Shared and individual tuning curves for social vision

Rekha S. Varrier[1], Zishan Su[1], Qi Liang[1], Tory Benson[1], Eshin Jolly[1] and Emily S. Finn[1]

[1]Department of Psychological and Brain Sciences, Dartmouth College, Hanover 03755

Corresponding authors:

Rekha S. Varrier: rvarrier@uni-bonn.de and

Emily S. Finn: emily.s.finn@dartmouth.edu

Abstract

A stimulus with light is clearly visual; a stimulus with sound is clearly auditory. But what makes a stimulus "social", and how do judgments of socialness differ across people? Here, we characterize both group-level and individual thresholds for perceiving the presence and nature of a social interaction. We take advantage of the fact that humans are primed to see social interactions—e.g., chasing, playing, fighting—even in very un-lifelike stimuli such as animations of geometric shapes. Unlike prior work using these stimuli, we exploit their most advantageous property, which is that their visual features are fully parameterizable. We use this property to construct psychophysics-inspired "social tuning curves" for individual subjects. Social tuning curves are stable within individuals, unique across individuals, and show some relationship to socio-affective traits. Results support the view that social information processing begins early in the perceptual hierarchy. Further, our approach lays the foundation for a generative account of social perception in single subjects.

Introduction

A hallmark of the human species is our extraordinary sociality, which depends on reading and responding to others' behavior in ways that are largely effortless and shared across the population. Yet despite this shared general framework, there are substantial idiosyncrasies in how people perceive, interpret, and react to social information[1–5]. Many of these individual differences simply reflect variation in personality traits and social styles. Yet, extreme deviations from typical social processing are also central to developmental conditions such as autism[6–10], as well as mental illnesses such as schizophrenia[11,12], paranoia[13] and depression[14–17].

Social information can take many forms, including linguistic cues, facial cues, and whole-body motion cues. While humans get nuanced information from linguistic and facial cues, motion cues are necessary for some of our most basic evolutionary social behaviors that are conserved across species: e.g., pursuing, evading, playing, fighting, and courting[18,19]. Humans are primed to perceive these types of interactions even in very un-lifelike stimuli: for example, when faced with videos of simple geometric shapes moving around the screen, even without prompting, most neurotypical people will construct narratives to explain the shapes' movements in terms of goals, beliefs, and desires. This highly robust observation dates back at least to Heider and Simmel[20] and has since been leveraged to study social perception in a wide variety of contexts and populations[21–24]. The effect holds across cultures, suggesting a biological origin[18] . Along with related phenomena such as pareidolia[25]—the tendency to perceive faces in inanimate objects—this suggests an automaticity to social information processing that belies its typical conceptualization as a high-level cognitive process. Indeed, recent work supports the notion that core components of a social interaction can be extracted by the human visual system using fast, bottom-up processes[26], which is likely evolutionarily adaptive for a species that depends heavily on its sociality for survival[27,28].

While using simple geometric-shape animations as experimental stimuli has yielded important insights into behavioral, cognitive, and neural aspects of social perception, most past work using these stimuli has substantial limitations. Studies typically use a small number of manually generated animations handcrafted by human experimenters to be either obviously social or obviously non-social, with no systematic variation or control over visual features[3,29,30]. Subjects' responses are then classified as accurate or inaccurate with respect to these "ground truth" experimenter labels. Furthermore, even given a stimulus deemed "social" by most people, different individuals may be perceiving different *types* of social interactions in that same stimulus; the *nature* of the perceived interaction is rarely probed (and if it is, it is usually assumed to have a ground truth label —e.g., helping versus hindering—for which the experimenters again have clear "ground truth" labels in mind)[31–33]. Together, these practices often produce ceiling effects and compress individual variability in behavior (at least in normative populations), which is unrealistic given that most real-world social scenarios are complex and may engender different interpretations across people.

Here, we exploit a highly advantageous yet hitherto under-used property of such animations—namely, that they are algorithmically controllable and amenable to principles of visual psychophysics—to characterize people's socio-perceptual tendencies at both the group and individual level. We study two processes: (1) how people detect the presence of an interaction and (2) how people discriminate between types of interactions. By parametrically varying motion attributes[22], we programmatically generate a large set of animations and use participants' subjectively reported percepts to construct "social tuning curves" capturing shared trends and individual differences. Throughout, we adopt the perspective that socialness is in the eye of the beholder: in other words, there is no "correct" and "incorrect"; whatever percept is reported is the ground truth for that trial for that participant. We *embrace* ambiguous stimuli—i.e., those that yield high variability in reported percepts—as a feature rather than a bug, as these offer an opportunity to probe the limits of what makes a stimulus social, and how these limits differ for different individuals. We use this framework to show that robust individual differences in socio-perceptual tendencies exist atop group-level trends, that these individual differences are reliable over a period of months, and that they may be related (albeit likely in complex, nonlinear ways) to traits indexing real-world social and affective function.

Methods

Table 1: Summary of the experiments

<table>
<tr><th colspan="2">Experiment type</th><th>Social detection</th><th>Social discrimination</th></tr>
<tr><td colspan="2">Task</td><td>Presence vs. absence of social interactions</td><td>Valence discrimination: play vs. fight</td></tr>
<tr><td colspan="2">Parametrized motion attribute</td><td>Chase directness</td><td>Charge speed</td></tr>
<tr><td rowspan="4">Sample size</td><td>Pilot experiments (free-text responses)</td><td>N = 60</td><td>N =103, with cover story
N =102, without cover story</td></tr>
<tr><td>Main experiments</td><td>N = 312, session 1
(N = 240 returned for session 2)</td><td>N = 319, session 1
(N = 269 returned for session 2)</td></tr>
<tr><td>Supplementary experiments (controlled for correlated motion)</td><td>N = 308</td><td></td></tr>
<tr><td>Within-subject mixed-task design</td><td colspan="2">N = 279</td></tr>
</table>

Participants

All data collection and analysis procedures were approved by the Committee for the Protection of Human Subjects of Dartmouth College. All data were collected online (http://www.prolific.com/). We used the following selection criteria: Participants had to (1) be fluent in English, (2) have their location set as the USA or UK and (3) not have participated in our previous studies with similar stimuli. For consistency in data quality, all studies – except for the retest sessions where participants (identified using their 24-character Prolific IDs) were invited to complete a second session – were typically launched at 9am Eastern US time on Prolific and closed when the desired sample size was reached (usually about 2pm Eastern US time). The retest sessions were launched 2 months (Mean=71.0 days, SD=5.6 days; detection task) or 1 month (Mean=32.8 days, SD=7.7 days; discrimination task) after their respective first sessions and were left open for about 6 weeks until no participants signed up for the task in at least 1 week, with the goal to encourage as many participants as possible to return. Note that the aim of the second sessions was to test the retest reliability (i.e., how similar behavior was on two independent sessions), so the exact time gap between the sessions did not need to be the same, we only required that the session not be so close that we would see effects of perceptual learning or task-specific memory.

Stimuli

Stimuli were simple animations generated using a custom JavaScript-based software called *psyanim* (https://github.com/thefinnlab/psyanim-2). Each animation had two circular agents (radius 12 px): one black and one gray, set against a white background in a world of size 800px x 600px (see Fig 1a and all the stimuli here; exact stimuli for the individual experiments will be linked in-text near the description of that behavior). At the start of an animation, the agents were on either side of the center of the screen (coordinates: 400, 300): left (coordinates: 250, 300) and right (coordinates: 550, 300). In each experiment, the black agent started on the left (gray on the right) in half of the animations, and vice versa in the other half. The animations were 6s (detection task) or 8s (discrimination task) long with a frame rate of 60 Hz. A key difference between this study and most past work on social perception using animations is that we generated our animations purely programmatically using quantifiable, parameterizable motion attributes (Fig 1a). Each experiment consisted of 7 levels of stimuli where one motion attribute of interest varied linearly while all other attributes were held constant. These attributes were chosen such that people's perception of a social scene on a scale from the most non-social to most social (detection task) or from most playful to most aggressive (discrimination task) varied along the attribute of interest. In the following sub-sections, we describe these motion attributes as well as the animations in more detail.

Detection task

To manipulate percepts as to the presence versus absence of a social interaction (detection task), we relied on the motion attribute *chase directness*, which governs the fidelity with which one agent (the "predator") chases the other agent (the "prey"). This attribute was originally described by Gao et al.[22], where it was called "chase subtlety" and was found to robustly influence people's ability to detect social interactions (in particular, chases). Chase subtlety was originally defined in angles, i.e., by how many degrees a predator could deviate from a perfect heat-seeking path between itself and the prey at each time step. Thus, a *chase subtlety* of 0° would indicate a very direct chase, a *chase subtlety* of 90° would indicate a somewhat noisy chase where the predator can go off-path by up to 90° clockwise or 90° counterclockwise, and *chase subtleties* > 90° would indicate very noisy chasing behaviors where the predator can occasionally even move away from the prey. Here, we reversed and normalized the subtlety angle to derive *chase directness* ($\frac{180-chaseSubtlety}{180}$), such that the higher the directness, the more obvious (detectible) the chase. Details on how this was implemented in *psyanim* is below.

<u>*Chase animations*</u>

In the animations we generated, one agent is the predator and the other is the prey. Predator/prey assignment was counterbalanced across stimuli in terms of both start position (left/right of the center) and color (gray/black). The predator was programed to chase the prey at varying levels of *chase directness* and the prey was programmed to flee from the predator when it was within a certain radius. When the predator agent was beyond its field of view (or the distance between them was greater than the "*safety distance*"), the prey agent simply wandered around the screen. We included the wandering behavior so as to prevent the fleeing behavior from looking too obvious, so that people did not make decisions purely based on the prey. All variable attributes other than *chase directness* governing the motions of the predator and prey were held constant over all animations (i.e., over all levels of chase directness). The relevant attributes are described below:

Chase directness: Every 350ms (set by an attribute called *subtlety lag* explained in the next paragraph), the predator picks a value from a uniform distribution that ranges from [–*chaseSubtlety*, *chaseSubtlety*], where *chaseSubtlety* ∈ {0°, 30°, 60°, 90°, 120°,150°}. (i.e., *chase directness* ∈ {1, 0.833, 0.667, 0.5, 0.333, 0.167}). These behaviors are illustrated in Fig 1a. To get a feel for what these values mean, we encourage readers to watch some of the animations, available on GitHub (the stimuli used for session 1 chase detection experiments here, and those for session 2 experiments are here).

Other attributes: All attributes besides *chase directness* were set to be constant across animations. Some of the relevant attributes that influenced how the animations were perceived were optimized by manual piloting during stimulus development. These were, for the predator: (i) *maximum chase speed* = 1.5px/frame (frame rate = 60), (ii) *maximum chase acceleration* = 0.1px/frame^2 and (iii) *subtlety lag* = 350ms (how often the agent recomputes its chase direction; lower values will make the animation more jittery). For the prey: (i) *flee subtlety* = 30° (the angle which an agent can deviate from the true direction away from the predator — this parameter helps to avoid fleeing looking very obvious), (ii) *safety distance* = 100px (distance below which the agent flees from the predator; above this, it wanders), (iii) *maximum flee speed* = 1.8px/frame, (iv) *maximum flee acceleration*: 0.15px/frame^2, (v) *maximum wander speed* = 1.5px/frame, *maximum wander acceleration* = 0.1px/frame^2, (vi) *maximum seek speed* = 3.5px/frame and (viii) *maximum seek acceleration* = 0.05px/frame^2 (the seek behavior is included in the prey algorithm to keep it away from the boundaries/walls/edges of the world so that it does not get stuck in a corner). The flee speed and acceleration of the prey were set to be slightly higher than those of the predator to ensure that the predator never actually catches the prey.

"Invisible chase" control condition

With this control, we sought to rule out an alternative possibility for how *chase directness* might influence socialness perception that is less related to a chase *per se* and more related to general motion contingency between the two agents. Specifically, observers may notice that the predator and prey trajectories are linked more tightly in time at higher *chase directness* (where immediately after the prey changes direction, the predator too will change direction) than at lower *chase directness* (where the predator will not change direction as quickly and obviously upon the prey changing direction). Participants may simply be using this heuristic—i.e., whether the prey changing direction prompts the predator to change direction—instead of the actual chase between the predator and prey. To test for this possibility, inspired by Gao et al.[22], in a subset of behavioral experiments we included an additional set of control stimuli, where the actual prey was initialized at a randomly chosen location on the screen for each animation but was made invisible. The predator and a visible "mimicking" agent each started at one of the two regular starting locations (left and right of center as described at the start of the **Stimuli** section). The mimicking agent copies the true prey's trajectory but with a 180° rotation (i.e., if the invisible prey moves up and to the right, the mimicking agent will move down and to the left). The animations used for this study are here. For illustrative purposes, the invisible (true) prey is shown in yellow in these exemplars (participants never saw the yellow dots!). There were only 2 relevant parameters for the mimicking agent in *psyanim*: (i*) name or ID of the agent* to mimic (i.e., the invisible true prey), (ii) *angleOffset* = 180° (how much to offset the movement in angles).

Wander behavior

We also generated animations where both the agents were wandering independently, meaning that there was no programmed contingency between their motion patterns. The speed and acceleration of the two wandering agents matched that of the predator and prey agents in the main chase animations (pseudo-predator and pseudo-prey agents, respectively) to make these animations as close as possible to the chase animations. These animations serve as an additional check as to whether participants use the speed of an agent as a heuristic to help identify the predator (since, in chase animations, the predator always moved slightly slower than the prey so as to avoid catching it as described above). Although these animations were generated differently to the chase animations described above (where one of the agents is designed to chase the other, however inefficiently), they are conceptually equivalent to a chase with *chase directness* = 0 (subtlety 180°; i.e., where the predator is equally likely to move in any direction irrespective of the prey's position). Hence, we use these as our experimental stimuli for *chase directness* = 0 in the social detection experiments. These stimuli can be seen here.

The relevant parameters for these animations in *psyanim* were: (i) *maximum wander speed* (for the pseudo-predator agent = 1.5px/frame, pseudo-prey agent = 1.8px/frame) , (ii) maximum wander acceleration (for the pseudo-predator agent = 0.1px/frame^2, for the pseudo-prey agent = 0.15px/frame^2), (iii) maximum angle change per frame = 35° (how much the movement direction can change from one frame to the next), (iv) minimum screen boundary distance = 50px (the distance agents try to maintain from the screen boundary).

Discrimination task

To study how people discriminate between positive and negative social interactions, we varied the attribute *charge speed* in a novel social interaction scene that is different from the chase detection task discussed above. This scene was inspired by the presence of physical contact in the real world for common positive as well as negative interactions (e.g., hugs, high-fives, physical fights)[34] and how speed can be a clue to valence, with slower movements being usually perceived as more peaceful/positive, and faster movements as more aggressive/negative[19].

When generating these animations, both agents were set to have the same goal: to wander for a certain period and then charge at the other agent at the predetermined *charge speed* which varies between 1.5px/frame and 9px/frame (stimuli in Fig 1a and here). Once one agent initiates a charge, the other agent will respond after a short delay; following contact, both agents will return to wandering.

As with the detection task, several features were kept at constant values after extensive in-lab piloting. These were: (i) *minTargetDistanceForCharg*e = 200px, *maxTargetDistanceForCharge* = 500px (the between-agent distance range within which collisions would be initiated), (ii) *mean charge delay* =200ms, *jitter* =100ms (how long the second agent waits after the first agent charges at it), (iii) *mean break duration*: 2000ms, *jitter*=200ms (the duration for which two agents wander between consecutive charges), (iv) maximum wander speed: 1.5px/frame, (v) maximum wander acceleration=0.1px/frame^2, (iii) wander panic distance=800px (minimum distance at which a wandering agent will charge back at another agent).

Quality-checking stimuli

We quality checked all generated animations for both the detection and the discrimination tasks. Animations were manually checked by at least two lab members and were removed if they contained glitchy/flickery movement patterns, if the agents got stuck in the corners or stuck to each other for extended periods, and/or if one or both agents went offscreen. At all stages, bad animations were replaced with new ones of the same type (e.g., same *chase directness* value, predator color and start position) to obtain the targeted number of animations. For the detection task, our final stimulus set included 84 chase animations (12 animations at each of 7 *chase directness* levels, including animations generated via the wander algorithm which served as *chase directness*=0) and 84 ‘invisible chase’ animations (control; 12 animations at each of the *chase directness* levels used in the chase animations plus animations generated via the wander algorithm which served as *chase directness* = 0). Within these sets, the starting position (left versus right of center) and the role of predator versus prey was counterbalanced between the gray and black agent across animations, so that when participants were presented with an animation, there was no expectation of what they would see next based on the position or the color of the predator.

For *charge speed* (discrimination) stimuli, we removed bad animations according to similar criteria as described for the detection stimuli. In addition, we ensured that all animations in the final set had the same number of actual collisions (two), since differences in the number of collisions could have influenced percepts independently of *charge speed*, which was the main motion attribute of interest. The final set of discrimination stimuli contained 20 animations at each of 7 *charge speed* levels for a total of 140 animations. Within this set, the initial position of the gray and black agents was counterbalanced across animations.

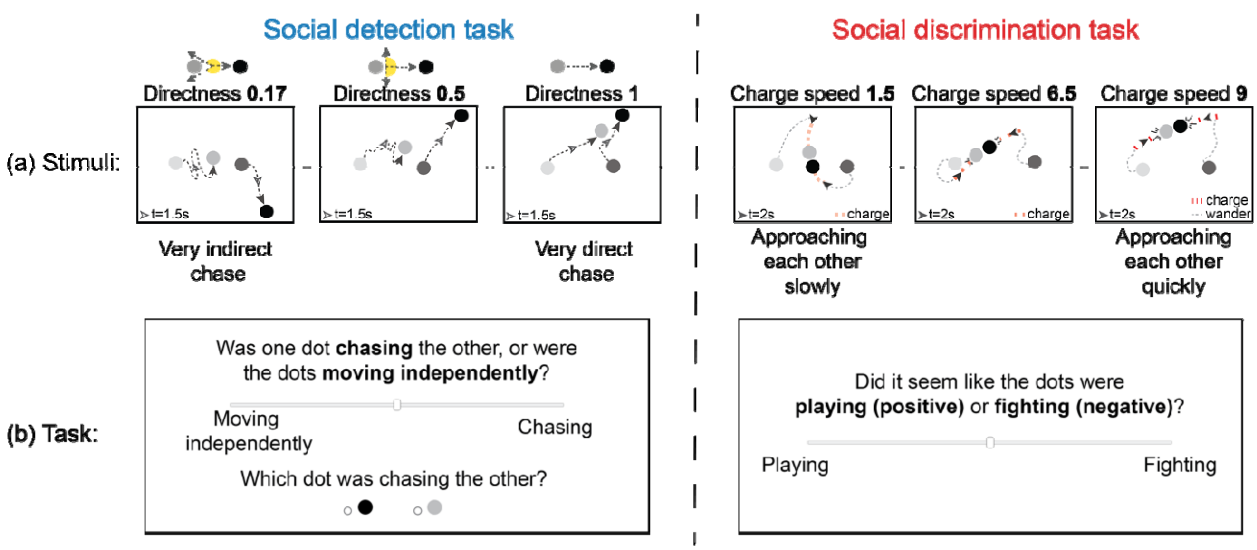

Fig 1: Social detection and discrimination tasks. (a) Static schematics of the animation stimuli for the detection (left) and discrimination (right) tasks to illustrate the effect of varying motion attributes (chase directness and charge speed, respectively) on agents’ trajectories. (b) The response screen that was presented following each animation in the main experiments. Both tasks required participants to rate their percept of the preceding animation on a continuous bar, and the detection task (left) additionally asked participants to identify the agent that was doing

the chasing. In both experiments, the positions of the slider labels ("Moving independently"/ "Chasing" or "Playing"/"Fighting") on the left versus right were counterbalanced across participants.

Animacy cover story

Attributing intentions to moving shapes entails judgments of two related yet distinct features – animacy and socialness. Animacy is a widely used concept that applies to entities that are considered alive[35]. These entities exhibit signs of self-propelled, non-Newtonian motion by seeming to engage in goal-directed behavior[36] and responding to their surroundings (e.g., changing speed or direction to avoid an obstacle). In our view, animacy is necessary but not sufficient for socialness: animacy can be detected in displays of single agents, in which by definition there is no social interaction present, but in multi-agent displays, to the extent that agents are perceived to be socially interacting, they must also be perceived as animate (i.e., as possessing a mind that would be motivated to engage in social behavior). Our goal here was to isolate the concept of *socialness* above and beyond animacy. Because differences in percepts of animacy might confound judgments of socialness, to encourage uniform perception of animacy across both animations and participants to the extent possible, we provided a cover story that the agents (dots) represented children in a public park. This was the exact story participants received: "*We recently videotaped a public park where nearby children go, with the goal of capturing the essence of children's behaviors within a familiar park setting. To protect the identities of these young people, we used an algorithm that represents* ***a pair of children as two dots****, each tracing the path of an individual child.*" We used this cover story to set the context in all of our experiments except in the early pilot experiments, since the goal of the latter was to evaluate how participants spontaneously interpret the animations even without any context.

Pilot experiments (open-ended responses)

All pilot and main experiments were programmed and run using the *jsPsych* platform[37] with a custom plugin (https://github.com/thefinnlab/psyanim-2) to present the *psyanim* animations.

Pilot experiment design

We first conducted a set of small-scale studies where participants could freely describe the animations. Before imposing explicit, constrained rating scales, our goal was to verify that these animations do in fact spontaneously evoke percepts that fall approximately along the intended axes from non-social to social (detection task) or playful to aggressive (discrimination task).

In these experiments, each participant was presented with 7 animations (1 animation for each level of the motion attribute as described under **Stimuli**). For the detection task, we did not use the cover story. For the discrimination task we ran two versions – one without and one with the cover story. In what follows, unless otherwise noted, we present data from the version *with* the cover story. After watching each animation, participants responded to the following prompt to indicate what the dots could have represented: "Briefly describe what the dots were doing. Guess if you do not know." (In the discrimination task with the cover story, the text varied slightly: "Describe what the dots were doing using a word or a short phrase"). The task lasted ~5-10 min overall.

Pilot experiment data analysis

We analyzed the free-response data using techniques from natural language processing. In each experiment (detection and discrimination tasks), we derived the average "meaning" of descriptions at each stimulus level using semantic embeddings. Specifically, we used Bidirectional Encoder Representations from Transformers (BERT)[38] language models as implemented in the Python library *SentenceTransformers* (https://huggingface.co/sentence-transformers/all-MiniLM-L6-v2). For each description, we get a 384-dimensional vector embedding. We then averaged across all embeddings at each stimulus level (detection task: 12 unique stimuli per level of *chase directness* x 5 observers per stimulus = 60 observations per level of *chase directness*; discrimination task: 20 stimuli per level of *charge speed* x 5 observers per stimulus = 100 observations per level of *charge speed*).

In an initial exploratory/data-driven analysis, we compared this mean vector to the embeddings of *all* 8432 English verbs from the natural language toolkit (nltk; https://www.nltk.org/howto/wordnet.html) to identify the 5 verbs whose embeddings it was closest to. To more clearly isolate the *differences* in percepts across levels, we then removed words that appeared in at least 6 of the 7 motion attribute levels within each experiment.

In a follow-up, more hypothesis-driven analysis, we quantified the change in percepts across stimulus levels by computing the similarity of mean embeddings at each level to our expected percepts at either ends of the response scale (detection task: "chasing", "moving independently"; discrimination task: "playing", "fighting"). We expected that, as the motion attribute value increased, descriptions' similarity to one extreme ("chasing" or "fighting") would increase and similarity to the other extreme ("moving independently" or "playing") would decrease. To quantify this, we took the difference between embeddings' similarity scores to both extrema ($difference\_score = score\_chasing - score\_moving\_independently$ for the detection task and $difference\_score = score\_fighting - score\_playing$ for the discrimination task), giving us one difference_score per trial. Later, scores were compared using a linear mixed effects model (LME; *pymer4* package[39]): $difference\_score \sim motion\_attribute + (1|subID)$, where for the detection and discrimination tasks, $motion\_attribute$ referred to *chase directness* and *charge speed*, respectively.

Finally, for the discrimination task only, we quantified how the valence, or "sentiment", of the descriptions varied across attribute levels. We used a RoBERTa-base model[40] (https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment-latest/tree/main) to automatically quantify sentiment, which yields a positive, negative, and neutral score for each description. Our dependent variable $difference\_score$ was the difference between the negative and positive sentiment scores for each description. Similar to the approach to semantic similarity described in the previous analysis, the effect of the motion attribute (namely, *charge speed*) on these values was computed using the LME: $difference\_score \sim charge\ speed + (1|subID) + (1|stimID)$.

We noted that free-text descriptions were overall biased toward positive sentiment (Fig S1c, left). This could likely be because of our cover story about the dots representing children in a park, which carries a strong prior toward playful interactions. To check this, we ran an additional small pilot batch without a cover story too (all else remained the same).

Results from the analyses of free text responses in these pilot experiments showed evidence favoring our hypotheses that (1) in the detection study, as *chase directness* increases, animations are seen as more social, and (2) in the discrimination study, as *charge speed* increases, interactions are seen as more aggressive and negatively valenced. Together, the pilot experiments confirmed that the stimuli we generated algorithmically could spontaneously—i.e., without prompting with explicit choices of possible interactions—evoke percepts along the intended axes, and this gave us the confidence to move forward with our main experiments using these axes to structure responses, described in the next section.

Main experiments

Main experiment design

Our first two main experiments consisted of only detection or only discrimination trials, respectively, while the third main experiment was a mixed-task design in which the same participants performed both the detection and discrimination tasks. We used the same cover story described above (about the dots being children in a public park) in all the main experiments.

For the detection study, there were 84 trials per participant (12 trials per *chase directness* level x 7 levels) with 7 optional breaks (one every 12 trials). In the supplementary experiments that also included the invisible chase control condition, there were 6 stimuli at each motion attribute level (2 conditions x 7 stimulus levels x 6 trials per level). After each trial, participants gave two responses: (1) they rated, on a continuous scale, to what degree one of the two dots was chasing the other versus moving independently (with the location of the labels on the left versus right extremes of the scale kept constant within participant, but counterbalanced across participants), (2) they identified the dot that was chasing the other, in a two-alternative forced choice. Both questions were presented on the same page, and the page timed out in 10s. The instruction specific to this task (after the cover story was presented) was as follows: "*Some videos depict a situation in which* ***one dot is chasing the other****; other videos depict a situation in which the dots are* ***moving independently****. You will be asked to* ***rate how much you think the dots are interacting (meaning one dot is chasing the other) versus moving independently****. For all videos, you will also be asked to determine* ***which dot was chasing the other****. If there was no chase in a particular video, just make a guess.*". Participants performed one practice trial before the experiment began.

For the discrimination study, there were 70 trials per participant (10 trials per *charge speed* level x 7 levels) with 7 self-timed breaks (one every 10 trials). After each trial, participants rated on a continuous scale to what degree the dots were engaged in a positive (playing) versus negative (fighting) interaction (as with the detection task, the label positions were counterbalanced across participants). This response page also timed out in 10s. The instruction specific to this task (after the cover story was presented) was as follows: "*Some videos depict a situation in which the children are engaged in a* ***positive interaction (e.g., playing)****; other videos depict a situation in which the children are engaged in a* ***negative interaction (e.g., fighting)****. After each video, you will be asked to* ***rate to what extent the children (dots) were engaged in a positive or negative interaction*** *on a continuous bar. There are no right or wrong answers here, so if you are unsure, just guess!*". Participants performed two practice trials before the experiment.

In the third main experiment (mixed-task design study), detection and discrimination animations were presented in six interleaved blocks: block sequence with 14 trials each (121212 or 212121, where 1=detection task and 2=discrimination task; one of the two block sequences was randomly selected for each participant. Here too, there were 84 trials in total: 42 detection trials (6 at each level of *chase directness*) and 42 discrimination trials (6 at each level of *charge speed*). The 42 animations from each task (detection or discrimination) were randomized *across* the 3 blocks of the task. Participants performed two practice trials each for the detection and discrimination experiments at the start of the experiment.

The primary task portion of all three experiments (detection, discrimination, or mixed-task design) lasted ~15-20 min. At the end of the primary task portion, we presented participants with the trait questionnaires described below under **Trait measures**. The sequence of the questionnaires was counterbalanced across participants.

Quality checks during data acquisition. We performed a few data quality checks during data acquisition to exclude poor participants within the first few minutes of the study. First, after the instructions, including the cover story about the dots representing children in a park, but before the start of the main experiment, we presented participants with a multiple-choice question as to what the dots represented. The options were "animals", "balls", "adults", "children", "magnets". The correct answer was "children" (as mentioned clearly in the cover story). If participants responded incorrectly, they were given one chance to correct their answer, and if their second response was also incorrect, they were immediately excluded from the study. (Note that this same question was asked again *after* the main task and used for a second quality-check analysis, see below). Participants were also warned that they may not be compensated if they missed (i.e., timed out on) more than 10% of trials. We also excluded participants who opened other tabs or had bad internet connections by including a demo animation on the very first page and checking playback duration in real time. If the duration of this page was much higher than 6 or 8s (actual duration of the animations), this meant that the animation did not play as normal, and either paused (because of other open tabs and the participant not paying attention) or played very slowly (possibly because of a slow internet connection). Participants who stayed on the demo animation page for longer than a liberal threshold of 20s were immediately excluded from the study. Besides these online quality checks *during* the experiments, we conducted further quality checks at the data analysis stage to exclude poor participants.

Main experiment analysis

Data analysis was similar across all main experiments unless specified otherwise.

Exclusion criteria

First, we excluded participants with bad or unreliable data, as defined by meeting one or more of the following criteria: (i) missing responses (i.e., timing out) on more than 5% of all trials; (ii) incorrect responses in the post-main-experiment debrief question asking them to identify what the dots represented shown (note that this question is identical to the question asked at the beginning of the main experiment, but the rationale here is that if by the end of the main experiment participants had forgotten what the dots represented, their perception and ratings could have been affected by whatever they assumed the dots to represent by the end); (iii) (for the detection task alone) incorrectly identifying the predator in more than one third of animations with directness=1 (rationale: the chase/predator identity is very obvious in these animations, so incorrect answers here are most likely failures of attention); (iv) lingering on the animation page for more than 20s in at least 5% of trials (each animation was only programmed to last 6 or 8s, and so the page should have lasted for a similar duration; any longer indicates that they may have clicked away from the experiment tab and/or had a slow internet connection); (v) failing to respond to $\geq$ 10% of items on one or more trait questionnaires; or (vi) missing at least one (out of five) attention-check items in the trait questionnaires. Trait questionnaires are described in detail in the **Trait measures** section.

Separate detection and discrimination experiments

Group-level analyses: We first analyzed data at the group level to ascertain shared tendencies in how motion attributes affect social percepts. Participants' ratings were coded on a 0–1 scale: (i) detection task: *moving independently* = 0, *chasing* = 1; (ii) discrimination task: *playing* = 0 and *fighting* = 1. For the detection task, we used the response to the predator identification question to compute accuracy (0 or 1 on each trial). We used linear mixed-effects analyses to quantify the effect of each motion attribute while controlling for confounding variables using the following model: $rating \sim motion\_attribute\_level\ +\ mean\_dist\ +\ trial\_number\ +\ (1|subID)\ +\ (1|stimID)$. Here $rating$ refers to participant responses indicating the level of $socialness$ (degree of *chasing*) or $aggressiveness$ (degree of *fighting*). The term $motion\_attribute\_level$ refers to degree of either $chase\ directness$ (detection) or $charge\ speed$ (discrimination) and could take one of 7 levels. Additional terms are: (i) $mean\_dist$, the distance between the two agents averaged across all frames; we included this term because past work has shown that agents that are closer together are more likely to be perceived as interacting[6]; (ii) trial number, indicating serial order over the course of the experiment (to check for any drift in ratings over time); and random-effects terms for (iii) subject identity and (iv) specific animation identity (exemplar). For the predator-identification question in the detection task, we ran a logistic regression model with $accuracy$ (0/1) as the dependent variable and the same main- and random-effects predictor terms as above.

Individual-level analyses: We next analyzed data at the individual level to determine the extent to which participants differed in their socio-perceptual tendencies, how robust these differences were across sessions, and how detection and discrimination tendencies covary with one another and with other socio-affective traits. Our primary approach to analyzing individual-level data was to compute single-subject "tuning curves" for detection and/or discrimination behavior. We averaged each participant's responses at each motion attribute level (*chase directness* or c*harge speed* level for detection and discrimination tasks, respectively). For each participant, we plotted motion attribute level (normalized to a 0-1 range; x-axis) against average rating across animations at that level (y-axis). Similar to the group-level results shown in Fig 3, visual inspection suggested that individual detection ratings followed a sigmoid shape, while discrimination ratings followed a more linear trend. We empirically tested both sigmoid and linear fits to evaluate which one better fit each type of data and verified this pattern: the Akaike Information Criterion (AIC) – which indicates which of two models fits the data better (lower AIC indicates better fits) – was lower for the sigmoid fits in the detection task (mean difference sigmoid – linear fits ≤ -7.84, $p < .001$ based on paired t-test in all the detection experiments) and higher for the sigmoid fits in the discrimination task (mean difference sigmoid – linear fits ≥ 3.47, $p < .001$ in all the discrimination experiments, Fig S2). Hence, we used sigmoid fits to characterize each participant's response data in the detection task and linear fits to characterize responses in the discrimination task.

The sigmoid curve-fitting equation was : $S(x) = \gamma + (1 - \gamma - \lambda) \frac{1}{1-e^{-\frac{(x-\alpha)}{\beta}}}$, where x = the value of the motion attribute (*chase directness* or *charge speed*); γ and λ = the lower/upper asymptote of the curve; α, β = center, slope. The linear equation was: $L(x) = c + m * x$, where x = the motion attribute (*chase directness* or *charge speed*); c = the lower intercept of the line; m = slope. For both functions, we calculated several key parameters from the fitted curve of each participant:

Shifts from extremes: Lower bias lb and upper bias ub (lower and upper intercepts, respectively) reflect ratings at the lowest and highest levels of the motion attribute — in other words, how close to the extremes of the rating scale a participant is willing to go. For linear fits, lb = intercept c.

Bias: This parameter was derived from the above-mentioned bias terms (lb and ub) as a comprehensive summary of people's bias that also factors in apparent biases due to differences in overall confidence or perceptual vividness. The bias term thus measures to what degree people avoid the lowest end of the scale relative to how much they avoid the two ends of the scale in general (which could reflect lower confidence overall or a less intense effect of stimuli overall). We quantified this as $\frac{lb}{lb+ub}$. A bias of 0.5 means that there is no bias towards one end of the scale, bias < 0.5 means that people are more biased towards the lower end, and bias > 0.5 means that people are more biased towards the upper end. In the detection task, bias > 0.5 can be interpreted as a predisposition to see things as social ("chasing") more than non-social ("moving independently"); in the discrimination task, bias > 0.5 can be interpreted as a predisposition toward seeing things as more like "fighting" than "playing".

Midpoints: Two types of midpoints can be derived from each curve: an objective midpoint ($x_{obj} = f^{-1}(S(x) = 0.5)$) and a subjective midpoint (subj_center, α). The objective midpoint is the motion attribute value at which the participant's rating crosses the absolute midpoint of the rating scale (0.5 for all participants). The subjective midpoint is the motion attribute value at which the participant's rating crosses the halfway point of *their own* behavioral curve (e.g., for a participant whose ratings vary between from 0.4 and 0.8, their mid-point would be the motion attribute level at which the tuning curve crosses 0.6). Here we mostly focus here on the objective midpoint, which we call the point of subjective equivalence (**PSE**) because it is similar in spirit to PSE as defined in traditional visual psychophysics work (i.e., the stimulus feature level at which options A and B are equally likely in a two-alternative forced-choice discrimination task [41]).

Range: The distance between the lowest and highest ratings on the Y-axis. This parameter can be interpreted in multiple ways: it quantifies how much of the total response scale participants use, how much participants differentiate between stimuli at extreme (lowest and highest end) motion attribute levels, and except in cases of strong response biases, might reflect how confident people are in their percepts (especially at the lowest and highest motion attribute values). This is defined as $1 - lb - ub$.

Sigma (σ; sigmoid fit only): Sigma or the inverse of the slope ($\frac{1}{\beta}$) determines the steepness of the sigmoid curve during its transition from perceiving something as "moving independently" (at lower *chase directness* levels) to "chasing" (at higher *chase directness* levels) in the detection task. A smaller sigma (or a higher slope) indicates that participants perceive something as more social (Δ_{rating}) with the smallest change in the sensory evidence (change in *chase directness* or $\Delta_{\text{chase directness}}$) — this can be interpreted as people exhibiting higher confidence when rating the intermediate, ambiguous stimuli. A higher sigma (or lower slope) indicates that participants need a lot more sensory evidence ($\Delta_{\text{chase directness}}$) to rate something as more social (Δ_{rating}) — this may reflect lower confidence and/or a more gradual evidence-based shift in the perceptual intensity of stimuli at the ambiguous middle levels.

We plotted covariance matrices (Pearson *r*) between the various parameters within the detection and discrimination tasks. Based on these covariances as well as test-retest reliability of the curve fit parameters (described in detail below), in what follows, we focus on three main curve-fit parameters that are both relatively reliable and not highly collinear with one another: PSE, range and bias. These terms are illustrated in Fig 2. Since PSE covaries tightly with range and bias in the linear fits specifically and it is not possible to change PSE without changing either the range or the bias, we only use range and bias for the discrimination task.

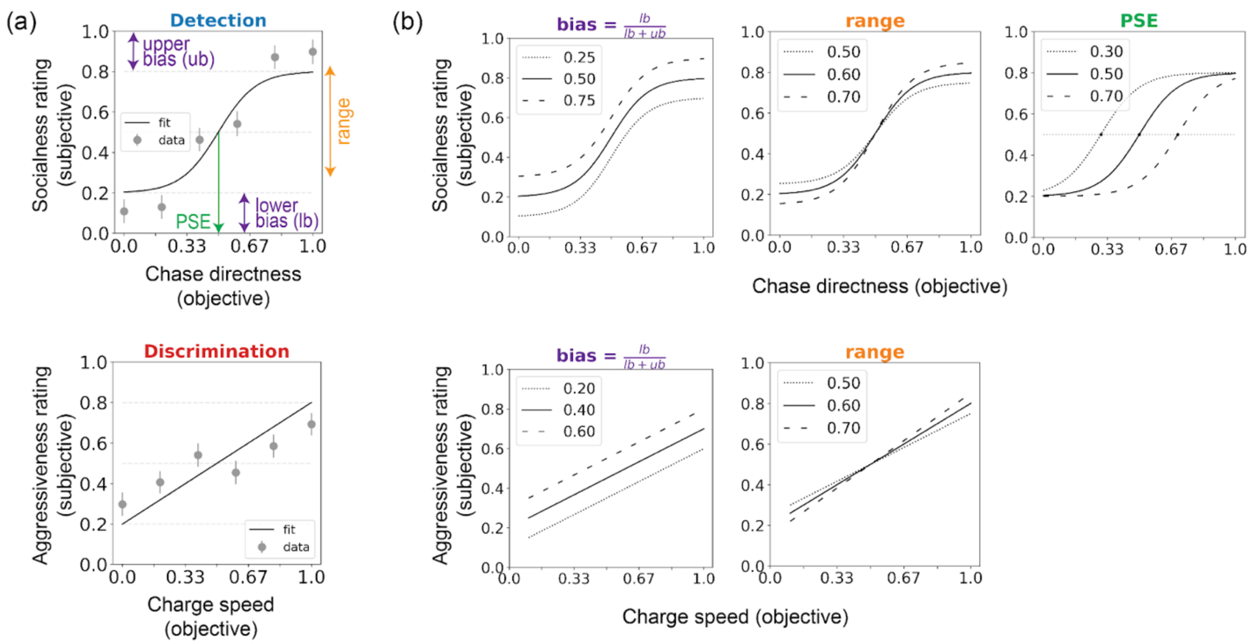

Fig 2: Fitting social tuning curves to individuals' reported percepts. (a) We fit sigmoid (detection experiments; top) or linear (discrimination experiments; bottom) curves to individual-participant rating data and used the resulting curve parameters to characterize individual participants. (b) Schematics of how each parameter can vary across participants. The upper rows show sigmoid fits as used for the detection task, and the lower rows show the linear fits used for the discrimination task. Each plot shows curves for three different hypothetical participants.

We tested the robustness of each person's tuning curve by using parameters calculated on data from their first session to fit the same participant's data in their second session 1–2 months later (and vice versa) and calculating the residual normalized root-mean squared error (NRMSE) between the curve and the true data points. We used a paired t-test to compare the NRMSE from this within-participant fit to the NRMSE from the mean across-participant fit (calculated by averaging the curve-fit parameter values across all other participants in the same session. This quantified the extent to which an individual's tuning curve offered a better prediction of their own held-out data than generic tuning curves based on group data. Further, we calculated the intra-class correlation coefficient (ICC) of each curve-fit parameter of interest to measure test-retest reliability across the two sessions (ICC2, single random raters, as implemented by the Python package *pingouin*).

Mixed-task experiments

In these experiments, the same participants performed both the detection and discrimination tasks. Data extraction and curve-fitting was performed similarly to the separate detection and discrimination experiments described above. For each participant, we obtained tuning curves for both detection and discrimination. We studied how socio-perceptual tendencies in the two tasks relate by calculating the Pearson correlation coefficient between all possible pairs of curve-fit parameters across the detection and discrimination tasks. Here, we restricted our analyses to only the most robust curve-fit parameters: bias, range and PSE for the detection task, and bias and range for the discrimination task (5 total).

Traits were scored as described under **Trait measures**, giving us 14 dimensions in total (5 for AQ, 2 for PANAS, 5 for NEO-FFI, 1 for loneliness and 1 for number of friends). As a first-level exploratory analysis, we first performed trait-behavior correlations (Pearson r) between each trait dimension and curve fit parameter (14 x 5 = 70 total correlations). We report both uncorrected results and results after correcting for multiple comparisons using the Benjamini-Hochberg procedure.

We next performed inter-subject representational similarity analysis (IS-RSA)[42] to test the second-order hypothesis that pairs of subjects who are more similar in their curve-fit parameters are also more similar in their pattern of trait scores. Specifically, we computed a set of subject-by-subject representational dissimilarity matrices (RDMs) reflecting the Euclidean distance between each pair of subjects': (i) pattern of trait scores (14x1 vectors), (ii) detection tuning curve parameters (3x1 vectors), (iii) discrimination tuning curve parameters (2x1 vectors), or (iv) combined tuning curve parameters across both tasks (5x1 vectors). We then correlated (Pearson r) the vectorized upper triangles of the trait RDM with each of the three curve-parameter RDMs. We assessed the significance of each IS-RSA r value non-parametrically using a Mantel test, in which subject-vector assignment is randomly shuffled in one of the two RDMs

5,000 times to generate a null distribution of *r* values expected by chance. The true *r* value is compared against this null distribution to derive a *p* value using the formula:

$$p = \frac{(\#\ \text{permutations exceeding true value} + 1)}{(\#\ \text{permutations} + 1)}.$$

Trait measures

We chose individual-difference measures of interest based on past findings that behavior on social perception and cognition tasks often differs between populations (e.g., neurotypical versus autistic, patients with depression versus healthy controls) and/or covaries in the normative population with socio-affective and personality traits. Past work has shown that people higher on autism-like phenotypes are less likely to detect intentions and interactions in social animation displays[6–10], while people with higher internalizing symptom scores (related to anxiety and social withdrawal) and a higher desire for social connection are *more* likely to detect intentions and interactions[5,43–46]. Other studies have associated depression with impaired emotion recognition of social stimuli (hypersensitivity to negative cues, hyposensitivity to positive cues)[14–17]. We assessed autism-like traits with the autism quotient (AQ) questionnaire[47], loneliness with the UCLA loneliness scale[48], and general affect with the Positive and Negative Affect Schedule (PANAS)[49]. We also administered the NEO five-factor inventory for multidimensional personality (NEO-FFI; more popularly known as the “Big five”)[50]. Finally, we asked participants to state the number of close friends they had, since this metric provides additional information about participants’ real-world social tendencies.

Our final battery thus consisted of five entities: (i) AQ, (ii) PANAS, (iii) NEO-FFI, (iv) UCLA loneliness scale and (v) the self-reported number of friends. Details of each questionnaire are given below.

The AQ consists of 50 total items measuring five subdomains: social skill deficits, communication deficits, attention-switching deficits, heightened attention to details and imagination deficits. For each item, participants had four response choices ("Definitely disagree", "Slightly disagree", "Slightly agree", "Definitely agree"). We reverse-scored the items that were intended to be as per the instructions from the creators; however, when assigning a score on each item, we assigned responses scores between 0 and 3 (where 3 → less neurotypical and more autistic) in place of binarizing responses (assigning 0 to the first two levels and 1 to the last two levels). Higher scores on each subdomain indicate more autism-like phenotypes (e.g., greater social skill deficits).

PANAS consists of 20 total items, 10 measuring positive affect and 10 measuring negative affect. Participants responded on a five-point scale ("Very slightly or not at all ", "A little", "Moderately", "Quite a bit", "Extremely"). Each response was coded between 0 and 4, and there were no reverse-scored items. This scale results in separate scores for positive and negative affect.

NEO-FFI consists of 60 total items measuring five dimensions: openness, extraversion, neuroticism, conscientiousness, agreeableness. Participants responded on a five-point scale ("Strongly disagree", "Disagree", "Neutral", "Agree", "Strongly agree") coded between 1 and 5. This scale yields a summary score for each of the five dimensions.

The UCLA loneliness scale consists of 20 total items that measure a single dimension. Participants responded on a 4-point scale ("Never", "Rarely", "Sometimes", "Always") scored between 1 and 4. Higher scores indicate higher loneliness. Lastly, participants also responded to the following question with an integer value: "*Please estimate the number of **close friends** that you have, where "close friends" are people that you feel at ease with and can talk to about private matters.*"

Within each questionnaire we also added one attention check question (e.g., for AQ, the question was this: "If you are doing your best to complete this survey honestly, choose 'Definitely agree'.") to confirm that participants were paying attention to the questions. The accuracy on these questions was used as a quality check criterion during data pre-processing (described under *Main experiment analysis*).

Code and data availability

All stimuli, data and code will be available upon publication at: https://github.com/thefinnlab/psyanim_behav_paper1

Results

We systematically studied how people perceive the presence and nature of a social interaction using sets of algorithmically generated fully parameterized animations. Each animation consisted of two agents—one gray and one black circle—that were programmed to move in certain ways with respect to one another. Critically, the animations varied parametrically along one motion attribute and were controlled for all other low-level visual features. We conducted three sets of experiments: (1) detection studies, in which participants rated the extent to which the agents appeared to be interacting (i.e., one agent chasing another) versus moving independently, (2) discrimination studies, in which participants rated the extent to which the agents appeared to be fighting versus playing, and (3) a mixed-task study where participants performed both the detection and discrimination experiments (see stimuli here and details in Fig 1). Below, we describe group-level and individual behavioral patterns for both social detection and discrimination, as well as how these two behaviors compare to one other and to self-reported social and affective traits.

Simple motion attributes influence the detection of both the presence and nature of social interactions

Varying simple motion attributes reliably affects social percepts at the group level

For detection experiments, we varied the attribute *chase directness*, which determines the fidelity with which the movement of one agent (the predator) is contingent on the other agent (the prey): more direct chases should look more obviously social, less direct chases should look more like agents moving independently. For discrimination experiments, in which the two agents come together and move apart in succession, we varied the attribute *charge speed*, which determines how fast the agents approach one another: slower should look more playful, faster should look more aggressive/fight-like.

Pilot experiments

We first performed pilot experiments to test that these animations could spontaneously evoke percepts along the intended continua without any explicit prompting. In these experiments, participants watched animations and gave free-response text descriptions, which we quantified using tools from natural language processing. Indeed, we found that varying *chase directness* elicited percepts along a continuum from moving independently (non-social) to chasing (social), while varying *charge speed* elicited percepts along a continuum from positive/playing to negative/fighting (see **Supplementary Results** and Fig S1). These results gave us confidence to move forward to our main experiments, in which we replaced free responses with these continua as predetermined rating scales.

Detection experiments

In our main social detection experiments (see Table 1 for sample sizes and other relevant information), after watching each animation, participants (1) rated how social it was on a continuous scale ranging from "moving independently" to "chasing" (henceforth referred to as "socialness rating"), and (2) identified the predator agent by color. In line with our expectations and past work [22], we found that as chase directness increased, ratings shifted towards "chasing" ($b=0.805$, $p<.001$; Fig 3a). Social perception did not seem to change with time (indexed as the trial number; $b=–0.001$, $p=.779$), suggesting that there was no measurable "drift" in ratings toward more social or more non-social over the course of the experiment. We also observed that, similar to subjective ratings, predator identification accuracy increased as chases became more direct (logistic regression $b=4.735$, $p<.001$; Fig S3a); this result further confirmed that participants were, on average, experiencing the chase in the expected way based on the generating algorithm (i.e., they correctly perceived its directionality, especially in the case of the more direct chases).

Participants might have been using visual features other than *chase directness* to form their judgments of socialness. For example, past work has shown that agents that are closer together are more likely to be perceived as interacting[6]. That more-direct chases also resulted in a narrowing of the distance between the two agents over time was an unavoidable consequence of our animation-generation algorithm (where the predator was programmed to chase after the prey); indeed, in our stimulus set, chase directness and mean distance between agents over the course of the animation were correlated across animations (Pearson r=–0.69, p<.001). Hence, to account for other visual features beyond *chase directness* that participants may have been using, we ran additional models including mean distance between agents as a covariate. While this term was also a significant predictor of socialness ratings (b=–0.432, p<.001) and this model fit the data better (AIC=–12325 compared to the AIC of the model without mean distance, –12293; lower AIC indicates a better fit), *chase directness* captured additional unique variance in socialness percepts (b=0.603, p<.001). Further, mean distance was not a predictor of the predator identification accuracy (b=–0.716, p=.179) and also did not meaningfully improve the model fit for accuracy (AIC without and with mean distance = –7986 and –7984, respectively).

Lastly, participants may have been relying on other heuristics to form socialness judgments. Since a chase typically results in correlated movement patterns between two agents (when the prey changes direction, the predator is likely to do so a moment later), this correlation will be stronger for more direct chases than less direct chases. However, agents can also show correlated motion without necessarily having one chase the other—i.e., one agent could change direction every time the other one does but be equally likely to turn away from (or orthogonal to) the path of the other. To test whether participants are simply relying on nonspecific correlated motion as a heuristic, also inspired by Gao et al.[22], we ran a separate experiment in an independent set of participants where we replaced half of the directness 0.167 to 1 chases (6 levels) with a non-social "invisible chase" control. In these trials, the predator was chasing a true prey agent that was made invisible to observers, while a visible "fake" prey mimicked the true prey's trajectory reflected over a 180° rotation. In this way, correlated motion between the two agents was preserved—when the true prey changed direction, so did both visible agents (the predator and the mimicking agent)—but not necessarily in a manner consistent with chasing. In line with our prediction, we saw that while in the true chase condition, socialness ratings increased with *chase directness* (b=0.471, p<.001), in the invisible chase condition, if anything, there was a slight trend in the opposite direction (socialness ratings decreased as directness increased; b=–0.146, p<.001; Fig S3b). This suggests that the increase in socialness with chase directness is not merely because of correlated motion in general, but rather correlated motion that is specifically consistent with pursuit behavior. In sum, in line with Gao et al.[22], these experiments show that the motion attribute *chase directness* influences how social stimuli are perceived to be at the group level.

Discrimination task

Once a social interaction is detected, the next step is to discriminate the *nature* of that interaction; past work[20,30] and real-life experience suggest that given the same sensory input, there may be even more variability across people in their percepts of *how* agents are interacting than simply *if* they are interacting. Interactions can be characterized along several dimensions, but a fundamental one is valence—i.e., how positive or negative is the interaction?[51] Using our parametric approach, we explored changes in the valence of social percepts between "playing" and “fighting”. Playing and fighting, which are preserved throughout much of the animal kingdom, involve two agents moving apart and coming back together in quick succession. We manipulated the speed with which our agents approached each other (*charge speed*) to determine whether this simple motion cue could reliably affect percepts of an interaction’s valence, with slower speeds looking more friendly (like “playing”) and higher speeds looking more aggressive (like “fighting”).

In these experiments (see Table 1), participants watched each animation and rated it on a continuous scale from “playing” to “fighting”. We found that as the *charge speed* increased, interactions were perceived as more aggressive ($b = 0.622$, $p < .001$; Fig 3b). The effect of *charge speed* persisted when controlling for the mean distance between the two agents (which also affected ratings such that higher mean distances predicted slightly less aggressive ratings; $b=-0.057$, $p=.008$)) and trial number as an index of time (for which we found that percepts of aggressiveness also weakly increased over the course of the experiment; $b=0.024$, $p<.001$).

Together, results indicate that social percepts—both the presence and nature of an interaction—can be manipulated by simple motion attributes in ways that are generally shared across people. Despite these commonalities, to what extent are there stable and meaningful individual differences in these socio-perceptual tendencies? We probe this question next.

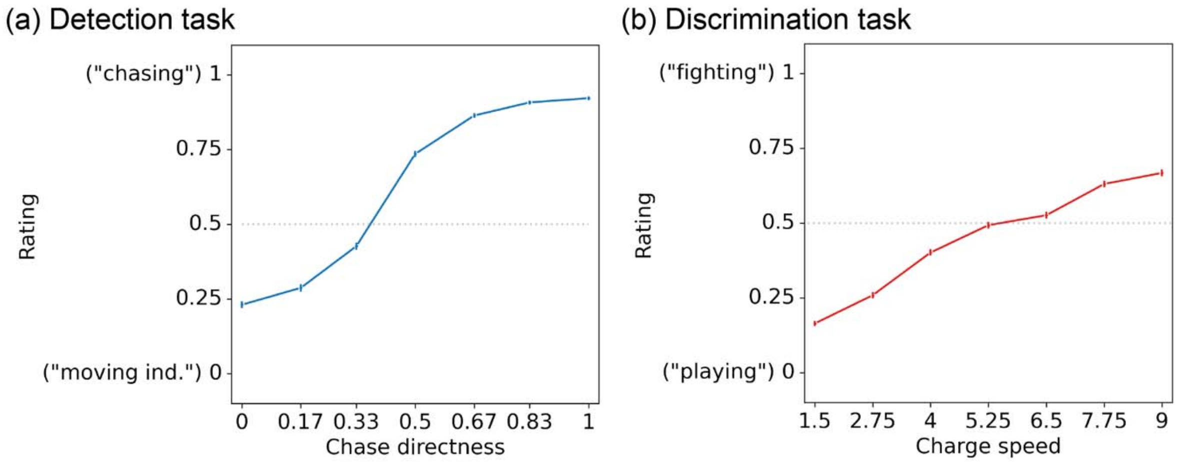

Fig 3: Group-level behavior on the detection and discrimination tasks. (a) As chases become more direct, observers were more likely to report percepts closer to the "chasing" (social interaction) end of the scale. At less direct chases, ratings were lower (i.e., closer to "moving independently ["moving ind."]). (b) As the charge speed (speed at which agents charge at each other) increased, interactions were perceived to be more aggressive (closer to the "fighting" end of the scale). At lower charge speeds, ratings were closer to "playing". N= 312 and 319 in panels (a) and (b), respectively. Errorbars represent the 95% confidence interval.

Robust individual differences in social perception exist atop group-level tendencies

Even given these shared general tendencies, individuals often vary in their percepts of social interactions, especially when faced with ambiguous scenarios. To quantify this across-subject variability, inspired by psychophysics approaches, we fit individual participants' rating data with a sigmoid (detection experiments) or linear function (discrimination experiments; see **Methods**) to derive individual "social tuning curves". For detection curves, we focused on three main parameters: (i) point of subjective equality (PSE), the value of *chase directness* at which the participant's percept crosses the midpoint of the rating scale (0.5); higher values indicate that more evidence is needed to declare something "social"; (ii) range, the difference between ratings at the lowest (0) and highest (1) levels of *chase directness*; higher values may reflect higher perceptual vividness, certainty or diversity in percepts; and (iii) bias, the extent to which ratings are skewed toward one end of the scale; higher values (> 0.5) indicate a bias toward social ("chasing") while lower values (< 0.05) indicate a bias toward nonsocial ("moving independently”). For discrimination curves, we focused on two main parameters: (i) range, the difference between ratings at the lowest (0) and highest (1) levels of *charge speed*; higher values may reflect higher perceptual vividness, certainty or diversity in percepts; and (ii) bias, the extent to which ratings are skewed toward one end of the scale; higher values (> 0.5) indicate a bias toward “fighting” while lower values (< 0.5) indicate a bias toward “playing”. See Fig 2 for a schematic of these parameters.

While the vast majority of subjects showed the same general directionality as the group-level trends, fine-grained properties of tuning curves differed between subjects (see Fig 4a for sample participants [all data in Fig S4 and S5] and Fig 4b for full distributions of parameters of interest). To test the stability of these tuning curves within participants, we had participants return for a second session 1–2 months later, in which the task design was identical to the first session except that we used previously unseen animations (generated using the same algorithms). Visual inspection showed that idiosyncrasies in behavior and tuning curves were largely preserved across sessions (e.g., Fig 4a).

We quantified the stability and uniqueness of tuning curves in two ways. First, we used parameters calculated on data from a participant’s first session to fit the same participant’s data in their second session (or vice versa). To test the extent to which curve parameters were both stable within people and distinct across people, we compared this within-participant fit to the mean across-participant fit calculated by using each participant’s curve to fit data from all other participants in the same session. Participants’ ratings were generally better predicted by their own curves from a different session than the average of everyone else’s curves in the same sessions (Fig 4c, middle and right violin plots within each subplot; outliers omitted for clarity; detection task: mean difference (MD)=0.04 and 0.05 when fitting session 2 data to session 1 and vice versa, both $p<.001$; discrimination task: MD=0.07 ($p=.006$) and 0.08 ($p=.06$) when fitting session 2 data to session 1 and vice versa). Second, we calculated the intra-class correlation coefficient between parameters fit to data within each session for each parameter of interest. In the detection task (sigmoid fit), PSE, range and bias showed moderate to good reliability; in the discrimination task (linear fit), slope and bias showed generally moderate reliability (Fig 4d; see Fig S6a-b for data on other parameters that were less reliable and/or redundant with the main curve-fit parameters of interest and Fig S6c for the covariance between all curve-fit parameters). Overall, then, individual tuning curves for both social detection and discrimination were both stable within subjects (reliable across sessions) and unique between subjects, suggesting a trait-like component.

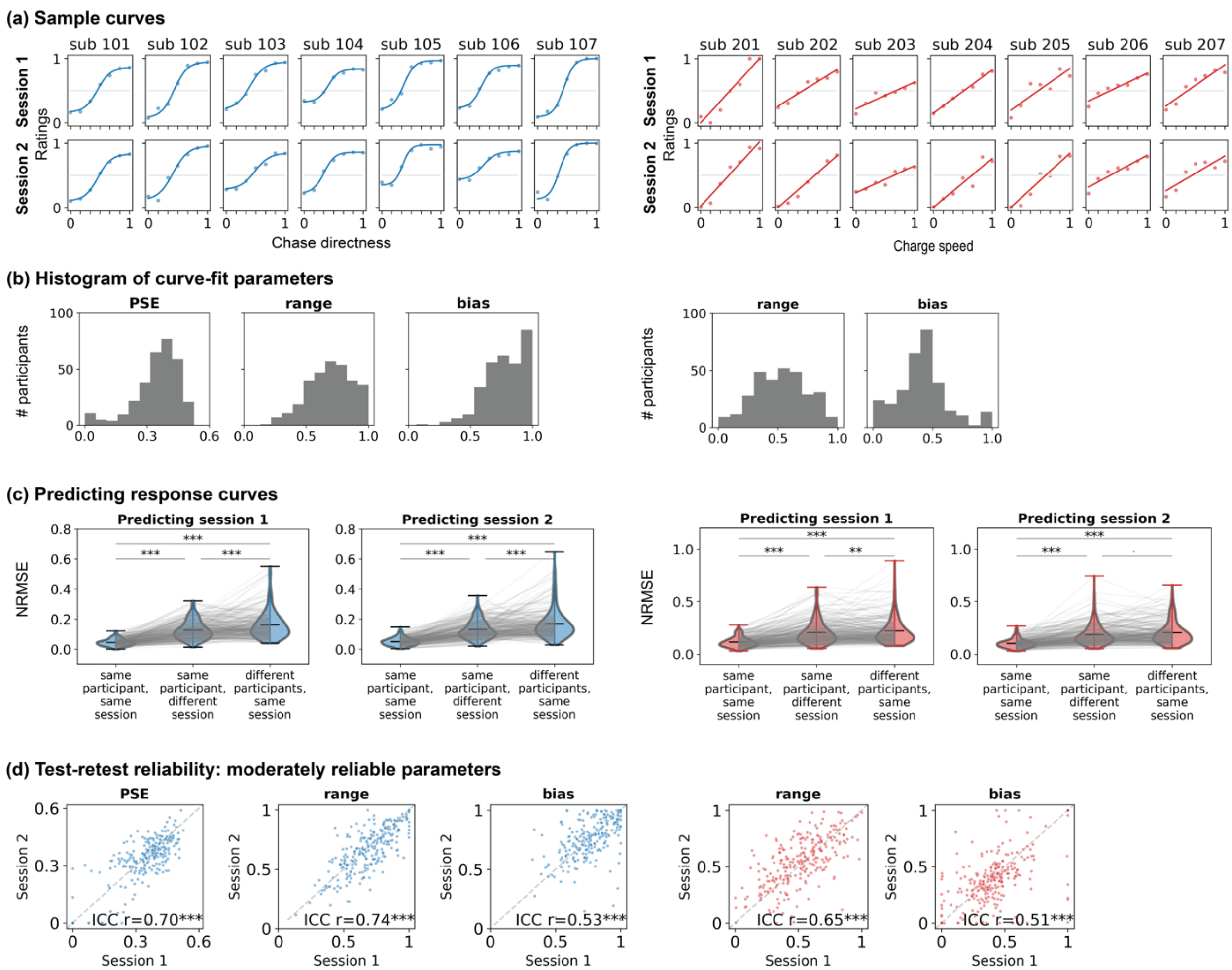

Fig 4: Individual differences in social tuning curves. (a) Tuning curves for sample participants in two distinct sessions. Dots represent mean ratings at each motion attribute level, and the line shows the best fitting sigmoid (detection) or linear curve (discrimination). For the full versions, see Fig S4 and S5. (b) Histograms showing the full distributions of the main curve-fit parameters of interest from all subjects in session 1 (left, detection experiment; right, discrimination experiment). PSE, point of subjective equality. (c) Predicting individuals' single-session ratings using curve parameters fit to their own data from the same session (left), their own data from a different session (middle), or the average parameters from all other participants in the same session (right). NRMSE=normalized root mean square error; lower values indicate better fits. (d) Test-retest reliability between sessions 1 and 2 of the main curve-fit parameters of interest.

Each dot represents a participant. Test-retest reliability for other parameters is shown in Fig S6.
***= $p < .001$, **= $p < .01$, $^{\cdot}$= $p<.1$.

Individual social detection and discrimination tendencies are only weakly related

Social detection and discrimination both exhibit shared tendencies as well as robust individual differences, but how do behaviors relate across the two tasks? In other words, are individuals' discrimination tendencies predictable from their detection tendencies (and vice versa)? For example, people who are more prone to seeing social interactions in the first place might also be more prone to seeing them in a more positive (or negative) light. To study this, we conducted a third experiment in which a new set of participants performed both the detection and discrimination tasks in interleaved blocks in a single session. We successfully replicated the group-level behavioral trends (Fig 3) in this new group of people (see Fig S7). We then fit sigmoid and linear curves to each individual's detection and discrimination data, respectively. As a sanity check, we verified that the covariances across curve-fit parameters within each task (detection task: PSE, range, bias; discrimination task: range, bias) were comparable to earlier experiments (Fig 5).

Next, we correlated curve-fit parameters across participants both within and across tasks. Focusing on between-task correlations (highlighted part of the matrix in Fig 5), the majority of pairwise correlations were weak, suggesting that detection and discrimination behavior are overall relatively independent. We did, however, find two significant relationships. First, the "range" parameter correlated moderately across tasks ($r=0.44$, $q<.05$). This indicates that people who distinguished social ("chasing") from non-social ("moving independently") more strongly also distinguished negative interactions ("fighting") more from positive interactions ("playing") more strongly. This could reflect people's general confidence/willingness to use extremes (people who are more confident may have used a wider range of the rating scale in both tasks) and/or the extent to which their percepts are sensitive to sensory evidence (people whose percepts vary more strongly with sensory evidence would have a higher range in both tasks). The second result was that people who showed a higher bias toward socialness ("chasing") in the detection task also showed a higher range (more distance between extremes) in the discrimination task ($r=.39$, $q<.05$). This suggests that people who are more predisposed to detecting social interactions may also be more sensitive to motion cues and/or more certain when discriminating between different types of interactions. Still, on the whole, social discrimination tendencies were largely not directly predictable from detection tendencies and vice versa. This suggests that behavior on the two tasks may be complementary in revealing individual differences in social perception, and combining information about detection and discrimination tuning curves likely better characterizes individuals' socio-perceptual tendencies than one task alone.

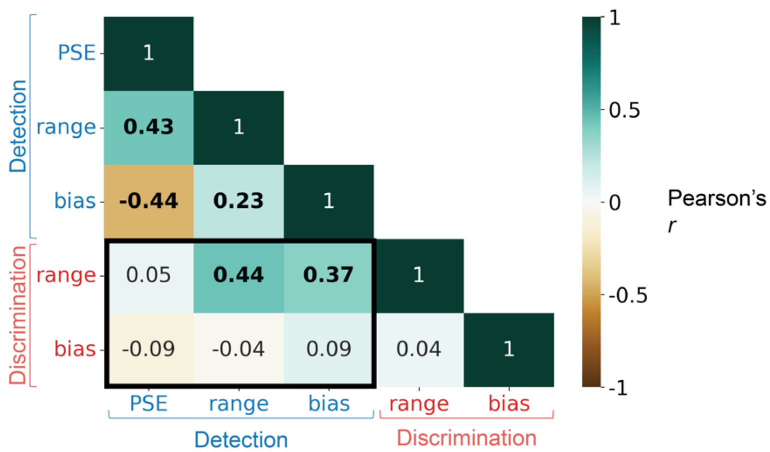

Fig 5: Correlation between the main curve-fit parameters of interest from the social detection and discrimination tasks in the mixed-design experiment (within-subject design). Across tasks (black box), only two moderate pairwise correlations emerged, suggesting that each task provides non-redundant information on individuals' socio-perceptual tendencies. Significant correlations (FDR q < .05) are displayed in bold text.

Combining individuals' social detection and discrimination behavior best relates to trait differences

How do socio-perceptual tendencies as measured by behavior on our detection and discrimination tasks relate to real-world variability in social function? In past work, behavior on related tasks has been found to differ in various clinical and subclinical conditions including autism[6–10], internalizing symptoms[5], depression[14–17] and loneliness[44,45]. In our final set of analyses, we studied if and how properties of individuals' social tuning curves covaried with social, affective, and personality traits as measured by established self-report scales (see **Methods** for details).

We first performed exploratory correlations between all trait scores (14 total) with all 5 (3 for detection, 2 for discrimination) curve-fit parameters. Although some significant correlations emerged, none survived multiple comparison corrections (Fig 6a). The uncorrected correlations suggested that: (1) more extraverted (E) individuals had lower social thresholds (PSE) in the detection task; (2) individuals with higher communication deficits (comm) had less of a bias toward “social” responses, and individuals with (3) higher openness (O) and (4) lower negative affect (neg) showed higher uncertainty (lower range) in discriminating between positive and negative interactions. The first two of these were in line with past work and our *a priori* hypothesis that individuals with social interaction deficits might have higher thresholds (i.e., need more evidence) to detect social information.

One possibility is that, rather than first-order relationships between single tuning-curve parameters and single trait dimensions, relationships between traits and social detection/ discrimination behavior are more complex—perhaps multivariate and/or nonlinear. To explore this possibility, we used inter-subject representational similarity analysis (IS-RSA[42]) to test for a second-order relationship: in other words, to test the hypothesis that individuals with more similar tuning curves in one or both tasks are also more similar in their pattern of trait scores. Results showed that indeed, we could recover such a second-order relationship, but the effect was significant only when combining information about tuning curves from both detection and discrimination tasks (Fig 6b). Thus, features of individuals’ social detection and discrimination behavior in our controlled experimental setting may carry a meaningful, albeit weak, signal as to self-reported real-world social functioning. Taken alongside the relatively weak correlations between detection and discrimination tuning curves seen in the previous section, these results suggest that social detection and discrimination behavior each carry unique information about an individual’s socio-perceptual tendencies.

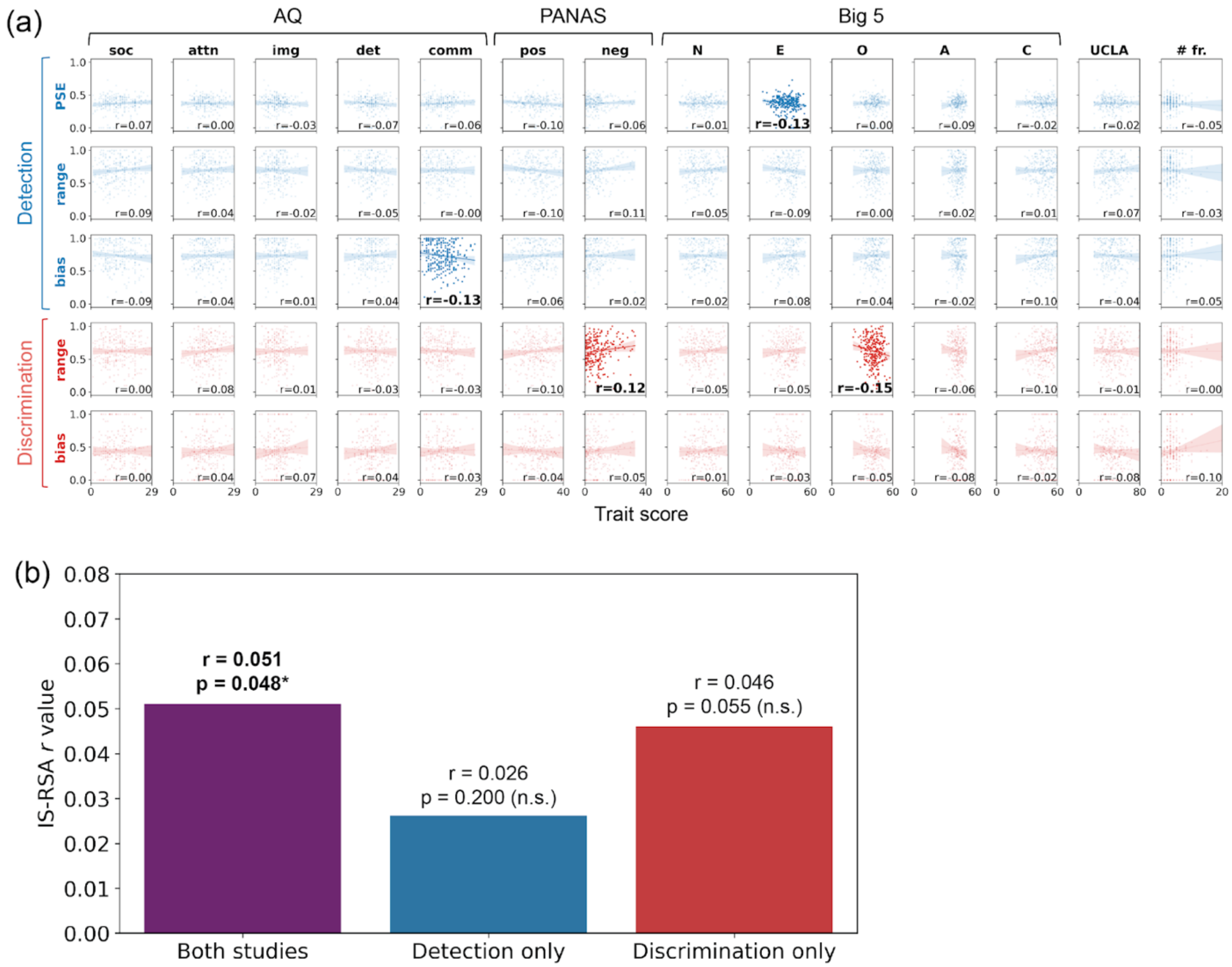

Fig 6: (a) Exploratory pairwise Pearson correlations between each trait score and each curve fit parameter in both the detection and discrimination tasks (exploratory analysis). Correlations that are significant at p < 0.05 (uncorrected for multiple comparisons) are shown in darker red/blue. (b) Results from an inter-subject representational similarity analysis (IS-RSA) testing the second-order hypothesis that pairs of participants with more similar socio-perceptual tendencies on our task(s) also have more similar patterns of trait scores. This relationship is significant only when combining parameters from both detection and discrimination tasks (purple), not when using parameters from a single task alone (blue and red). Abbreviations: Autism Quotient (AQ) questionnaire subscales – 'soc': social skill deficits, 'attn': attention-switching deficits, 'img': imagination deficits, 'det': heightened attention-to-detail and 'comm': communication deficits; Positive and Negative Affect Schedule (PANAS) subscales – 'pos': positive affect and 'neg': negative affect; Big 5 (NEO-FFI) subscales – 'N': neuroticism, 'E':

extraversion, 'O': openness, 'A': agreeableness and 'C': conscientiousness; 'UCLA': UCLA loneliness scale; '# fr.': number of friends.

Discussion

Here, we used a psychophysics-inspired approach to characterize both group-level tendencies and individual differences in social perception. We found both strong commonalities in how people use relatively low-level motion attributes to arrive at percepts of the presence (detection experiments) and nature (discrimination experiments) of a social interaction, and robust individual differences that were replicable over a period of months and showed some—albeit weak and complex—relationships to trait phenotypes.

Our approach lends a level of rigor and precision to the study of social perception. While some notable past work has used parametric stimulus manipulations[22,52–56], here, we extend this approach to individual-subject data to recover single-person social tuning curves that are both reliable and unique. Generating stimuli algorithmically makes our approach simultaneously more objective and more subjective: more objective because we can create parametric manipulations using quantitatively defined features (rather than relying on handcrafted stimuli created and labeled via experimenter intuition[30]) and thereby generalize beyond item-level effects, and more subjective because we are eschewing any notions of a ground truth and classifying behavior according to observers' own reports[57], which better reflects what happens in the real world—where different people can and do interpret the same social situation differently.

Social cognition is traditionally considered a high-level abstract cognitive process, but more recent work has found evidence that recognizing and processing social information begins earlier in the perceptual hierarchy than previously thought[26,56,58–60], and artificial intelligence can extract basic cues as to the presence and nature of social information using fast, automatic, visually-based processes[56]. The recently proposed "third visual pathway" in the brain, which runs along the lateral surface from early visual regions into the superior temporal sulcus and is specialized for extracting social information from dynamic cues, embodies the theory that our visual system might be especially attuned to social information, given its evolutionary importance[59]. Our work adds to this growing body of work by showing that parametrically varying meaningful "mid-level" visual motion attributes[61] even in very stripped-down, simplified stimuli can directly modulate social percepts at both the group and individual level, thus confirming that social perception involves visual evidence accumulation with both an objective and subjective component. Of note, autism is a condition marked by deficits with social cognition, but also altered basic visual processing[62,63]; the idea that perception is the fundamental starting point for social cognition might lead us to discover hierarchical links between aberrations in these two domains.

While the motion attributes we used here, *chase directness* and *charge speed*, were sufficient to evoke varying social detection and discrimination percepts respectively, we also acknowledge that social percepts are governed by many more dimensions than these two. We propose the idea of individual social perception "landscapes" that can be conceptualized as a multidimensional space spanned by objectively defined axes (e.g., motion attributes such as our *chase directness* and *charge speed*, plus many others) where the dimensions themselves are fixed, but sensitivity to these dimensions varies across people and can be expressed in terms of tuning curve parameters. In Fig 7, we show a two-dimensional schematic of what these landscapes might look like and how they might vary. In this example, the horizontal axis represents an attribute that influences *if* a social interaction is perceived (detection; e.g., *chase directness*) and the vertical axis represents an attribute that influences *how* a social interaction is perceived (discrimination; e.g., *charge speed*). A healthy neurotypical individual (Fig 7a) might show moderate sensitivity to objective evidence for socialness (indicated by the saturation gradient along the horizontal axis); once information is generally deemed social, percepts of the *nature* of that information are generally balanced between positive and negative interactions (even vertical distribution of pink and blue). On the other hand, someone with an autism-like phenotype (Fig 7b) might have lower sensitivity to social information—in other words, they might require higher doses of objective evidence to detect a social interaction. For someone with a depression-like phenotype (Fig 7c), detection sensitivity may be largely normal, but discrimination might be skewed toward negative percepts. Lastly, someone with psychosis/paranoia-like traits (Fig 7d) might show both heightened sensitivity to socialness and a bias towards negative percepts—in other words, a proneness to read social intentions, particularly nefarious ones, into scenarios that others might perceive as non-social or, at most, social yet neutral. While we explored only two possible dimensions here, and did not include clinical populations, we see the present set of experiments as a first step toward discovering and characterizing these landscapes—which, because they are based on more implicit behavioral readouts, are possibly less prone to overt bias than self-report measures and therefore a useful complement to existing trait scales. Importantly, our results showed that relationships with classical traits emerged only when combining the two tasks (detection and discrimination), indicating that each axis carries unique variance; adding more dimensions (i.e., using more complex, yet still parameterized stimuli) will likely enhance our ability to characterize real-world social and affective function.

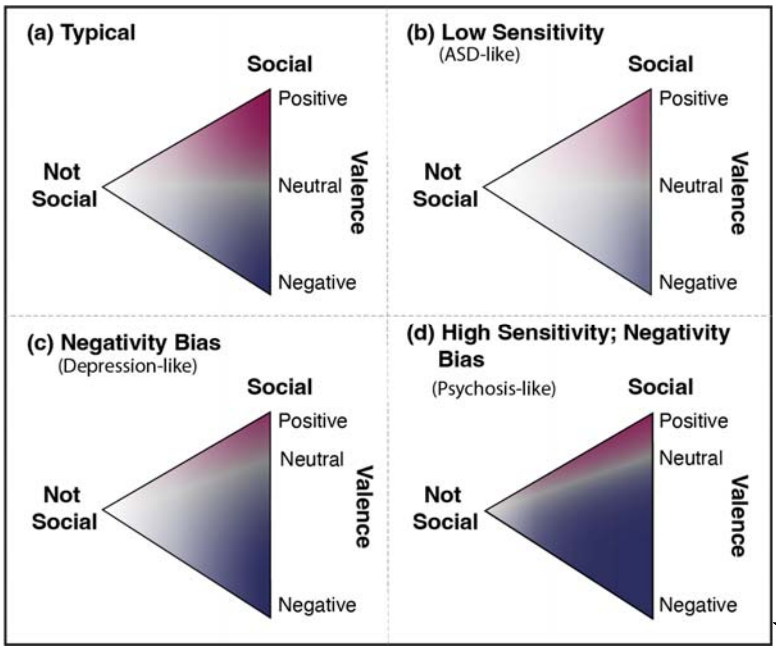

Fig 7. 2D projection of social perception landscapes with detection (not social ⟷ social) and discrimination (positive ⟷ negative) along the horizontal and vertical axes, respectively. How subjective percepts of the presence and nature of social interactions may present in (a) a neurotypical individual versus individuals with (b) autistic, (c) depressive or (d) psychotic traits who show, respectively, reduced progression towards social with sensory evidence (large faded area), increased biases towards negative percepts (large blue zone even at typically neutral or positive evidence), and an increased social sensitivity as well as bias towards negative (leftward saturation combined with expanded blue territory), respectively.

Future work can combine our psychophysics-inspired task framework with additional readouts such as reaction times, eye-tracking, physiological measures and/or neuroimaging to yield a more comprehensive picture of the evidence accumulation and decision-making strategies, attentional processes, and other computations underlying individuals' social-perceptual judgments. There may be a role for generative AI in bridging the gap between the highly impoverished stimuli used here and the full complexity of real-world social information—i.e., we may be able to use generative AI to create more naturalistic-feeling stimuli that are nevertheless still parameterized along known axes. In closing, we note that while the vast majority of past work on social perception and cognition has focused on passive (third person) perception of others' interactions—which is indeed an important part of social cognition—many of our most salient and important social experiences are ones in which we are an active (first-person) participant. One final advantage of our framework is that it can be easily adapted to a first-person context, in which participants themselves are controlling one of the agents and the other agents are programmed to behave in a certain way toward them. This opens the door to generating and comparing social tuning curves between passive and active scenarios, as well as extracting more latent behavioral readouts such as participants' movement trajectories, which we anticipate will provide an even richer and more useful picture of social perception at both the group and individual levels.

Acknowledgements

This work was funded by R01MH129648 (E.S.F.), Neukom CompX Faculty Grant (E.S.F.). We thank Tommy Botch and Clara Sava-Segal for their support and guidance with NLP models and several analyses suggestions, Ahmed Elyamani and Danrae Pray for their help with developing the software, Peng Liu and Jordan Selesnick for their inputs in the preliminary stages of the experiments, undergraduate RAs Alison Sasaki and Eliana Stanford for help with quality-checking stimuli, piloting the experiments and testing the *psyanim* software, and Beyond Bounds Creative for their help with schematics in Fig 7.

References

1. Abassi, E. & Papeo, L. Behavioral and neural markers of visual configural processing in social scene perception. *NeuroImage* **260**, 119506 (2022).
2. Hull, K., Van Hedger, K. & Van Hedger, S. C. Absorption relates to individual differences in visual face pareidolia. *Curr Psychol* **43**, 4458–4474 (2024).
3. Nguyen, M., Vanderwal, T. & Hasson, U. Shared understanding of narratives is correlated with shared neural responses. *NeuroImage* **184**, 161–170 (2019).
4. Ratajska, A., Brown, M. I. & Chabris, C. F. Attributing Social Meaning to Animated Shapes: A New Experimental Study of Apparent Behavior. *The American Journal of Psychology* **133**, 295–312 (2020).
5. Varrier, R. S. & Finn, E. S. Seeing Social: A Neural Signature for Conscious Perception of Social Interactions. *J. Neurosci.* **42**, 9211–9226 (2022).
6. Rasmussen, C. E. & Jiang, Y. V. Judging social interaction in the Heider and Simmel movie. *Quarterly Journal of Experimental Psychology* **72**, 2350–2361 (2019).
7. Zwickel, J., White, S. J., Coniston, D., Senju, A. & Frith, U. Exploring the building blocks of social cognition: spontaneous agency perception and visual perspective taking in autism. *Social Cognitive and Affective Neuroscience* **6**, 564–571 (2011).
8. Klin, A. & Jones, W. Attributing social and physical meaning to ambiguous visual displays in individuals with higher-functioning autism spectrum disorders. *Brain and Cognition* **61**, 40–53 (2006).
9. Vandewouw, M. M. *et al*. Do shapes have feelings? Social attribution in children with autism spectrum disorder and attention-deficit/hyperactivity disorder. *Transl Psychiatry* **11**, 1–11 (2021).

10. Lisøy, R. S. *et al.* Seeing minds – a signal detection study of agency attribution along the autism-psychosis continuum. *Cognitive Neuropsychiatry* **0**, 1–17 (2022).
11. Langdon, R., Boulton, K., Connaughton, E. & Gao, T. Perceiving and attributing intentionality in schizophrenia. *Cognitive Neuropsychiatry* **25**, 269–280 (2020).
12. Horan, W. P. *et al.* Disturbances in the spontaneous attribution of social meaning in schizophrenia. *Psychological Medicine* **39**, 635–643 (2009).
13. Castiello, S., Ongchoco, J. D. K., van Buren, B., Scholl, B. J. & Corlett, P. R. Paranoid and teleological thinking give rise to distinct social hallucinations in vision. *Commun Psychol* **2**, 1–12 (2024).
14. Kohler, C. G., Hoffman, L. J., Eastman, L. B., Healey, K. & Moberg, P. J. Facial emotion perception in depression and bipolar disorder: A quantitative review. *Psychiatry Research* **188**, 303–309 (2011).
15. Liu, W., Huang, J., Wang, L., Gong, Q. & Chan, R. C. K. Facial perception bias in patients with major depression. *Psychiatry Research* **197**, 217–220 (2012).
16. Krause, F. C., Linardatos, E., Fresco, D. M. & Moore, M. T. Facial emotion recognition in major depressive disorder: A meta-analytic review. *Journal of Affective Disorders* **293**, 320–328 (2021).
17. Kaletsch, M. *et al.* Major Depressive Disorder Alters Perception of Emotional Body Movements. *Front. Psychiatry* **5**, (2014).
18. Barrett, H. C., Todd, P. M., Miller, G. F. & Blythe, P. W. Accurate judgments of intention from motion cues alone: A cross-cultural study. *Evolution and Human Behavior* **26**, 313–331 (2005).

19. Blythe, P. W., Todd, P. M. & Miller, G. F. How motion reveals intention: Categorizing social interactions. in *Simple heuristics that make us smart* 257–285 (Oxford University Press, New York, NY, US, 1999).
20. Heider, F. & Simmel, M. An Experimental Study of Apparent Behavior. *The American Journal of Psychology* **57**, 243–259 (1944).
21. Abell, F., Happé, F. & Frith, U. Do triangles play tricks? Attribution of mental states to animated shapes in normal and abnormal development. *Cognitive Development* **15**, 1–16 (2000).
22. Gao, T., Newman, G. E. & Scholl, B. J. The psychophysics of chasing: A case study in the perception of animacy. *Cognitive Psychology* **59**, 154–179 (2009).
23. Isik, L., Koldewyn, K., Beeler, D. & Kanwisher, N. Perceiving social interactions in the posterior superior temporal sulcus. *Proceedings of the National Academy of Sciences* **114**, E9145–E9152 (2017).
24. Scholl, B. J. & Tremoulet, P. D. Perceptual causality and animacy. *Trends in Cognitive Sciences* **4**, 299–309 (2000).
25. Zhou, L.-F. & Meng, M. Do you see the “face”? Individual differences in face pareidolia. *Journal of Pacific Rim Psychology* **14**, e2 (2020).
26. McMahon, E. & Isik, L. Seeing social interactions. *Trends in Cognitive Sciences* (2023) doi:10.1016/j.tics.2023.09.001.
27. Dunbar, R. I. M. The social brain hypothesis – thirty years on. *Annals of Human Biology* **51**, 2359920 (2024).
28. Frith, C. & Frith, U. *What Makes Us Social?* (MIT Press, 2023).

29. Barch, D. M. *et al.* Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* **80**, 169–189 (2013).
30. Castelli, F., Happé, F., Frith, U. & Frith, C. Movement and Mind: A Functional Imaging Study of Perception and Interpretation of Complex Intentional Movement Patterns. *NeuroImage* **12**, 314–325 (2000).
31. Hamlin, J. K., Wynn, K. & Bloom, P. Social evaluation by preverbal infants. *Nature* **450**, 557–559 (2007).
32. Rutherford, M. D. & Kuhlmeier, V. A. *Social Perception: Detection and Interpretation of Animacy, Agency, and Intention*. (MIT Press, 2013).
33. Ullman, T. *et al.* Help or Hinder: Bayesian Models of Social Goal Inference. in *Advances in Neural Information Processing Systems* vol. 22 (Curran Associates, Inc., 2009).
34. Neri, P., Luu, J. Y. & Levi, D. M. Meaningful interactions can enhance visual discrimination of human agents. *Nat Neurosci* **9**, 1186–1192 (2006).
35. Trompenaars, T., Kaluge, T. A., Sarabi, R. & de Swart, P. Cognitive animacy and its relation to linguistic animacy: evidence from Japanese and Persian. *Language Sciences* **86**, 101399 (2021).
36. Schultz, J. & Frith, C. D. Animacy and the prediction of behaviour. *Neuroscience & Biobehavioral Reviews* **140**, 104766 (2022).
37. Leeuw, J. R. de, Gilbert, R. A. & Luchterhandt, B. jsPsych: Enabling an Open-Source Collaborative Ecosystem of Behavioral Experiments. *Journal of Open Source Software* **8**, 5351 (2023).
38. Wang, W. *et al.* MiniLM: Deep Self-Attention Distillation for Task-Agnostic Compression of Pre-Trained Transformers. Preprint at https://doi.org/10.48550/arXiv.2002.10957 (2020).

39. Jolly, E. Pymer4: Connecting R and Python for Linear Mixed Modeling. *Journal of Open Source Software* **3**, 862 (2018).

40. Loureiro, D., Barbieri, F., Neves, L., Anke, L. E. & Camacho-Collados, J. TimeLMs: Diachronic Language Models from Twitter. Preprint at https://doi.org/10.48550/arXiv.2202.03829 (2022).

41. Gescheider, G. A. *Psychophysics: Method, Theory, and Application*. (Taylor & Francis Group, 1985).

42. Finn, E. S. *et al.* Idiosynchrony: From shared responses to individual differences during naturalistic neuroimaging. *NeuroImage* **215**, 116828 (2020).

43. Epley, N., Akalis, S., Waytz, A. & Cacioppo, J. T. Creating Social Connection Through Inferential Reproduction: Loneliness and Perceived Agency in Gadgets, Gods, and Greyhounds. *Psychol Sci* **19**, 114–120 (2008).

44. Gardner, W. L., Pickett, C. L., Jefferis, V. & Knowles, M. On the Outside Looking In: Loneliness and Social Monitoring. *Pers Soc Psychol Bull* **31**, 1549–1560 (2005).

45. Powers, K. E., Worsham, A. L., Freeman, J. B., Wheatley, T. & Heatherton, T. F. Social Connection Modulates Perceptions of Animacy. *Psychol Sci* **25**, 1943–1948 (2014).

46. Tomova, L. *et al.* Acute social isolation evokes midbrain craving responses similar to hunger. *Nat Neurosci* **23**, 1597–1605 (2020).

47. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J. & Clubley, E. The Autism-Spectrum Quotient (AQ): Evidence from Asperger Syndrome/High-Functioning Autism, Males and Females, Scientists and Mathematicians. *J Autism Dev Disord* **31**, 5–17 (2001).

48. Russell, D., Peplau, L. A. & Ferguson, M. L. Developing a Measure of Loneliness. *Journal of Personality Assessment* (1978) doi:10.1207/s15327752jpa4203_11.

49. Watson, D., Clark, L. A. & Tellegen, A. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology* **54**, 1063–1070 (1988).
50. Franić, S., Borsboom, D., Dolan, C. V. & Boomsma, D. I. The Big Five Personality Traits: Psychological Entities or Statistical Constructs? *Behav Genet* **44**, 591–604 (2014).
51. Santavirta, S., Malén, T., Erdemli, A. & Nummenmaa, L. A taxonomy for human social perception: Data-driven modeling with cinematic stimuli. *J Pers Soc Psychol* (2024) doi:10.1037/pspa0000415.
52. Horovitz, S. G., Rossion, B., Skudlarski, P. & Gore, J. C. Parametric design and correlational analyses help integrating fMRI and electrophysiological data during face processing. *NeuroImage* **22**, 1587–1595 (2004).
53. Johansson, M., Mecklinger, A. & Treese, A.-C. Recognition Memory for Emotional and Neutral Faces: An Event-Related Potential Study. *Journal of Cognitive Neuroscience* **16**, 1840–1853 (2004).
54. Looser, C. E. & Wheatley, T. The Tipping Point of Animacy: How, When, and Where We Perceive Life in a Face. *Psychol Sci* **21**, 1854–1862 (2010).
55. Schultz, J. & Bülthoff, H. H. Perceiving animacy purely from visual motion cues involves intraparietal sulcus. *NeuroImage* **197**, 120–132 (2019).
56. Malik, M. & Isik, L. Relational visual representations underlie human social interaction recognition. *Nat Commun* **14**, 7317 (2023).
57. Peters, M. A. K. Introspective psychophysics for the study of subjective experience. *Cerebral Cortex* bhae455 (2024) doi:10.1093/cercor/bhae455.

58. Gandolfo, M. *et al.* Converging evidence that left extrastriate body area supports visual sensitivity to social interactions. *Current Biology* **34**, 343-351.e5 (2024).

59. Pitcher, D. & Ungerleider, L. G. Evidence for a Third Visual Pathway Specialized for Social Perception. *Trends in Cognitive Sciences* **25**, 100–110 (2021).

60. Thieu, M. K., Ayzenberg, V., Lourenco, S. F. & Kragel, P. A. Visual looming is a primitive for human emotion. *iScience* **27**, (2024).

61. McMahon, E., Bonner, M. F. & Isik, L. Hierarchical organization of social action features along the lateral visual pathway. (2023).

62. Choi, Y. B., Mentch, J., Haskins, A. J., Van Wicklin, C. & Robertson, C. E. Visual processing in genetic conditions linked to autism: A behavioral study of binocular rivalry in individuals with 16p11.2 deletions and age-matched controls. *Autism Research* **16**, 831–840 (2023).

63. Robertson, C. E. & Baron-Cohen, S. Sensory perception in autism. *Nat Rev Neurosci* **18**, 671–684 (2017).