are humanoid robots processed more like humans or mannequins, or bipedal animals, such as apes? And to what extent are humanoid robots processed like objects with which they share functional features, such as the ability to grasp and manipulate different musical instruments and tools?

Rather than being a dichotomy, we can conceptualise the neuro-cognitive mechanisms supporting the distinction between living and non-living entities along a continuum between perceptually similar objects also varying in their degree of attributed human psychological traits (e.g., the ability to think and plan: *agency*). Several lines of evidence suggest that when forming a representation of a person, brain circuits for representing another person's physical appearance, their psychological traits, and social stereotypes are engaged (Quadflieg et al., 2011; Greven et al., 2016; Ramsey, 2018). For example, Thorat et al. (2019) showed that posterior brain areas of the ventral-temporal cortex (VTC) may be responsible for analysing the perceptual features of a

stimulus, while more anterior parts may be responsible to process its perceived agency traits. It is worth noting that when making judgments about humans and robots, people tend to consider robots as less capable of thinking and experiencing events like humans do (Gray et al., 2007). In this sense, the perception of robots may not only be influenced by visual processes alone, but also by the human qualities people think they possess (Jack et al., 2013).

Here, given that humanoid robots may share (dis)similar features across a variety of objects, rather than focus at superordinate level of object classification (e.g., living vs non-living), we investigate the similarity in the perceptual processes that cover the classification of a wide range of object classes and features. The benefit is that a similarity space should be uncovered between humanoid robots and wider range of agents and objects, which should yield new insights into the underlying cognitive and perceptual systems that support the perception of humanoid robots. Importantly, given the novelty of the tasks and stimuli proposed in this study, we used headless stimuli to ensure shape and pose of the observed agent were based only from bodily cues (Yovel et al., 2010).

To probe these outstanding questions, the current study used a range of perceptual tasks since people may explicitly categorise objects at three levels of abstraction (Mervis & Rosch, 1981). For example, we may indicate whether two stimuli are perceptually identical (exemplar level: are these two stimuli the same exact person? Are they the same exact tool?), if they reflect two different types within the same category (intermediate level: are these two stimuli both humans? Are they both tools?), or if the stimuli are part of different superordinate more abstract categories (general level: living or non-living). Previous studies suggested that tasks at an intermediate stage may reveal neural (Rogers et al., 2005) and behavioural (Wiggett et al., 2009) differences between living and non-living stimuli. Thus, we developed three match-to-category tasks (Experiments 1–3) and a priming task (Experiments 4) to study the categorisation of a perceptual stimulus at the intermediate level (humans, humanoid robots, apes, mannequins, tools, and non-tool objects).

First, in the match-to-category task participants saw a sample category and then indicated which of two probes belonged to the same category.[1] If perceptual information is shared by two intermediate categories, then we would expect interference (e.g., accuracy drop, longer reaction times) when making a judgment (Experiments 1 and 2). However, when the relevance of low-level information is reduced, interference could also reflect the influence of superordinate information (Experiments 2 and 3). We also reasoned that if any interpretation using the match-to-category task is correct, then we should reach similar conclusions from results obtained using a different task. In this case, we chose a priming task. In the priming task (Experiment 4), participants first observed one stimulus of a category and then indicated whether the target image was a human or robot. As opposed to the match-to-category task, the more two stimuli rely upon the same processes to categorise, the more we should expect a facilitatory effect in a priming task. Finally, in Experiment 5, we asked participants to rate the stimuli along several dimensions (see Methods) to further support our findings.

We expected people to process humanoid robots as mindless objects (Contini et al., 2020; Sacino et al., 2022) like tools or non-tool objects. We also expected humanoid robots to be processed more like mannequins than humans or non-human animals. Findings are discussed with respect to the cognitive and neural mechanisms supporting the perception of humans and humanoid robots, and how we might expect robot perception to change and develop with experience.

---

[1] Note that the match-to-category task used in this study can also be conceptualised as a cued visual search task with only two probes.

2. Materials and Methods

2.1. Transparency and openness

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study. All data, analysis code, and research materials are available at https://osf.io/586mh/. Data were analysed using R 4.0.2 (R Core Team, 2022) and the packages we used are detailed in the "Measures, Data Processing and Statistical Approach" section. All experiments' designs and analyses were pre-registered.[2]

2.2. Participants

Across five separate pre-registered experiments, a total of 368 English-speaking adults (see Table 1) were recruited from the University of Hull (Experiments 1, 2, and 4) and via online platforms (Experiments 3, 4, and 5; Palan and Schitter, 2018; Experiment 3, Call For Participants: https://www.callforparticipants.com) in exchange for partial course credit and monetary compensation, respectively.

Each participant completed a single experiment. Prior sample size calculations (G*Power; Faul et al., 2007) were performed to have sufficient power to detect medium to large effect sizes for Experiments 1 and 2 (dz = 0.60, Alpha = 0.05, Beta = 0.80, minimum sample size of 24). Experiment 3 was performed online, and the sample size was adjusted to detect small effect sizes (dz = 0.25, Alpha = 0.05, Beta = 0.80, minimum sample size of 128). As we observed effect sizes higher than 0.3 in the online Experiment 3, we increased the minimum effect size of interest for Experiment 4 (dz = 0.3, Alpha = 0.05, Beta = 0.80, minimum sample size of 90) and Experiment 5 (dz = 0.35, Alpha = 0.05, Beta = 0.80, minimum sample size of 67). Anticipating that some participants might be removed (e.g., outliers) especially in the online experiments, we slightly increased the target sample size for online data collection to avoid reduced statistical power.

The task, procedure, and methodology were reviewed and approved by the institutional review boards of the University of Hull (Experiments 1–3: protocol number FHS150; Experiment 4: protocol number FHS286; Experiment 5: protocol number FHS404) and carried out in accordance with the standards set by the Declaration of Helsinki. All participants were naïve to the task and purpose of the experiment and informed consent was obtained before starting the task.

Table 1

Descriptive measures of the sample for each experiment. We recruited 120 and 30 participants from 'Prolific' and 'Call For Participants' for Experiment 3, respectively. We recruited 66 and 34 participants from the University of Hull and 'Prolific' for Experiment 4, respectively.

Experiment	Sample size	Age mean ± s.e.m., range [min – max]
1 (Lab)	n = 24 [female = 13, male = 11, prefer not to say = 0]	22.63 ± 1.30, [18–43]
2 (Lab)	n = 24 [female = 12, male = 12, prefer not to say = 0]	20.38 ± 0.55, [18–29]
3 (online)	n = 150 [female = 63, male = 86, prefer not to say = 1]	27.88 ± 0.81, [18–73]
4 (online)	n = 100 [female = 55, male = 39, prefer not to say = 6]	22.71 ± 0.67, [18–50]
5 (online)	n = 70 [female = 33, male = 36, prefer not to say = 1]	27.29 ± 0.90, [18–53]

2.3. Apparatus and Task

Across the five experiments, participants observed stimuli grouped into six categories (Fig. 1): tools (i.e., forks, golf club, hammer, pen, saw, tennis racket; labelled as "tools" during instructions), non-tool-objects (i.e., accordion, saxophone, trumpet, violin, guitar, tambourine; labelled as "musical instruments"), and headless "humans", headless humanoid "robots", headless "monkeys", and headless wooden "mannequins". We used headless stimuli because heads are processed in cortical areas that are separate from bodies (Downing et al., 2001), to reduce the role of the head for pose estimation (Yovel et al., 2010), and that facial features might alter the processing of the body (Welsh et al., 2023). Each category had 20 different stimuli. Since our goal was to compare how different human-like bodies are processed, we presented tools and non-tool-objects as separate categories as they can prime and evoke hand movements equally but differ in their ability to extend the body (Bracci and op de Beeck, 2016). For human bodies, we adopted stimuli used to find body-selective brain areas (Downing et al., 2001). Humanoid robots had two legs, two arms, a recognisable texture (e.g., metal, plastic), and their overall human-likeness ratings ranged between 37 and 66 (ABOT Database, accessed 04/09/2019; Phillips et al., 2018; see Supplementary Table S1 for the list or humanoid robots used).

In Experiments 1–3, participants performed a two-choice match-to-category visual categorisation task, in which they decided which of two pictures matched the category presented previously (Fig. 2a; Moro et al., 2008; Yeh & Peelen, 2022). In Experiments 1 and 2 we used human, humanoid robot, non-tool object, and tool categories. In Experiment 3 we used human, humanoid robot, monkey, and mannequin categories. In each trial, participants observed the sample category for 150ms. After that, a visual mask was shown for 500ms (see below how masks were created), followed by the appearance of two probes: a picture different from the sample but belonging to the same category as the target stimulus, and a picture belonging to another category as a distractor. For example, if the sample category were "humans", after the mask participants may have observed a humanoid robot image as distractor and a different human body as target. The sample category appeared in the centre of the screen and the target-distractor pair appeared along the central vertical axis against a black background. The location (up, down) of the target (and distractor) was randomised each trial. The target-distractor pair remained on screen until keypress. Participants pressed the "k" and "m" with the index and third finger of the right hand to indicate whether the target category was "up" or "down" respectively. There were 12 target-distractor combinations for a total of 240 trials for Experiments 1–2 (20 trials for target-distractor pair) and 192 trials for Experiment 3 (16 trials for target-distractor pair) pseudorandomised across 4 blocks. The intertrial interval randomly ranged between 1,000ms and 1,400ms for Experiments 1–2 and between 900ms and 1,100ms for Experiments 3–4.

In Experiment 4, participants performed a two-choice priming task, in which they decided whether a target image depicted a human or a humanoid robot (Fig. 2b). Participants observed for 300ms a picture of a prime (mask, human, humanoid robot, non-tool-object, tool, monkey, mannequin). After prime presentation, the screen remained blank for 100ms. This was followed by the presentation of the target (i.e., either a picture of a human or humanoid robot body) for 300ms. Then, the screen turned blank and the trial ended after the participant made a keypress. There were 14 prime-target combinations for a total of 280 trials (20 trials for prime-target pair) pseudorandomised across 4 blocks. In those cases where the prime and target categories were identical (prime-human, target-human), the target image differed from the prime image. The prime and target stimuli appeared in the centre of the screen against a black background. Participants pressed the "k" or "l" to indicate which target category they perceived with the index and third finger of the right hand. The association of a key to a human or robot category was randomised across participants. The intertrial interval randomly ranged between 900ms and 1,100ms.
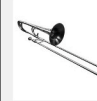
Fig. 1. Examples of the categories across the five experiments. See on the OSF repository for the stimuli used across the experiments before matching contrast and luminance, the overall shapes for each experiment, and the scripts used to generate them. Numbers below each category indicate the experiment(s) that category was presented. We also show a stimulus sample and the overall shape for each category. In Experiment 1 we used several images of tools and musical instruments. In Experiments 2, 4, and 5 we used guitars and hammers only.
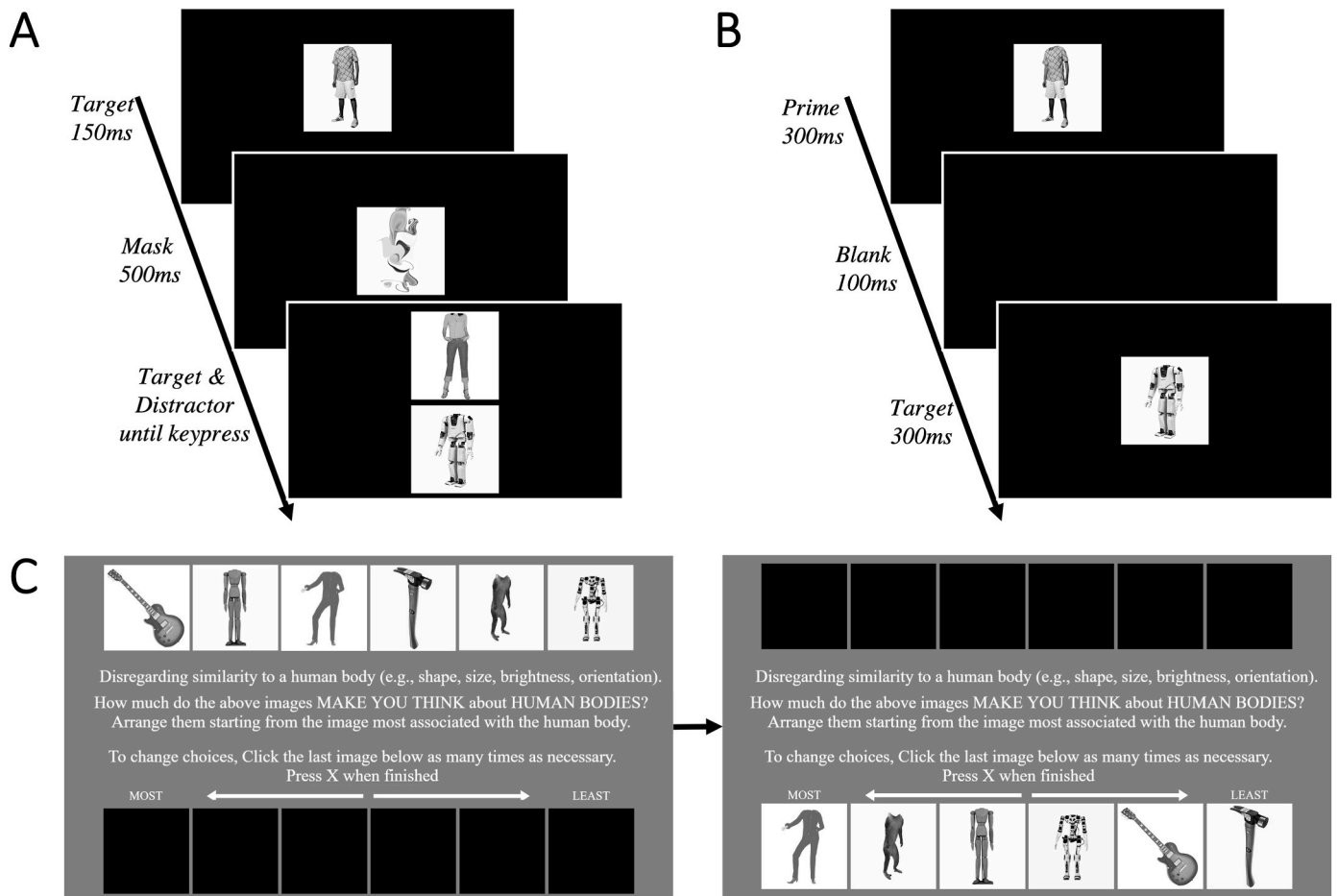
Fig. 2. Trial timeline of the match-to-category task (A) and the priming task (B). In the arrangement task (C), participants saw in the upper part of the screen the six categories (left side) and ranked them based on the instructions presented at the centre of the screen. Once the ranking was complete participants could continue by pressing the 'x' key.

In Experiment 5, we asked participants to observe six images, each belonging to a different category (human, humanoid robot, non-tool-object, tool, monkey, mannequin), and to arrange them based on four different criteria in four separate blocks (Fig. 2c). In particular, we asked participants to disregard the "similarity in low-level image features (e.g., size, brightness, orientation)" and arrange the images based on how much they were "perceptually SIMILAR to a HUMAN SHAPE" (*Perceived Similarity*), how much they belonged "to a LIVING CATEGORY"

(*Perceived Liveliness*), how much they made them "THINK about HUMAN BODIES" (*Body Association*), and how much they made them "THINK about HUMAN ACTIONS" (*Action Association*). The order of the questions was randomised across participants and the categories were top-aligned to the screen with the location of each category randomly selected (e.g., a human category could appear as the left-most image for the first question and the right-most image for the second question). The stimuli for each category were randomised for each question. There was

no time pressure to complete each question and participants pressed the 'x' key when satisfied with their arrangement to start answering the next question.

For each experiment we tried to match the overall number of pixels across all stimuli (see OSF folder "01"). Then, all grayscale stimuli were matched for luminance and contrast using lumMatch function in SHINE toolbox (Willenbockel et al., 2010) independently for each experiment (Experiments 1–5). Masks were created with a custom Matlab script for generating diffeomorphed images (40 distortion steps, distortion level = 160; Stojanoski and Cusack, 2014) from the luminance and contrast matched images independently for each experiment (Experiments 1–4).

In all experiments, we never showed robots with human skin-like textures. We preferred to have clearly distinguishable categories between humans and mechanical humanoid robots. We reasoned that the match-to-category and priming tasks performances may have been negatively affected when humans and humanoid robots with highly realistic textures were presented simultaneously (Experiments 1–3; high RTs and low accuracies given the high resemblance of the two categories and the brief presentation of the sample stimulus; Experiment 4: less ability to discriminate the role of robots as primes given their high resemblance with humans).

Experiments 1–2 were designed using Matlab 2018R and Psychotoolbox 3 (Brainard, 1997; Kleiner et al., 2007). Experiments 3–5 were hosted on Pavlovia (Pavlovia.org) and designed using Psychopy 3 (Peirce et al., 2019).

2.4. Procedure

Participants were invited to read the information sheet and communicate any questions to the experimenter if needed. After providing informed consent, participants were explained (Experiments 1–2) or read (Experiments 3–5) the experimental instructions and performed a quick practice session (20 trials in Experiments 1, 2, and 4; 12 trials in Experiment 3; 1 trial in Experiment 5). The online practice session for Experiment 3 provided feedback to participants for their first 6 practice trials. For Experiment 5, participants familiarised with the arrangement task by ordering from the highest to the lowest a predefined sequence of images depicting numbers (i.e., 4, 5, 6, 2, 1, 3). After the practice session, participants completed 4 experimental blocks. Then, participants rated their exposure to media robotic content ("How often do you watch movies, TV series, or play videogames where robots are involved?") using a nominal scale (1 = Never, 2 = Once every Year, 3 = Once every 6 months, 4 = Once every 3 months, 5 = Once every month, 6 = More than once every month) for Experiments 1–4 only. These ratings entered the statistical model as scaled predictor of non-interest (see next section). Finally, participants were debriefed as to the purpose of the experiment.

In Experiments 1 and 2 only, after the main task and before the debrief session, participants also completed a computerised series of questionnaires related to attitudes and perception of robots. They reported their level of agreement along a horizontal bar to the items of the Negative Attitudes towards Robots Scale (NARS) and the Robot Anxiety Scale (RAS; Nomura et al., 2008). Moreover, participants answered to some items of the anthropomorphism, animacy, and intelligence subscales of the Godspeed questionnaire (Bartneck et al., 2008) by indicating their position on a scale between two bipolar words (e.g., human-like, machine-like). Furthermore, we assessed participants' opinions about how close they perceive themselves and other humans to robots, tools, and objects, using a modified and computerised version of the Inclusion of the Other in the Self (Schubert & Otten, 2002). These questionnaires were collected to have insights about the perception of robots within the sample and were not analysed further (see Supplementary Table S2).

2.5. Measures, Data Processing and Statistical Approach

We collected task Accuracy and Response Times (RTs; expressed in seconds) as performance measures in Experiments 1–4 and each category final rank for all questions in Experiment 5. We specified how data would be processed in the online pre-registration files. Preregistered exclusion criteria slightly varied across Experiments 1–4. Specifically, moving from laboratory to online testing (from Experiment 2 to Experiment 3), we decided to exclude participants with an accuracy <65%. Moreover, for Experiment 4 we also specified that trials deemed too fast (<0.150 s) or too slow (>3.000) would be removed. This was necessary as we noticed from Experiment 3 that excluding trials based only one criterion (trials with a RT exceeding 2.5 standard deviations, SD, from the mean within each block) was not sufficient to prevent performance to be potentially affected by responses that were too fast (e. g., anticipating the answer) or too slow (e.g., being distracted).

In order to facilitate comparison across experiments, we decided to adopt the same data processing approach across Experiments 1–4. So, we plotted the probability distribution of the obtained dataset for each experiment after removing trials with RTs deemed too fast (<0.150 s), too slow (>3.000), and exceeding 2.5SD from the mean RT within each block. Visual inspection of the plots suggested that responses may be influenced by slow RTs (Fig. S1). Hence, the 3.0 s criterion may not have been adequate to exclude trials that would affect statistical analyses. Therefore, after computing the value exceeding 2.5SD from the average of all trials for each experiment (Experiment 1: 1.046 s; Experiment 2: 0.943 s; Experiment 3: 1.301 s; Experiment 4: 1.046 s), we selected 1.5 s as a new cut-off value for Experiments 1–4 (Fig. S1 shows the distribution probability plot for each experiment using 3.0 s and 1.5 s as RTs cut-off criterium).

To sum up, we combined different pre-registered data processing approaches into one to facilitate comparison across experiments. First, we excluded trials <0.150 s and >1.500 s from Experiments 1–4. Then, we excluded trials whose RTs fell above or below 2.5SDs of the overall mean within each block for each participant. At this stage of data processing, we excluded participants whose overall accuracy was below 65%. Finally, we excluded participants whose performance (in RTs or Accuracy) fell above or below 2.5SDs of the overall mean across conditions of the remaining participants.

Statistics for Experiments 1–4 were performed using R 4.0.2 (R Core Team, 2022) run on the University of Hull High-Performance facility VIPER (http://hpc.wordpress.hull.ac.uk/home/). Specifically, we used the lme4 package (v1.1.27.1; Bates et al., 2015) to perform Linear Mixed Models (LMM) with fixed effects and complex random intercepts (CRIs) as scalar random effects (Scandola & Tidoni, 2024). Model reduction started from the full-CRIs LMM (Scandola & Tidoni, 2024) with all main effects and interaction of interests and scaled predictors of non-interest (e.g., trial number; see Supplementary Table S3). If the model overfitted, the CRI with the lowest variance was removed until a convergent non-singular model was found (for the final model see Supplementary Table S3). For LMMs on RT of correct answers, we also report the partial eta-squared as a measure of effect size (effectsize v0.4.5; Ben-Shachar et al., 2020). For all LMMs, we computed the conditional R2 (for lme4lmer performance v0.7.3, Lüdecke et al., 2021; for lme4glmer MuMIn v1.43.17, Barton, 2016). Throughout the paper, we report the p-values computed on the estimates of the simplified LMM. For each multiple comparison, we report the individual Bonferroni corrected p-value computed from the final LMM using emmeans (v1.6.2–1; Lenth, 2019).

We also performed several confirmatory analyses. For Experiment 4, we performed an ANOVA on mean-aggregated Accuracy data and ANOVAs on both mean- and median-aggregated RT data (confirmatory ANOVAs were not possible for Experiments 1–3 as they were not fully crossed designs). To support results from the multiple comparisons conducted using the LMM approach, we ran multiple Bonferroni-corrected pairwise t-test comparisons on mean-aggregated data for

Accuracy and both mean- and median-aggregated for RT data. For each comparison, we computed the absolute value of the Cohen's d (|d|) and the Bayes Factor (BF10; default Cauchy prior of 0.707; JASP Team, 2022; Version 0.14) to further facilitate the reader in assessing the strength of the evidence. Classically, BF10 is interpreted as showing very strong evidence towards the alternative hypothesis when greater than 150, strong evidence when equal or greater than 20, positive evidence when equal or greater than 3, and with weak or negligible evidence when between 1 and 3 (Raftery, 1995). The inverse of these values (1/150, 1/20, 1/3) can be interpreted as BF10 showing very strong, strong, or positive evidence towards the null hypothesis. Importantly, we considered an effect as statistically significant when the p-values obtained from the LMM model and aggregated data were all below 0.05. In case one or all p-values were greater than 0.05 but all lower than 0.10, we considered that finding as a tendency. We considered an effect as statistically non-significant, or non-conclusive, when at least one p-value obtained from either the LMM or the aggregated data was greater than 0.10. If not stated otherwise, the confirmatory analyses supported the results obtained from the LMM model.

Given the considerable number of comparisons across main and supporting analyses, reporting information for all performed contrasts would affect the readability of the manuscript. In the main text, we report the p-values and the averages for the compared conditions (mean ± s.e.m.). In dedicated supplementary tables for each experiment (Tables S4–S8) we report all statistical information (e.g., estimates and confidence intervals of the LMM, Cohen's d and Bayes Factor computed on mean- and median-aggregated data).

The interested reader can find on OSF a dedicated folder with the reanalyses of Experiments 1–4 using the Inverse Efficiency index (interpretation of the results does not change).

Analyses for Experiment 5 were performed using JASP as each category for each question was ranked only once. Given the absence of prior comparisons of interests, we computed Bonferroni-corrected post-hocs from the JASP statistical model.

3. Experiment 1 – is a humanoid robot perceived as a human body (part 1)?

Participants completed the match-to-category task with headless human bodies, headless humanoid robots, tools, and non-tool-objects categories.

One trial was excluded as a participant needed a break within a block. We also removed trials with RTs deemed too fast or too slow (1.20%). Then, trials with RTs falling above or below 2.5SD of the overall mean within each block of each participant were removed (3.15%). No participants' performance was <65% and no participants had a performance above 2.5SD of the overall mean across conditions of the remaining participants (final sample n = 24; average trial per condition in the final dataset [min, max]: 19.14 ± 0.07 s.e.m. [13, 20]).

3.1. Results

We analysed performance measures (see Fig. 3) with Target (human, robot, non-tool-object, object) and Distractor (human, robot, non-tool-object, object) as within-subject fixed effects of a linear mixed model (LMM; see Table S3 in Supplementary Materials for details on the fixed and random-effects structure of all statistical models). In the case of a two-way Target by Distractor interaction, we performed eighteen multiple paired comparisons of interest. Specifically, we compared the effect of each distractor within each target category (12 comparisons: e.g., Target-Human/Distractor-Robot vs Target-Human/Distractor-Object), and when two categories were paired but their target-distractor relationship changed (6 comparisons: e.g., Target-Human/Distractor-Robot vs Target-Robot/Distractor-Human). We also performed confirmatory
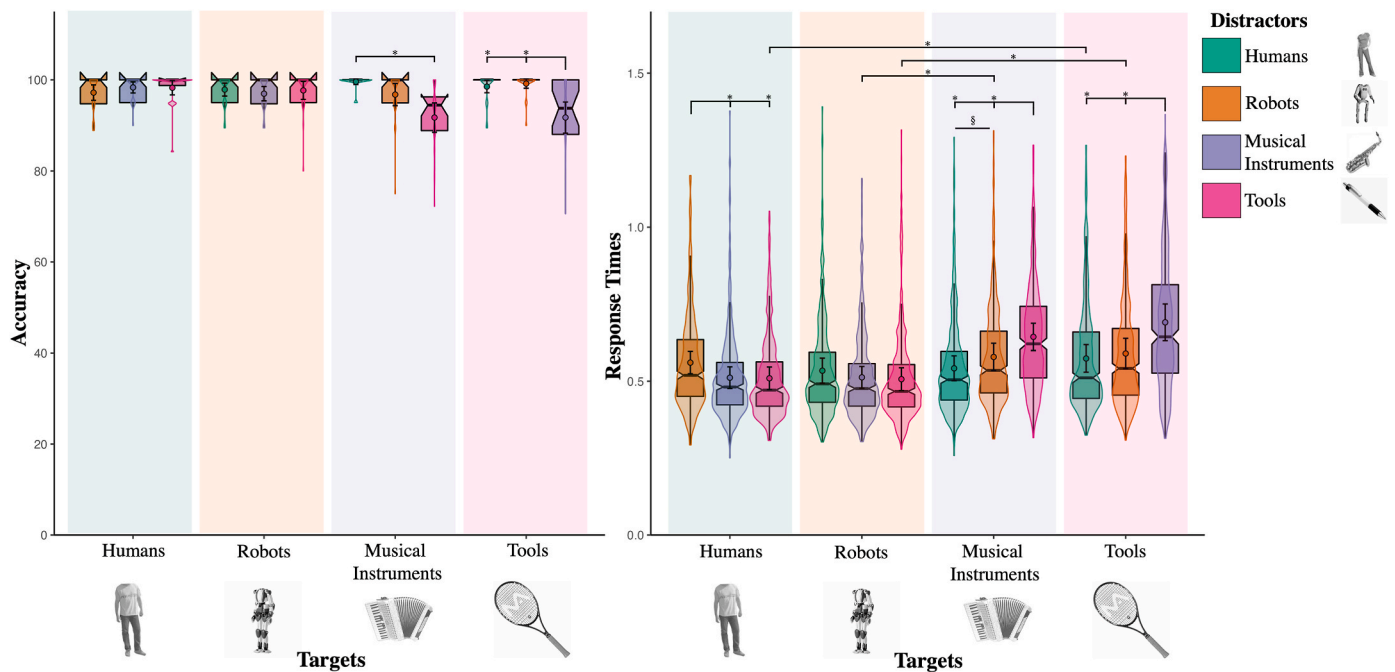
Fig. 3. Participants' performance (Accuracy percentage on the left, Response Times expressed in seconds on the right) in Experiment 1. To provide a comprehensive overview of collected data, raw data from each experimental condition are visualised as median boxplots (with lower and upper hinges corresponding to the 25th and 75th percentile and whiskers extending no further than 1.5 * "Interquartile Range" from the hinge) and violin probability density. The notched boxplot gives a roughly 95% confidence interval for comparing medians. If the notches of two boxes do not overlap, this suggests that the medians are significantly different. The circles inside each median bar plot indicate the average of the by-subject mean-aggregated data for that condition. Error bars represent 95% confidence intervals of the mean based on subject-aggregated data. Data visualisation has been possible by adapting the open-source R code "RainCloudPlots" (Allen et al., 2021). Asterisks denote the significant differences (p < 0.05) for both the LMM and the comparisons on aggregated data as reported in the main text. Section sign symbol (§) denotes a tendency (e.g., one or all p-values from the LMM and confirmatory analyses were greater than 0.05 but all lower than 0.10).

Bonferroni-corrected pairwise *t*-test comparisons on mean-aggregated Accuracy data as well as on mean- and median-aggregated RT data. Non-conclusive findings (i.e., less robust, more fragile) are highlighted whenever the LMM and the confirmatory analyses yield discordant results (see Method section for further details; see Table S4 summarising all comparisons and effect sizes in Experiment 1).

For Accuracy, we observed no main effect of Target or Distractor, $\chi^2(3) < 7.773$, $p > 0.051$, and a significant Target by Distractor interaction, $\chi^2(5) = 20.266$, $p < 0.001$. The latter suggested that participants were less accurate in detecting a non-tool-object when the distractor was a tool ($91.72 \pm 1.55\%$) compared to a human ($99.58 \pm 0.29\%$; $p = 0.001$). Moreover, participants were less accurate in detecting a tool when the distractor was an object ($91.72 \pm 1.64\%$) compared to a human ($98.51 \pm 0.65\%$; $p = 0.002$) or a robot ($99.17 \pm 0.49\%$; $p = 0.001$). No other Bonferroni corrected p-values were lower than 0.05 for multiple comparisons computed on the estimates of the simplified LMM and using pairwise t-tests on aggregated data (see Methods for data analysis approach).

For RTs, we removed incorrect answers (2.94%) from the final dataset. We observed a main effect of Target, $F(3, 132) = 28.693$, $p < 0.001$, $\eta p2 = 0.683$, a main effect of Distractor, $F(3,108) = 6.538$, $p < 0.001$, $\eta p2 = 0.339$, and a significant Target by Distractor interaction, $F(5, 99) = 33.864$, $p < 0.001$, $\eta p2 = 0.637$. Comparisons within each the target category revealed that participants were slower in detecting a human when the distractor was a robot (0.560 ± 0.018 s) compared to an object (0.511 ± 0.017 s; $p = 0.001$) or a tool (0.510 ± 0.018 s; $p < 0.001$). Participants were also slower in detecting an object when the distractor was a tool (0.644 ± 0.021 s) compared to a human (0.542 ± 0.020 s; $p < 0.001$) or a robot (0.579 ± 0.021 s; $p < 0.001$). We also observed a tendency for participants to be slower in detecting an object when the distractor was a robot compared to a human ($p = 0.057$; $p_{mean} = 0.025$; $p_{median} = 0.009$). Moreover, participants were slower in detecting a tool when an object (0.691 ± 0.029 s) compared to a human (0.574 ± 0.022 s; $p < 0.001$) or a robot (0.590 ± 0.024 s; $p < 0.001$) acted as distractors.

Comparisons between conditions where the target-distractor relationship changed revealed that detecting a tool when humans and robots were the distractors took more time than detecting a human or a robot when tools were the distractors ($p < 0.001$). Participants were also slower in detecting an object when robots were the distractor than detecting a robot when objects were the distractor ($p < 0.001$). No other Bonferroni corrected p-values were lower than 0.05 for all multiple comparisons computed on the estimates of the simplified LMM and confirmatory multiple comparisons using pairwise t-tests on aggregated data (see Methods for data analysis approach).

3.2. Discussion experiment 1

We observed lower accuracies and higher RTs when tools and non-tool-objects were presented simultaneously. This may suggest that processing these categories may rely on similar neural and cognitive mechanisms.

We also observed that categorising a human took more time when the distractor was a robot compared to other categories. Rather than thinking that the two categories belong to the same superordinate category (e.g., living entities), results may be explained by the shape similarity between humans and robots. In other words, participants may have mainly relied on the visual shape of the stimulus to correctly classify humans.

We did not observe a similar pattern when participants had to classify a robot with human stimuli as distractors. We found RTs did not differ across the three distractors. This may indirectly indicate that the mechanisms responsible for categorising a robot may benefit from the combined engagement of processes analysing its human-like shape and its object-like visual textures. However, our task was a speeded-choice reaction time task. Thus, it is possible that familiar objects of human

bodies favoured the use of a subset of criteria (humanlike shape), leading to greater reaction times when an object with a similar structure was present. In contrast, when the target was a robot, participants may have paid attention to and integrated additional cues (shape and texture), leading to a reduced time to recognise robots.

Furthermore, we observed a tendency of longer RTs to classify non-tool-objects (i.e., musical instruments) when robots rather than humans acted as distractors. We also noted that accuracies did not differ when participants had to categorise non-tool-objects with tools and robots as distractors. This may further suggest a hybrid recruitment of cognitive and neural resources to process and represent humanoid robots (part human and part objects).

Moreover, comparisons between conditions where non-human-like stimuli and robots were paired (i.e., Target-Object/Distractor-Robot vs Target-Robot/Distractor-Object; Target-Tool/Distractor-Robot vs Target-Robot/Distractor-Tool) revealed some search asymmetries (Treisman & Gormican, 1988). That is, participants were faster or more accurate in searching one category over the other. RTs were slower when participants had to identify the non-human-like categories rather than the robots. Similar asymmetries were observed between tools and human bodies (e.g., Target-Human/Distractor-Object vs Target-Object/Distractor-Human). Although, search asymmetries are interpreted as evidence that the faster (or more accurate) categorisation of a target stimulus indicates that such stimulus carries more salient or deviant information than the distractor (Treisman & Gormican, 1988; Gandolfo & Downing, 2020), a more plausible explanation for the asymmetries in this experiment is that musical instruments and tools greatly differed in their orientations and shapes. In other words, musical instruments and tools varied greatly compared to human and robotic bodies and may have slowed their perceptual categorisation.

To sum up, perceptual processing of human-like robots possessing object-like features may require processes responsible for analysing both living and non-living categories. Moreover, since we found that a robot distractor had a different effect on musical instruments and tool targets, it is unlikely that our results can only be explained by familiarity of robot stimuli and visual texture similarities across robots, musical instruments, and tools. Rather, it is possible to assume that superordinate information (e.g., the non-living nature of robots, the functions of tools) may have modulated the behavioural performance. However, some of the results may have been affected by the lack of visual-form predictability of musical instruments and tools (i.e., greater RTs for recognising non-human-like categories).

For these reasons, we tried to replicate and extend the results by increasing visual shape predictability of non-human-like categories.

4. Experiment 2 – is a humanoid body perceived as a human body (part 2)?

Participants performed the same task detailed in Experiment 1. Stimuli of humans and robots were identical to Experiment 1. However, we reduced the shape variability for non-tool-objects and tools. We expanded the set of guitar and hammer images for the non-tool-objects and too ls categories respectively (see Fig. 1 for the overall shape of the stimuli used in Experiment 2). We selected guitars and hammers as their shapes are recognisable and distinctive: guitars have a wider bottom than hammers, hammers have a wider top than guitars.

One trial was excluded as a participant needed a break within a block. We removed trials with RTs deemed too fast or too slow (1.18%). Then, trials with RTs falling above or below 2.5SD of the overall mean within each block of each participant were removed (2.99%). No participants' performance was <65%. A participant with performance above 2.5SD of the overall mean across conditions of the remaining participants was excluded (final sample n = 23; average trial per condition in the final dataset [min, max]: 19.20 ± 0.06 s.e.m. [15, 20]).

4.1. Results

For Accuracy (see Fig. 4), we observed no main effect of Distractor, $\chi^2(3) = 2.757$, p = 0.431, a main effect of Target, $\chi^2(3) = 8.353$, p = 0.039, and a significant Target by Distractor interaction, $\chi^2(5) = 20.795$, p = 0.001. However, no Bonferroni corrected p-values were lower than 0.05 for multiple comparisons computed on the estimates of the simplified LMM and using pairwise t-tests on aggregated data (see Methods for data analysis approach; see Table S5 summarising all comparisons and effect sizes in Experiment 2).

For RTs, we removed incorrect answers (1.79%) from the final dataset. We observed a main effect of Target, $F(3, 129) = 35.412$, p < 0.001, $\eta p2 = 0.264$, a main effect of Distractor, $F(3, 116) = 12.898$, p < 0.001, $\eta p2 = 0.039$, and a significant Target by Distractor interaction, $F(5, 106) = 31.250$, p < 0.001, $\eta p2 = 0.597$. The latter suggested that participants were slower in detecting a human when the distractor was a robot (0.581 ± 0.021 s) compared to a guitar (0.519 ± 0.017 s; p < 0.001) or a hammer (0.525 ± 0.015 s; p < 0.001). Similarly, but to a lesser extent, participants were slower in detecting a robot when the distractor was a human (0.549 ± 0.019 s) compared to a guitar (0.513 ± 0.015 s; p = 0.010) or a hammer (0.517 ± 0.017 s; $p_{LMM} = 0.056$, $p_{mean} = 0.001$; $p_{median} = 0.045$). Participants were also slower in detecting a guitar when the distractor was a hammer (0.537 ± 0.014 s) compared to a human (0.482 ± 0.014 s; p < 0.001) or a robot (0.497 ± 0.015 s; p = 0.005). Finally, participants were slower in detecting a hammer when a guitar (0.550 ± 0.018 s) compared to a human (0.487 ± 0.019 s; p < 0.001) or a robot (0.479 ± 0.016 s; p < 0.001) acted as distractor.

Participants were faster in detecting hammers and guitars when a human was a distractor than detecting a human when hammers and guitars acted as distractors (all ps < 0.019). However, participants were only faster in detecting hammers when a robot was a distractor than detecting a robot when hammers were the distractor (p = 0.009). No other Bonferroni corrected p-values were lower than 0.05 for all multiple comparisons computed on the estimates of the simplified LMM and confirmatory multiple comparisons using pairwise t-tests on aggregated data.

4.2. Discussion experiment 2

Increasing the visual shape recognisability of the non-tool-object (i.e., guitars) and tool (i.e., hammers) categories successfully reduced performance variability across targets. Experiment 2 confirms that discriminating the non-tool-object and tool categories may have relied on similar processes. Indeed, we observed increased RTs when guitars and hammers were presented simultaneously. Since visual textures were not predictive of the two categories, this result may suggest that shape predictability (i.e., participants knew the shape to look for after the sample category) and dissimilarity (i.e., hammers had a larger top compared to guitars that were more rounded in the lower part) alone cannot explain these findings. Contrary, being part of the same super-ordinate category (e.g., non-living entities) may explain the increased RTs.

Improving the visual predictability of non-human-like stimuli may have also helped participants to rely more on the visual shape of the observed stimulus. Indeed, we did not see increased RTs when participants had to identify guitars with robots rather than humans as distractors. Importantly, while we confirmed slower RTs in categorising humans when robots rather than non-human-like stimuli acted as distractor, the same differences revealed to be smaller when participants had to categorise robots (see Table S5). This may indirectly confirm that the cognitive and neural representation of robots may include processes related to the analyses of both living and non-living categories.

Furthermore, we observed asymmetries between human and robots when they were paired with guitars and hammers (see Discussion). In particular, while participants were faster in identifying a hammer with robots as distractors rather than a robot with a hammer as distractor, such asymmetry was not observed for guitars. This further support the idea that results were not merely driven by low-level (dis)similarities across categories and that robots may belong to multiple superordinate categories (i.e., non-living non-tool entities).

Overall, Experiment 2 suggests that humanoid robots may share properties typical of living and non-living entities. To further test this prediction, we devised an experiment with four categories sharing a human-like shape but characterised by a different visual texture. If robots and non-tool-objects are indeed processed by the same
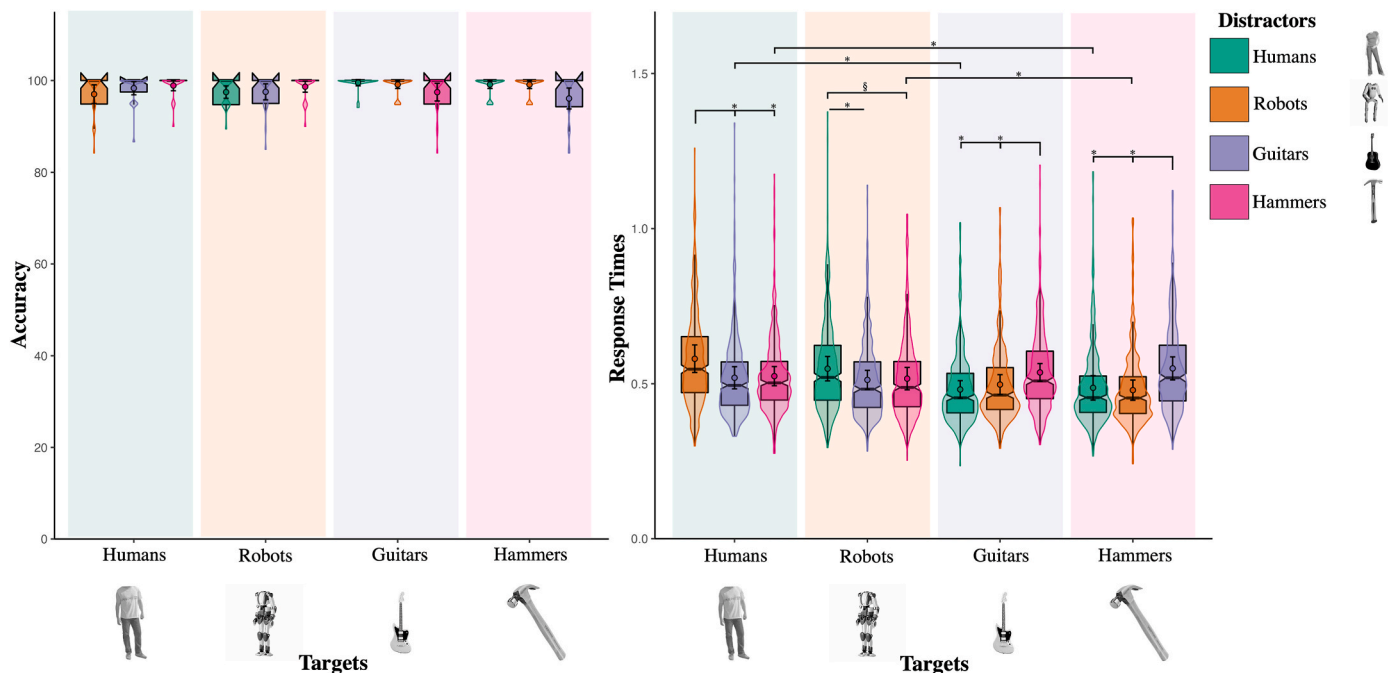

Fig. 4. Participants' performance (Accuracy percentage on the left, Response Times expressed in seconds on the right) in Experiment 2. See Fig. 3 for a detailed explanation of our data visualisation approach.

neurocognitive mechanisms, then we should observe specific interferences only when a distractor belongs to the same superordinate category.

5. Experiment 3 – is a humanoid robot perceived as a mannequin?

Participants performed the same task detailed in Experiment 1. Stimuli of humans and robots were identical to Experiment 1. However, we replaced guitars and hammers with images of other living and non-living human-like entities, namely monkeys and wooden mannequins. This way, while visual form was consistent across categories, only the visual texture of the object could clearly differentiate each stimulus (Fig. 1).

We removed trials with RTs deemed too fast or too slow (4.78%). Then, trials with RTs falling above or below 2.5SD of the overall mean within each block of each participant were removed (2.91%). Twelve participants' performance was <65%. Ten participants with a performance above 2.5SD of the overall mean across conditions of the remaining participants were excluded (final sample n = 128; average trial per condition in the final dataset [min, max]: 15.20 ± 0.02 s.e.m. [10, 16]).

5.1. Results

For Accuracy (see Fig. 5), we observed a main effect of Target, $\chi^2(3)$ = 24.098, p < 0.001, a main effect of Distractor, $\chi^2(3)$ = 66.427, p < 0.001, and a significant Target by Distractor interaction, $\chi^2(5)$ = 12.718, p = 0.026. The latter suggested that participants were less accurate in detecting a human when the distractor was a mannequin (95.39 ± 0.62%) compared to an ape (97.45 ± 0.43%; p = 0.014). Participants were less accurate in detecting a robot when the distractor was a mannequin (93.77 ± 0.60%) compared to a human (97.28 ± 0.45%; p < 0.001) or an ape (97.72 ± 0.43%; p < 0.001). Moreover, participants were less accurate in detecting a mannequin when the distractor was a robot (94.59 ± 0.62%) compared to a human (97.34 ± 0.44%; p <

0.001) or an ape (97.56 ± 0.37%; p < 0.001). We observed an asymmetry with participants tending to be more accurate in the detecting a mannequin with human bodies (not robotic bodies) as distractor than detecting a human with a mannequin as distractor (p_{LMM} = 0.036, p_{mean} = 0.081). No other Bonferroni corrected p-values were lower than 0.05 for multiple comparisons computed on the estimates of the simplified LMM and using pairwise t-tests on aggregated data (see Methods for data analysis approach; see Table S6 summarising all comparisons and effect sizes in Experiment 3).

For RTs, we removed incorrect answers (3.39%) from the final dataset. We observed a main effect of Target, F(3, 182) = 11.596, p < 0.001, $\eta p2$ = 0.247, a main effect of Distractor, F(3,166) = 22.350, p < 0.001, $\eta p2$ = 0.486, and a significant Target by Distractor interaction, F (5, 135) = 10.517, p < 0.001, $\eta p2$ = 0.287. The latter suggested that participants were slower in detecting a human when the distractor was a mannequin (0.659 ± 0.008 s) rather than a robot (0.616 ± 0.008 s; p < 0.001) or an ape (0.623 ± 0.007 s; p < 0.001). Similarly, participants were slower in detecting a robot when the distractor was a mannequin (0.665 ± 0.010 s) rather than a human (0.600 ± 0.008 s; p < 0.001) or an ape (0.606 ± 0.008 s; p < 0.001).

We observed asymmetries when apes, humans and robots were paired, with participants being faster in detecting an ape with a human and a robot as distractor (ape-human: 0.584 ± 0.007 s; ape-robot: 0.606 ± 0.008 s) than detecting a human or a robot (ape-human vs human-ape, p = 0.004; ape-robot vs robot-ape, p = 0.015). Similar asymmetries were observed when mannequins, humans and robots were paired, with participants being faster in detecting a mannequin paired with a human and a robot (mannequin-human: 0.607 ± 0.007 s; mannequin-robot: 0.616 ± 0.009 s) than detecting a human and a robot (mannequin-human vs human-mannequin, p < 0.001; mannequin-robot vs robot-mannequin, p < 0.001). No other Bonferroni corrected p-values were lower than 0.05 for all multiple comparisons computed on the estimates of the simplified LMM and confirmatory multiple comparisons using pairwise t-tests on aggregated data (see Methods for data analysis approach).
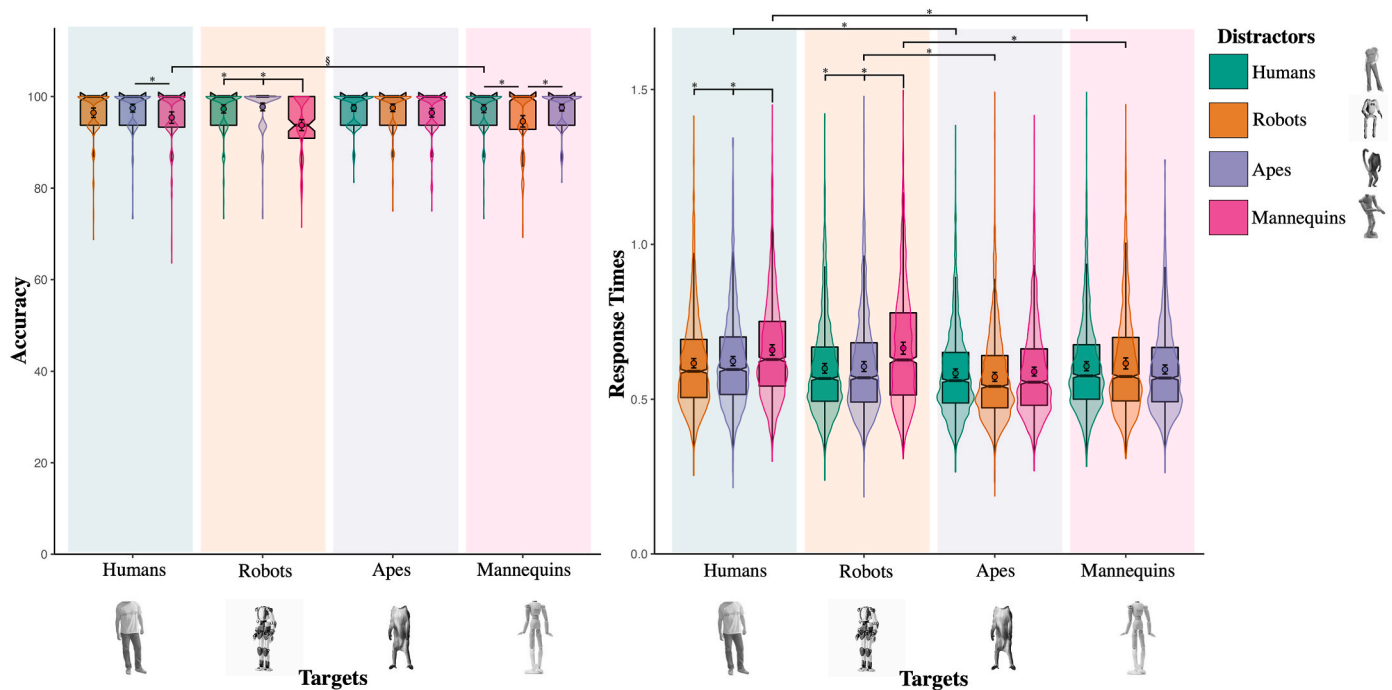

Fig. 5. Participants' performance (Accuracy percentage on the left, Response Times expressed in seconds on the right) in Experiment 3. See Fig. 3 for a detailed explanation of our data visualisation approach.

5.2. Discussion experiment 3

Results confirmed that humanoid robots may be considered a hybrid category. Although categorising humans with mannequins rather than robots and apes as distractors slowed RTs, categorising robots with mannequins rather than humans and apes as distractors revealed greater numerical differences and effect sizes (Table S6). Moreover, we observed decreased accuracies when participants had to identify both a robot with a mannequin as distractor and a mannequin with a robot as distractor. The fact that a similar result was not observed for monkeys cannot be attributed to their visual texture (i.e., fur) as all stimuli could be identified by their visual textures (humans: skin and clothes; robots: metal and plastic; apes: skin and fur; mannequin: wood). In other words, only the belonging of a visual texture to the same superordinate non-living category may explain the results. Thus, the fact that greater effects sizes and lower accuracies were observed when the mannequin and robot stimuli were paired may indicate that humanoid robots are indeed closer to a non-living human-like category rather than a living human-like category as monkeys are. Furthermore, the asymmetries in accuracy and RTs suggest that apes and mannequins may have features that deviate from the canonical processing of a human and robotic bodies (see Discussion).

To further test this interpretation and expand our understanding of human perception of humanoid robots, we combined all stimuli of Experiments 1–3 and tested them in a priming task. If the neurocognitive pathways responsible for processing humanoid robots are shared across non-living categories, then we should have expected faster RTs when robots are preceded by robots and other non-living objects (i.e., mannequins, non-tool-objects) rather than living ones (e.g., humans, apes).

6. Experiment 4 – Priming human and humanoid bodies

We combined the categories used in Experiments 2 and 3 (20 stimuli for six categories: humans, humanoids, monkeys, mannequins, guitars, hammers). We also added a control condition using masks as neutral prime (see Apparatus and Task). Humans and humanoid robot masks were excluded, and the neutral prime was randomly selected from a pool of 80 masks obtained from the remaining categories.

We removed trials with RTs deemed too fast or too slow (1.79%). Then, trials with RTs falling above or below 2.5SD of the overall mean within each block of each participant were removed (2.66%). No participants' performance was <65%. Four participants with a performance above or below 2.5SD of the overall mean across conditions of the remaining participants were excluded. Despite this data management approach, one participant had 0% accuracy when the robot prime preceded a human body and when the human prime preceded a robotic body, suggesting a misunderstanding of the task. Moreover, a participant had very low accuracy (12%) when the robot prime preceded a human body, suggesting more difficulty in completing that condition. Thus, we removed also those participants (final sample n = 94; average trial per condition in the final dataset [min, max]: 19.17 ± 0.03 s.e.m. [13, 20]).

6.1. Results

For Accuracy (see Fig. 6), we observed no main effect of Prime, $\chi^2(6) = 10.449$ p = 0.107, a main effect of Target, $\chi^2(1) = 6.420$, p = 0.011, and a significant Prime by Target interaction, $\chi^2(6) = 66.763$, p < 0.001. The latter suggested that participants were less accurate in detecting a human when the prime was a robot (92.12 ± 0.96%) compared to a mask (96.02 ± 0.51%; p < 0.001), a human (96.76 ± 0.52%; p < 0.001), and an ape (95.50 ± 0.64%; p = 0.004). Moreover, participants were less accurate in detecting a robot when the prime was a human (92.58 ± 0.97%) compared to a mask (96.28 ± 0.65%; p < 0.001), a robot (96.31 ± 0.58%; p < 0.001), a mannequin (96.46 ± 0.56%; p < 0.001), a guitar (96.22 ± 0.57%; p < 0.001), and a hammer (96.23 ± 0.57%; p < 0.001). Unsurprisingly, human and robot primes produced opposite results. Participants were more accurate in categorising a human than a robot preceded by a human prime (p < 0.001), and in categorising a robot than a human preceded by a robot prime (p < 0.001). No other Bonferroni corrected p-values were lower than 0.05 for multiple comparisons computed on the estimates of the simplified LMM and using pairwise t-tests on aggregated data (see Methods for data analysis approach; see Table S7 summarising all comparisons and effect sizes in
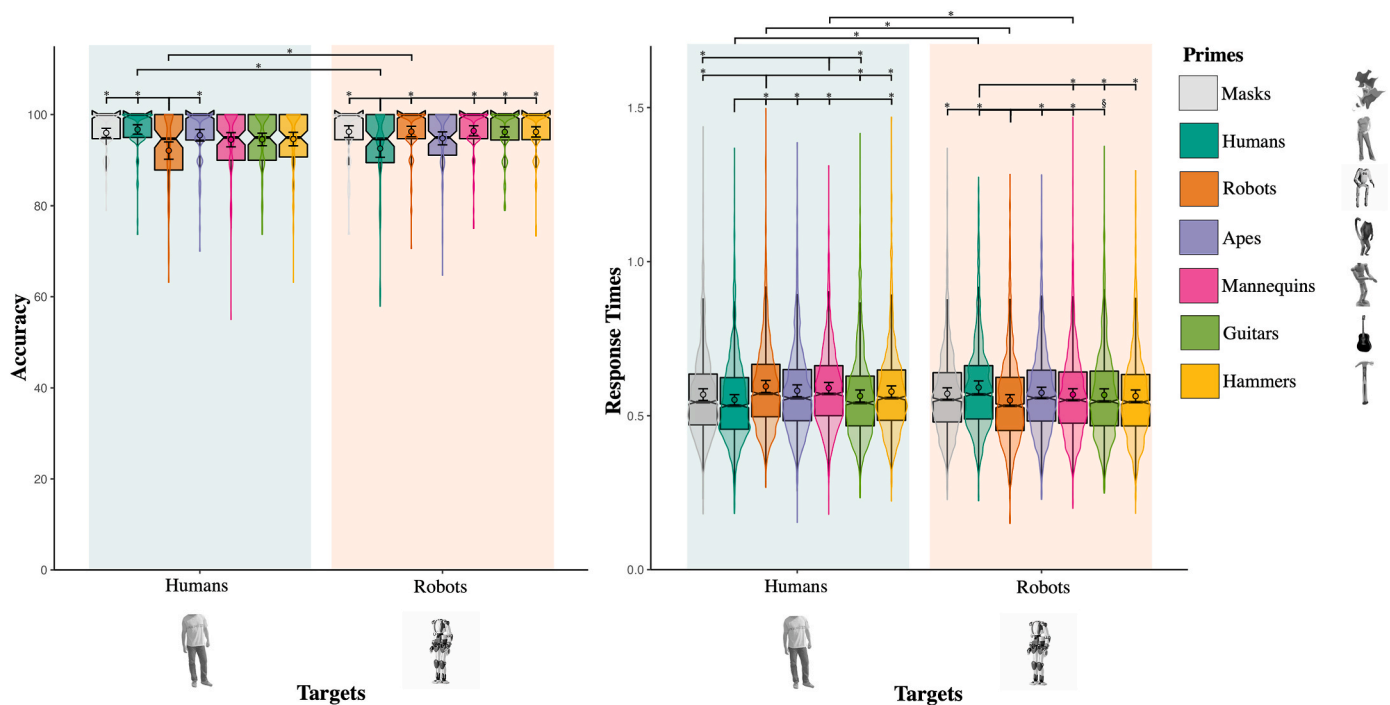

Fig. 6. Participants' performance (Accuracy percentage on the left, Response Times expressed in seconds on the right) in Experiment 4. See Fig. 3 for a detailed explanation of our data visualisation approach.

Experiment 4).

For RTs, we removed incorrect answers (4.73%) from the final dataset. We observed a main effect of Prime, $F(6, 202) = 4.183$, $p = 0.001$, $\eta p2 = 0.112$, no main effect of Target, $F(1, 70) = 1.378$, $p = 0.244$, $\eta p2 = 0.019$, and a significant Prime by Target interaction, $F(6, 1111) = 33.204$, $p < 0.001$, $\eta p2 = 0.158$. The latter suggested that participants were faster in detecting a human when the prime was a human (0.552 ± 0.008 s) compared to a robot (0.596 ± 0.010 s; $p < 0.001$), an ape (0.582 ± 0.010 s; $p < 0.001$), a mannequin (0.591 ± 0.009 s; $p < 0.001$), and a hammer prime (0.579 ± 0.009 s; $p < 0.001$). Participants were also slower in detecting a human when it was preceded by a robot prime rather than a mask (0.570 ± 0.010 s; $p < 0.001$), a guitar (0.565 ± 0.010 s; $p < 0.001$), and a hammer prime ($p = 0.010$). Moreover, participants were also slower in detecting a human when it was preceded by a mannequin prime rather than a mask and a guitar (both $p < 0.001$).

When a robot was preceded by a robot, participants were faster (0.552 ± 0.009 s) compared to a mask (0.573 ± 0.010 s; $p < 0.001$), a human (0.593 ± 0.011 s; $p < 0.001$), an ape (0.576 ± 0.009 s; $p < 0.001$), a mannequin (0.570 ± 0.010 s; $p = 0.003$), and tended to be faster compared to a guitar prime (0.569 ± 0.010 s; $p_{LMM} = 0.011$; $p_{mean} = 0.098$; $p_{median} = 0.017$). Participants were also slower when the robot was preceded by a human prime compared to a mannequin ($p < 0.001$), a guitar (0.568 ± 0.010 s; $p < 0.001$), and a hammer (0.565 ± 0.010 s; $p < 0.001$).

Finally, human, robot, and mannequin primes had opposite effects on human and robot targets. Specifically, participants were faster in detecting a human rather than a robot target with a human prime ($p < 0.001$), and were slower in detecting a human rather than a robot with a robot and mannequin primes (robot prime, $p < 0.001$; mannequin prime, $p = 0.025$).

6.2. Discussion experiment 4

Results suggest that recognising a human body is facilitated when participants are primed with a human body and reduced when non-human primes are displayed (i.e., robots, apes, mannequins, hammers). A similar facilitation in the RTs was observed when the prime was a guitar. Contrary, recognising a robot is facilitated by robot primes and gradually hindered the more the prime moves from a non-human-like non-living object (i.e., hammer, guitar) towards a human-like living entity (i.e., apes, humans) with the mannequin prime differing both from the robot and human prime. Crucially, as the perception of human and robot primes prior to the (human or robot) target may have pre-activated the manual response, accuracy and RTs in those conditions may not highlight semantic effects. The same limitation cannot be said for the other primes that belonged to other categories and were not mapped into any response selection (Henson et al., 2014).

Overall, while results when the target was the robot may appear straightforward (i.e., the more a prime resembles a human shape and possesses the visual textures of a living entity, the more it interferes with robot categorisation), it may be less straightforward to explain why a guitar prime did not hinder the recognition of human targets. However, it is reasonable to assume that functional associations evoked by some primes may have favoured the categorisation of a human body. In this specific case, guitars may have been associated with human bodies as guitars are played close to trunk and legs, while hammers may have been associated to body parts only (i.e., hands). This interpretation may explain why guitars did not interfere with the categorisation of humans but tended to produce slower answers when participants had to categorise a robot (i.e., guitars evoking more the image of a human body than hammers).

To test this interpretation, we ran a final experiment to understand whether people may associate guitars to human bodies more than hammers.

7. Experiment 5 –Perceptual and categorical dimensions of human-like and non-human-like stimuli

Stimuli were the same used in Experiment 4 (20 stimuli for each of the 6 categories: humans, humanoids, apes, mannequins, guitars, hammers). Participants saw six images, one for each category, and arranged them based on four criteria in separate blocks (see Apparatus and Task for further details). Four Participants did not pass our pre-registered exclusion criteria (i.e., a minimum rank of 4 for human stimuli to the *Perceived Similarity* and *Perceived Liveliness* questions; final sample n = 66).

7.1. Results

We analysed the ranking score given to each category with Category (human, robot, non-tool-object, object, mannequin, ape) as within-subject factor (see Fig. 7). In the case of a main effect, we performed post-hoc comparisons on the statistical model (see Methods for data analysis approach; see Table S8 summarising all comparisons and effect sizes in Experiment 5).

The ANOVA on the *Perceived Similarity* to a human shape revealed a main effect of Category, $F(3.27, 212.54) = 417.281$, $p < 0.001$, $\eta p2 = 0.865$. Each category differed from one another with participants ranking human bodies more like human bodies (4.985 ± 0.015) than mannequins (3.424 ± 0.081), robots (3.000 ± 0.097), apes (2.500 ± 0.108), guitars (0.879 ± 0.067), and hammers (0.212 ± 0.063).

The ANOVA on the attribute *Liveliness* of each category revealed a main effect of Category, $F(3.04, 197.83) = 253.401$, $p < 0.001$, $\eta p2 = 0.796$. Each category differed from one another, exception made between robots and mannequins, with participants ranking human bodies as belonging to a living category (4.818 ± 0.048) more than apes (3.939 ± 0.094), mannequins (2.394 ± 0.101), robots (2.348 ± 0.129), guitars (1.152 ± 0.102), and hammers (0.348 ± 0.082).

The ANOVA on the ranking scores from the *Body Association* question, revealed a main effect of Category, $F(3.57, 231.94) = 354.698$, $p < 0.001$, $\eta p2 = 0.845$. Each category differed from one another, exception made between robots and apes, with participants ranking human bodies as evoking more images of a human body (4.924 ± 0.050) than mannequins (3.500 ± 0.100), robots (2.833 ± 0.096), apes (2.621 ± 0.107), guitars (0.924 ± 0.069), and hammers (0.197 ± 0.062).

The ANOVA on the ranking scores from the *Action Association* question, revealed a main effect of Category, $F(2.34, 151.96) = 20.298$, $p < 0.001$, $\eta p2 = 0.238$. Human stimuli differed from all categories with participants ranking human bodies as evoking more images of a human actions (4.273 ± 0.168) than mannequins (2.470 ± 0.159), robots (2.364 ± 0.160), apes (2.212 ± 0.181), guitars (2.000 ± 0.197), and hammers (1.682 ± 0.234). Ranking scores attributed to mannequins, robots, apes, guitars, and hammers did not differ from one another.

7.2. Discussion experiment 5

Results confirmed our initial guesses that guitar stimuli may have been associated to whole human bodies more than hammers. Importantly, guitars and hammers did not differ in their association to human actions. This suggests that the findings in the previous experiments were driven by the perceptual and superordinate representations of the observed categories rather than their associations to motor actions.

Experiment 5 may also explain why mannequins may have interfered with the processing of humans and robots in Experiments 3. Indeed, mannequins were perceived more like a human shape than apes and robots (Q1) and evoked an image of a human body more than robots (Q3). However, mannequins and robots did not differ in their belonging to a non-living category (Q2). Hence, accuracy levels in Experiment 3 may reflect the perceived belonging to the same superordinate non-living category (we observed an accuracy drop when the mannequin was the distractor for robots and viceversa). Moreover, similar ratings
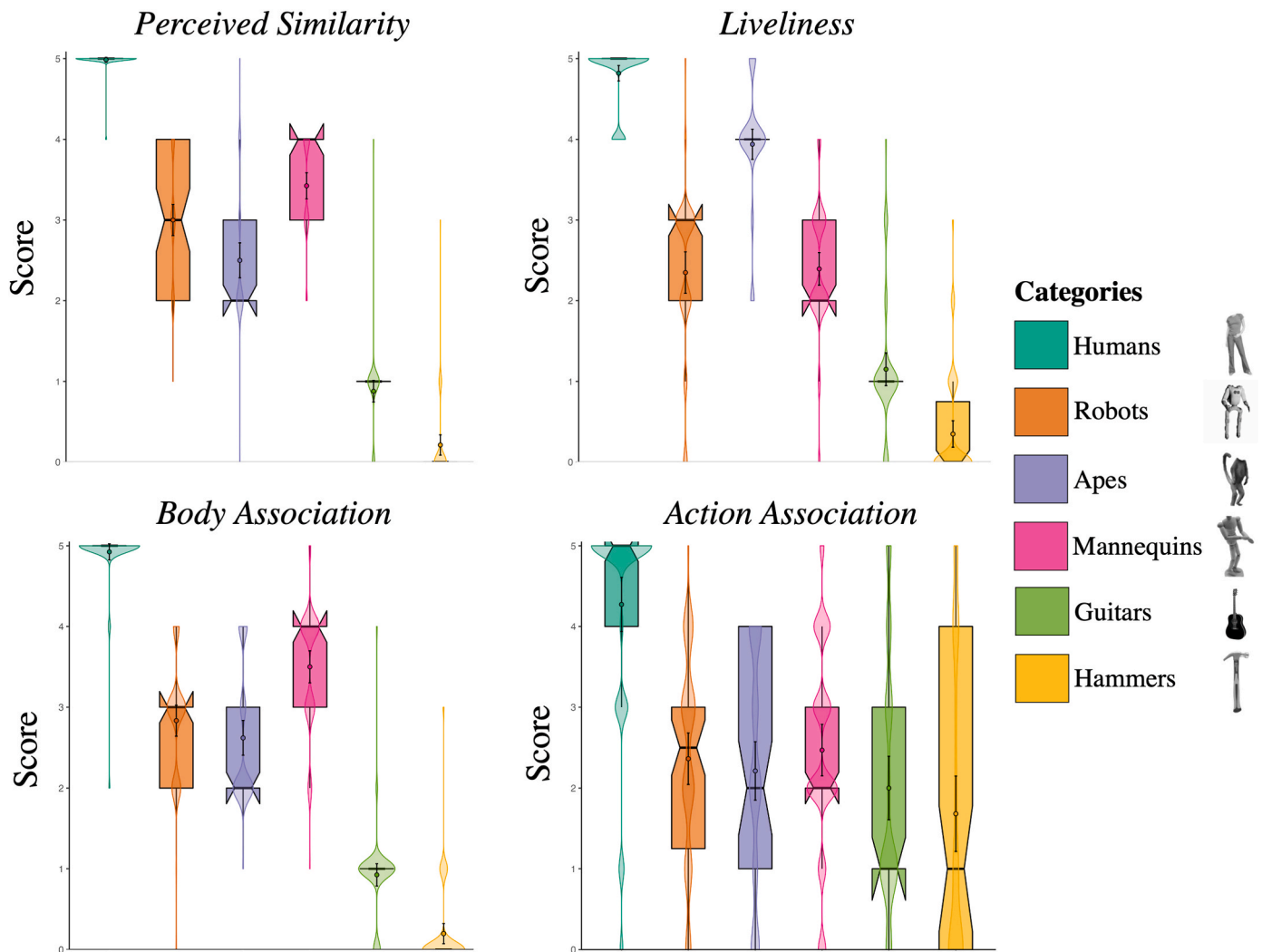
Fig. 7. Participants' ranking score for each question and category. Participants ranked six categories, therefore the minimum-maximum ranking score is 0–5. See Fig. 3 for a detailed explanation of our data visualisation approach. Ratings did not differ between Robots and Mannequins for the Liveliness question, and between Robots and Apes for the Body association Question. Only Humans were associated more to a human Action for the Action association question.

for the perceived liveliness may explain the greater effect size in the RTs of Experiment 3 for categorising a robot rather than a human when a mannequin was the distractor.

8. Discussion

In the coming decades, we expect to see an increased presence of embodied artificial agents, like humanoid robots, in society. Will people perceive robots with a human-like shape as human agents, as objects, or something between a human-like animal and an object? Across five behavioural experiments, we tested whether processing humanoid robots shares the same mechanisms active during the recognition of (non-) human-like (non-)living entities. Results suggested that humanoid robots are processed more like mindless mannequins than non-human animals. Although we tested only humanoid robots with no human-like visual textures (i.e., without human skin-like appearance), results suggest that the perceptual processes responsible for a first visual impression of a human-like robot may not be so different from the mechanisms responsible for analysing a non-living and inanimate (i.e., unable to self-propel) objects.

It is unlikely that our findings can be explained by the idea that processing humanoid robots is more effortful. In the first three experiments, the time taken to categorise a robot was never greater than the

time necessary to categorise a human. Contrary, numerical values suggested the opposite (i.e., faster RTs to categorise robots rather than humans). Moreover, the priming task in Experiment 4 showed no differences in categorising a human and a robot when the prime was a neutral mask.

In the following sections, we discuss how processing mid-level features alone (i.e., shape, visual textures, number of corners) and higher-order information may explain our data and help to visually categorise an object.

8.1. Categorising humans and humanoid robots

Results from a match-to-category task (Experiments 1–2) confirm the importance of visual shape analyses for object categorisation (Baldassi et al., 2013) and suggest that recognising a humanoid robot may only in part rely on the same processes active when people perceive a living category.

Furthermore, in Experiment 2, categorising non-human-like objects (i.e., hammers and guitars) was faster than categorising a human with hammers and guitars as distractors. The same was not observed when guitars were presented simultaneously with robots. These asymmetries (Treisman & Gormican, 1988) suggest that while identifying a non-human-like object with a predictable shape is faster than processing

a human body (i.e., non-human-like objects have features that deviate from the canonical processing of a human body, hence people are faster when looking for them), processing hammers but not guitars was faster than processing humanoid robots (i.e., guitars may have features that do not deviate from the canonical processing of a robotic body). It may be argued that the perception of tools evoked a motor response that speeded RTs, hence, the absence of such difference between humanoid robots and non-tool objects should not be considered as evidence that these two categories may share similar neurocognitive representations. However, we note that tools and non-tool-objects were matched for the ability to evoke manual actions (see Bracci & Peelen, 2013 and Experiment 5 in this study).

Building on these considerations, visual texture similarity (i.e., hammers and guitars shared similar visual textures: metal, plastic, and woods) or automatic motor responses evoked by the perceived stimulus cannot fully explain our data. On the other hand, it is plausible that participants relied also on shared and more abstract qualities of the perceptual stimuli to correctly categorise humans and humanoid robots (e.g., humans as living vs robots, guitars, and hammers as non-living categories).

In this sense, Experiment 3 confirmed that robots may be processed as objects belonging to a non-living category rather than a non-human animal one. Indeed, robots and mannequins interfered with each other, and no search asymmetry was not observed. This reciprocal interference was the only one across the four stimuli used (humans, apes, robots, mannequins). Moreover, the fact that apes (a living category) did not interfere more than robots when participants looked for humans suggests that different neurocognitive pathways may help categorising human-animals for humans and non-human-animals for apes (Mur et al., 2013). This result is in line with recent studies suggesting different neural correlates for processing bodies belonging to different species (Chesley et al., 2024). On the other hand, a search asymmetry between human bodies and mannequins (with greater accuracy and faster RTs for the latter) and between human bodies and apes (with faster RTs for the latter) may suggest that agents with human-like shape but considered lacking mental capacities (Leach et al., 2023) may have been processed as deviations from the canonical processing of a human body. However, a similar asymmetry was also observed for robotic bodies (with faster RTs for the apes and mannequins) suggesting that apes and mannequins may have been processed as deviations from the canonical processing of a robotic body too. More studies are needed to assess how mind attribution to humans, monkeys, and robots (Gray et al., 2007; Leach et al., 2023) and their perceptual differences (e.g., fur, having the tail, different body ratios between humans, robots, and apes) may influence search asymmetries.

Finally, the priming task in Experiment 4 showed a gradual increase in RTs to recognise a humanoid robot the more a prime moved from a mechanical object towards a living human. It is of note that a non-human animal prime (i.e., apes) slowed participants responses for both human and humanoid robot targets. This may further suggest that a human body may be treated as a special living category separate from nonhuman animals (Chesley et al., 2024), and that humanoid robots may share little of the qualities that human-like animals have (e.g., being alive, having a mind). Moreover, a guitar prime followed by a human target did not delay participants responses as a hammer prime did. This finding is in line with the idea that the recognition of an object is likely based on a more complex network of associations learned through experience (Cross et al., 2012).

To summarise, our results have two opposite implications. While human bodies may be less associated to non-human human-like animals, despite perceptual dissimilarity they may be more associated to non-human-like objects that are functionally related to the human body (i. e., in our case guitars) thus suggesting a top-down or distributed semantic knowledge influencing object categorisation (Ralph et al., 2017). On the other hand, while monkeys, robots, and mannequins may share some (mostly visual) features with humans, their mental capacities are

likely perceived very differently (Gray et al., 2007; Leach et al., 2023) and consequently are also likely to be represented as separate categories. The latter result seems to indirectly confirm a hierarchical processing of object categorisation. That is, the closer the visual resemblance between two categories, the more multiple high-order information (i.e., not only the belonging to a living vs non-living category) may affect task performance (Yeh & Peelen, 2022).

8.2. Are humanoid robots represented as mindless mannequins?

In Experiment 3, the visual form was very similar across stimuli and each category had a specific visual texture (Fig. 1). We observed that when the target was a human, the accuracy decreased and RTs increased with a mannequin as distractor. This result suggest that common mid-level features between the mannequins and robots (e.g., the body junctions were more visible in robots and mannequins than in humans) could not explain participants' performance. However, the mannequin's smooth edges may have promoted the perception of the stimulus as a living entity (Levin et al., 2001; Long et al., 2017; Zachariou et al., 2018). In this sense, Experiment 5 showed that mannequins may have been associated to a human body more than robots. However, in Experiment 3 human distractors did not reduce the ability to identify mannequins and we observed a greater interference when robots and mannequins were paired. Thus, our results are not a by-product of some shared mid-level features across humans, mannequin, and robots. Moreover, if shared mid-level features between robots and mannequins could fully explain our results, in Experiment 4 we should have seen mannequins facilitating the recognition of robots more than all other primes. Contrary, RTs for mannequins' primes were at an intermediate stage between human and robot primes and RTs after a hammer prime did not differ from a robot prime.

All in all, it is plausible that mannequins, humans, and robots shared the same body schema (Orlandi & Proverbio, 2020) and further processing of their categorical properties (living vs non-living) may have been required (Yeh & Peelen, 2022). However, only robots and mannequins shared to the same higher-level properties (non-living, non-human mind).

A recent framework suggests how semantic knowledge may support non-verbal behaviour (Binney and Ramsey, 2020; Ralph et al., 2017), in our study categorising objects using manual responses (i.e., non-verbal responses). According to that model, categorising a human body in the way participants did in our Experiments may not rely on a single localised brain area, but instead on higher-order information distributed across the cortex. Moreover, contextual information like task demands may influence which conceptual properties are more relevant to process, as well as how they interact with domain-specific neural networks representing an object. In other words, executive task demands may control how higher-order information associated with an object and modality-specific brain areas support object categorisation.

Our results fit well within such a model. In the match-to-category task (Experiments 1–3), the presentation of the target may have prioritised more the processing of specific low-level features (e.g., the shape of the visual object). Contrary, in the priming task (Experiment 4), performance changed based on the degree of association with the target stimulus (human or robot) and the prime (which was irrelevant to solve the task) pre-activated a widely distributed semantic network that then facilitated the categorising of humans and robots.

If this interpretation is correct, its implication for future studies on human—robot interaction is to assess the extent to which previous conceptions of robots may impact a task. For example, based on our results, if even human-like robots are perceived as mindless objects, it will be fundamental to ensure robots are first perceived as having a mind (i.e., increasing their socialness; Jastrzab et al., 2024) at the start of a social interaction to increase the chance of more intuitive and effective collaborations.

8.3. Limitations of the current study

In this series of experiments, we focused on humanoid robots with machine-like visual textures. However, we did not find clear evidence that participants used visual texture cues more to categorise robots than humans. We also did not use robots with human-like skin or robots with animal-like shape, or nor did we study mechanical-looking robots. So, our conclusions are limited to the adopted set of stimuli. Moreover, we presented stimuli with neutral postures. Future studies may investigate how robots with unusual postures (e.g., legs with orientations different from human legs' orientation) or indeed, even dynamic displays of robots in action, may affect the processing of humanoid robots as social agents. We also note that participants made few errors in completing the tasks and findings rely more on reaction times. Harder tasks are necessary to further confirm our results. Moreover, we do not have direct evidence on which theoretical model of visual object categorisation best explains our results or whether the visual processing of robots was hierarchical or influenced by a top-down more distributed semantic network. Finally, as our tasks focused on the explicit categorisation of the observed stimulus (i.e., processing the stimulus category was always relevant to solve the tasks), future studies will be required to investigate the network of cortical areas responsible for integrating mid-level and superordinate information during robot perception (Ekman et al., 2020).

9. Conclusions

Some humanoid robots may have the potential to be visually confused for humans, due their visual form. However, they can also be readily categorised as non-living agents based on other perceptual features (e.g., visual textures) and superordinate dimensions (e.g., not having a human mind). The results we report here expand current literature investigating how people perceive robots (Geiger & Balas, 2021) and have implications for how people perceive robots' actions (e. g., increase robot's perceived socialness; Tidoni et al., 2022; Scandola et al., 2023; Jastrzab et al., 2024) by confirming that robots with a human-like form may be perceived as a hybrid category, and that robots are presently more closely associated to mindless objects than living agents. Based on the malleability of perceptual processes with experience (Cross et al., 2012; Frank et al., 2023) our tasks open new possibilities to explore how direct and vicarious experience engaging or interacting with robots may change these perceptual representations.

CRediT authorship contribution statement

Emmanuele Tidoni: Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Emily S. Cross:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Methodology, Conceptualization. **Richard Ramsey:** Writing – review & editing, Writing – original draft, Visualization, Validation. **Michele Scandola:** Writing – review & editing, Writing – original draft, Validation, Formal analysis.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbah.2024.100105.

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research, 4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2

Ayzenberg, V., & Behrmann, M. (2022). Does the brain's ventral visual pathway compute object shape? *Trends in Cognitive Sciences, 26*(12), 1119–1132. https://doi.org/10.1016/j.tics.2022.09.019

Baldassi, C., Alemi-Neissi, A., Pagan, M., Dicarlo, J. J., Zecchina, R., & Zoccolan, D. (2013). Shape similarity, better than semantic membership, accounts for the structure of visual object representations in a population of monkey inferotemporal neurons. *PLoS Computational Biology, 9*(8), Article e1003167. https://doi.org/10.1371/journal.pcbi.1003167

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2008). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics, 1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Barton, K. (2016). *MuMIn: Multi-model inference* (Vol. 1, pp. 1–15). R Packag version. https://cran.r-project.org/web/packages/MuMIn/index.html.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1). https://doi.org/10.18637/jss.v067.i01

Ben-Shachar, M., Lüdecke, D., & Makowski, D. (2020). effectsize: Estimation of effect size indices and standardized parameters. *Journal of Open Source Software, 5*(56), 2815. https://doi.org/10.21105/joss.02815

Binney, R. J., & Ramsey, R. (2020). Social Semantics: The role of conceptual knowledge and cognitive control in a neurobiological model of the social brain. *Neuroscience & Biobehavioral Reviews, 112*, 28–38. https://doi.org/10.1016/j.neubiorev.2020.01.030

Bracci, S., Daniels, N., & op de Beeck, H. (2017). Task context overrules object- and category-related representational content in the human parietal cortex. *Cerebral Cortex, 27*(1), 310–321. https://doi.org/10.1093/cercor/bhw419

Bracci, S., & op de Beeck, H. (2016). Dissociations and associations between shape and category representations in the two visual pathways. *Journal of Neuroscience, 36*(2), 432–444. https://doi.org/10.1523/JNEUROSCI.2314-15.2016

Bracci, S., & Op de Beeck, H. P. (2023). Understanding human object vision: A picture is worth a thousand representations. *Annual Review of Psychology, 74*(1), 113–135. https://doi.org/10.1146/annurev-psych-032720-041031

Bracci, S., & Peelen, M.v. (2013). Body and object effectors: The organization of object representations in high-level visual cortex reflects body-object interactions. *Journal of Neuroscience, 33*(46), 18247–18258. https://doi.org/10.1523/jneurosci.1322-13.2013

Bracci, S., Ritchie, J. B., Kalfas, I., & op de Beeck, H. P. (2019). The ventral visual pathway represents animal appearance over animacy, unlike human behavior and deep neural networks. *Journal of Neuroscience, 39*(33), 6513–6525. https://doi.org/10.1523/JNEUROSCI.1714-18.2019

Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision, 10*(4), 433–436. https://doi.org/10.1163/156856897X00357

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain: The animate-inanimate distinction. *Journal of Cognitive Neuroscience, 10*(1), 1–34. https://doi.org/10.1162/089892998563752

Chesley, J., Riecke, L., Lu, J., Vogels, R., & De Gelder, B. (2024). Theta activity discriminates high-level, species-specific body processes. *Imaging Neuroscience, 2*, 1–12. https://doi.org/10.1162/imag_a_00150

Contini, E. W., Goddard, E., Grootswagers, T., Williams, M., & Carlson, T. (2020). A humanness dimension to visual object coding in the brain. *NeuroImage, 221*, Article 117139. https://doi.org/10.1016/j.neuroimage.2020.117139

Cross, E. S., Cohen, N. R., Hamilton, A. F. de C., Ramsey, R., Wolford, G., & Grafton, S. T. (2012). Physical experience leads to enhanced object perception in parietal cortex: Insights from knot tying. *Neuropsychologia, 50*(14), 3207–3217. https://doi.org/10.1016/j.neuropsychologia.2012.09.028

Cross, E. S., & Ramsey, R. (2021). Mind meets machine: Towards a cognitive science of human-machine interactions. *Trends in Cognitive Sciences, 25*(3), 200–212. https://doi.org/10.1016/j.tics.2020.11.009

Downing, P. E., Jiang, Y., Shuman, M., & Kanwisher, N. (2001). A cortical area selective for visual processing of the human body. *Science, 293*(5539), 2470–2473. https://doi.org/10.1126/science.1063414

Ekman, M., Roelfsema, P. R., & de Lange, F. P. (2020). Object selection by automatic spreading of top-down attentional signals in V1. *Journal of Neuroscience, 40*(48), 9250–9259. https://doi.org/10.1523/JNEUROSCI.0438-20.2020

Epley, N., Waytz, A., & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychology Review, 114*(4), 864–886. https://doi.org/10.1037/0033-295X.114.4.864

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

Frank, C., Kraeutner, S. N., Rieger, M., & Boe, S. G. (2023). Learning motor actions via imagery—perceptual or motor learning? *Psychological Research.* https://doi.org/10.1007/s00426-022-01787-4

Gandolfo, M., & Downing, P. E. (2020). Asymmetric visual representation of sex from human body shape. *Cognition, 205*, Article 104436. https://doi.org/10.1016/j.cognition.2020.104436

Geiger, A. R., & Balas, B. (2021). Robot faces elicit responses intermediate to human faces and objects at face-sensitive ERP components. *Scientific Reports, 11*(1), Article 17890. https://doi.org/10.1038/s41598-021-97527-6

Gray, H. M., Gray, K., & Wegner, D. M. (2007). Dimensions of mind perception. *Science, 315*(5812). https://doi.org/10.1126/science.1134475, 619–619.

Greven, I. M., Downing, P. E., & Ramsey, R. (2016). Linking person perception and person knowledge in the human brain. *Social Cognitive and Affective Neuroscience, 11* (4), 641–651. https://doi.org/10.1093/scan/nsv148

Grill-Spector, K., & Weiner, K. S. (2014). The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience, 15*(8), 536–548. https://doi.org/10.1038/nrn3747

Groen, I. I. A., Dekker, T. M., Knapen, T., & Silson, E. H. (2022). Visuospatial coding as ubiquitous scaffolding for human cognition. *Trends in Cognitive Sciences, 26*(1), 81–96. https://doi.org/10.1016/j.tics.2021.10.011

Henschel, A., Hortensius, R., & Cross, E. S. (2020). Social cognition in the age of human-robot interaction. *Trends in Neurosciences, 43*(6), 373–384. https://doi.org/10.1016/j.tins.2020.03.013

Henson, R. N., Eckstein, D., Waszak, F., Frings, C., & Horner, A. J. (2014). Stimulus–response bindings in priming. *Trends in Cognitive Sciences, 18*(7), 376–384. https://doi.org/10.1016/j.tics.2014.03.004

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

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

JASP Team. (2022). JASP (version 0.16.4)[computer software]. https://jasp-stats.org/.

Jastrzab, L. E., Chaudhury, B., Ashley, S. A., Koldewyn, K., & Cross, E. S. (2024). Beyond human-likeness: Socialness is more influential when attributing mental states to robots. *iScience, 27*(6), Article 110070. https://doi.org/10.1016/j.isci.2024.110070

Kleiner, M., Brainard, D. H., & Pelli, D. (2007). ECVP '07 abstracts. *Perception, 36*(1_suppl), 1–235. https://doi.org/10.1177/03010066070360S101

Leach, S., Sutton, R. M., Dhont, K., Douglas, K. M., & Bergström, Z. M. (2023). Changing minds about minds: Evidence that people are too sceptical about animal sentience. *Cognition, 230*, Article 105263. https://doi.org/10.1016/j.cognition.2022.105263

Lenth, R. (2019). Emmeans: Estimated marginal means. In *R packag* version 1.6.2. .

Levin, D. T., Takarae, Y., Miner, A. G., & Keil, F. (2001). Efficient visual search by category: Specifying the features that mark the difference between artifacts and animals in preattentive vision. *Perception & Psychophysics, 63*(4), 676–697. https://doi.org/10.3758/BF03194429

Long, B., Stormer, V. S., & Alvarez, G. A. (2017). Mid-level perceptual features contain early cues to animacy. *Journal of Vision, 17*(6), 20. https://doi.org/10.1167/17.6.20

Lowet, A. S., Firestone, C., & Scholl, B. J. (2018). Seeing structure: Shape skeletons modulate perceived similarity. *Attention, Perception, & Psychophysics, 80*(5), 1278–1289. https://doi.org/10.3758/s13414-017-1457-8

Lüdecke, D., Ben-Shachar, M., Patil, I., Waggoner, P., & Makowski, D. (2021). performance: An R package for assessment, comparison and testing of statistical models. *Journal of Open Source Software, 6*(60), 3139. https://doi.org/10.21105/joss.03139

Mervis, C. B., & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology, 32*(1), 89–115. https://doi.org/10.1146/annurev.ps.32.020181.000513

Moro, V., Urgesi, C., Pernigo, S., Lanteri, P., Pazzaglia, M., & Aglioti, S. M. (2008). The neural basis of body form and body action agnosia. *Neuron, 60*(2), 235–246. https://doi.org/10.1016/j.neuron.2008.09.022

Mur, M., Meys, M., Bodurka, J., Goebel, R., Bandettini, P. A., & Kriegeskorte, N. (2013). Human object-similarity judgments reflect and transcend the primate-IT object representation. *Frontiers in Psychology, 4*, 128. https://doi.org/10.3389/fpsyg.2013.00128

Nomura, T., Kanda, T., Suzuki, T., & Kato, K. (2008). Prediction of human behavior in human–robot interaction using psychological scales for anxiety and negative attitudes toward robots. *IEEE Transactions on Robotics, 24*(2), 442–451. https://doi.org/10.1109/tro.2007.914004

Orlandi, A., & Proverbio, A. M. (2020). ERP indices of an orientation-dependent recognition of the human body schema. *Neuropsychologia, 146*, Article 107535. https://doi.org/10.1016/j.neuropsychologia.2020.107535

Palan, S., & Schitter, C. (2018). Prolific.ac—a subject pool for online experiments. *Journal of Behavioral and Experimental Finance, 17*, 22–27. https://doi.org/10.1016/j.jbef.2017.12.004

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods, 51*(1), 195–203. https://doi.org/10.3758/s13428-018-01193-y

Phillips, E., Zhao, X., Ullman, D., & Malle, B. F. (2018). What is Human-like? *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, 105–113. https://doi.org/10.1145/3171221.3171268

Pitcher, D., & Ungerleider, L. G. (2021). Evidence for a third visual pathway specialized for social perception. *Trends in Cognitive Sciences, 25*(2), 100–110. https://doi.org/10.1016/j.tics.2020.11.006

Pobric, G., Jefferies, E., & Lambon Ralph, M. A. (2010). Category-specific versus category-general semantic impairment induced by transcranial magnetic stimulation. *Current Biology, 20*(10), 964–968. https://doi.org/10.1016/j.cub.2010.03.070

Proklova, D., & Goodale, M. A. (2022). The role of animal faces in the animate-inanimate distinction in the ventral temporal cortex. *Neuropsychologia, 169*. https://doi.org/10.1016/j.neuropsychologia.2022.108192

Proklova, D., Kaiser, D., & Peelen, M.v. (2016). Disentangling representations of object shape and object category in human visual cortex: The animate-inanimate distinction. *Journal of Cognitive Neuroscience, 28*(5), 680–692. https://doi.org/10.1162/jocn_a_00924

Quadflieg, S., Flannigan, N., Waiter, G. D., Rossion, B., Wig, G. S., Turk, D. J., & Macrae, C. N. (2011). Stereotype-based modulation of person perception. *NeuroImage, 57*(2), 549–557. https://doi.org/10.1016/j.neuroimage.2011.05.004

R Core Team. (2022). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. url = https://www.R-project.org/.

Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology, 25*, 111. https://doi.org/10.2307/271063

Ralph, M. A. L., Jefferies, E., Patterson, K., & Rogers, T. T. (2017). The neural and computational bases of semantic cognition. *Nature Reviews Neuroscience, 18*(1), 42–55. https://doi.org/10.1038/nrn.2016.150

Ramsey, R. (2018). Neural integration in body perception. *Journal of Cognitive Neuroscience, 30*(10), 1442–1451. https://doi.org/10.1162/jocn_a_01299

Reed, C. L., Stone, V. E., Bozova, S., & Tanaka, J. (2003). The body-inversion effect. *Psychological Science, 14*(4), 302–308. https://doi.org/10.1111/1467-9280.14431

Rogers, T. T., Hocking, J., Mechelli, A., Patterson, K., & Price, C. (2005). Fusiform activation to animals is driven by the process, not the stimulus. *Journal of Cognitive Neuroscience, 17*(3), 434–445. https://doi.org/10.1162/0898929053279531

Sacino, A., Cocchella, F., de Vita, G., Bracco, F., Rea, F., Sciutti, A., & Andrighetto, L. (2022). Human- or object-like? Cognitive anthropomorphism of humanoid robots. *PLoS One, 17*(7). https://doi.org/10.1371/journal.pone.0270787

Scandola, M., Cross, E. S., Caruana, N., & Tidoni, E. (2023). Body form modulates the prediction of human and artificial behaviour from gaze observation. *International Journal of Social Robotics, 15*(8), 1365–1385. https://doi.org/10.1007/s12369-022-00962-2

Scandola, M., & Tidoni, E. (2024). Reliability and feasibility of linear mixed models in fully crossed experimental designs. *Advances in Methods and Practices in Psychological Science, 7*(1). https://doi.org/10.1177/25152459231214454.

Schubert, T. W., & Otten, S. (2002). Overlap of self, ingroup, and outgroup: Pictorial measures of self-categorization. *Self and Identity, 1*(4), 353–376. https://doi.org/10.1080/152988602760328012

Stojanoski, B., & Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creating controlled scrambled images using diffeomorphic transformations. *Journal of Vision, 14*(12). https://doi.org/10.1167/14.12.6

Thorat, S., Proklova, D., & Peelen, M.v. (2019). The nature of the animacy organization in human ventral temporal cortex. *Elife, 8*. https://doi.org/10.7554/eLife.47142

Tidoni, E., Holle, H., Scandola, M., Schindler, I., Hill, L., & Cross, E. S. (2022). Human but not robotic gaze facilitates action prediction. *iScience, 25*(6), Article 104462. https://doi.org/10.1016/j.isci.2022.104462

Treisman, A., & Gormican, S. (1988). Feature analysis in early vision: Evidence from search asymmetries. *Psychological Review, 95*(1), 15–48. https://doi.org/10.1037/0033-295X.95.1.15

Wagemans, J., Elder, J. H., Kubovy, M., Palmer, S. E., Peterson, M. A., Singh, M., & Von Der Heydt, R. (2012). A century of gestalt psychology in visual perception: I. Perceptual grouping and figure–ground organization. *Psychological Bulletin, 138*(6), 1172–1217. https://doi.org/10.1037/a0029333

Wagemans, J., Feldman, J., Gepshtein, S., Kimchi, R., Pomerantz, J. R., Van Der Helm, P. A., & Van Leeuwen, C. (2012). A century of gestalt psychology in visual perception: II. Conceptual and theoretical foundations. *Psychological Bulletin, 138*(6), 1218–1252. https://doi.org/10.1037/a0029334

Welsh, T. N., McDougall, L., & Paulson, S. (2014). The personification of animals: Coding of human and nonhuman body parts based on posture and function. *Cognition, 132* (3), 398–415. https://doi.org/10.1016/j.cognition.2014.05.003

Welsh, T. N., Patel, S., Pathak, A., & Jovanov, K. (2023). "The clothes (and the face) make the Starman": Facial and clothing features shape self-other matching processes between human observers and a cartoon character. *Cognition, 230*, Article 105281. https://doi.org/10.1016/j.cognition.2022.105281

Wiggett, A. J., Pritchard, I. C., & Downing, P. E. (2009). Animate and inanimate objects in human visual cortex: Evidence for task-independent category effects. *Neuropsychologia, 47*(14), 3111–3117. https://doi.org/10.1016/j.neuropsychologia.2009.07.008

Willenbockel, V., Sadr, J., Fiset, D., Horne, G. O., Gosselin, F., & Tanaka, J. W. (2010). Controlling low-level image properties: The SHINE toolbox. *Behavior Research Methods, 42*(3), 671–684. https://doi.org/10.3758/BRM.42.3.671

Yeh, L.-C., & Peelen, M. V. (2022). The time course of categorical and perceptual similarity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance, 48*(10), 1069–1082. https://doi.org/10.1037/xhp0001034

Yovel, G., Pelc, T., & Lubetzky, I. (2010). It's all in your head: Why is the body inversion effect abolished for headless bodies? *Journal of Experimental Psychology: Human Perception and Performance, 36*(3), 759–767. https://doi.org/10.1037/a0017451

Zachariou, V., del Giacco, A. C., Ungerleider, L. G., & Yue, X. (2018). Bottom-up processing of curvilinear visual features is sufficient for animate/inanimate object categorization. *Journal of Vision, 18*(12). https://doi.org/10.1167/18.12.3

Zlotowski, J., & Bartneck, C. (2013). The inversion effect in HRI: Are robots perceived more like humans or objects? *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 365–372. https://doi.org/10.1109/HRI.2013.6483611