https://doi.org/10.26599/AIR.2022.9150010 CAAI Artificial Intelligence Research

Artificial Social Intelligence: A Comparative and Holistic View

Lifeng Fan[1], Manjie Xu[1, 2], Zhihao Cao[1, 3], Yixin Zhu[4] ✉, and Song-Chun Zhu[1, 3, 4] ✉

ABSTRACT

In addition to a physical comprehension of the world, humans possess a high social intelligence—the intelligence that senses social events, infers the goals and intents of others, and facilitates social interaction. Notably, humans are distinguished from their closest primate cousins by their social cognitive skills as opposed to their physical counterparts. We believe that artificial social intelligence (ASI) will play a crucial role in shaping the future of artificial intelligence (AI). This article begins with a review of ASI from a cognitive science standpoint, including social perception, theory of mind (ToM), and social interaction. Next, we examine the recently-emerged computational counterpart in the AI community. Finally, we provide an in-depth discussion on topics related to ASI.

KEYWORDS

social intelligence; theory of mind (ToM); communication; human-machine teaming

What is artificial intelligence (AI)? In contrast to the erroneous belief that AI is solely an engineering subject, it has been a science subject since its inception; John McCarthy defined AI as "the science and engineering of making intelligent machines, especially intelligent computer programs"[1]. AI, like all other scientific disciplines, investigates natural phenomena; in this context, AI focuses on both the physical and social aspects of intelligence.

1 Dawn of Artificial Social Intelligence

Despite controversies[2], the measuring of AI has a long history of employing human-like behavior tests, originating from the Turing test (originally called the imitation game)[3]: a test is administered to determine whether a person is conversing with a real person or a computer program simulating a human. Herbert Simon defined AI similarly with a focus on human-like behaviors: "We call programs intelligent if they exhibit behaviors that would be regarded as intelligent if they were exhibited by human beings."[4] Although modern AI has achieved human-level intelligence in some tasks using data-driven methods[5], it continues to advocate human-like tasks and evaluations[6–9].

Computationally, efforts towards human-like intelligence can be divided into physical intelligence and social intelligence[10], analogous to the developmental psychology ideas of intuitive physics and intuitive psychology[8,11]. In the literature, physical intelligence[12–14] has been studied systematically and extensively in AI[15,16], not only in terms of intuitive physics[17–26] and its applications to challenging AI problems (e.g., in computer vision[27–35] and robotics[36–41]), but also in terms of more abstract forms of physical knowledge[8,9]—causality[42–49] and problem-solving[7,49–56].

Despite its rapid growth in psychology[12,57–60], artificial social intelligence (ASI) has been mostly disregarded in the AI community, with only scattered applications. Notably, cognitive skills for interacting with the social world rather than the physical world distinguish 2.5-year-old human children (prior to reading and schooling) from chimpanzees[61]; humans exhibit significantly more advanced social-cognitive skills than their closest animal cousins. Thus, the research of ASI is essential for the future generation of AI.

To address the aforementioned deficiency, this article highlights a promising AI direction, the ASI, from a computational perspective. In contrast to the mechanical and abstract nature of physical intelligence, ASI involves many subfields that are currently studied separately, such as social perception, theory of mind (ToM), and social interaction[62,63], with varying emphasis on perception, cognitive components, behavior, and even psychometric methods to measure social skills[64]. We intend to provide a comparative and holistic perspective on (1) the gap between existing AI systems and human intelligence, (2) current issues, and (3) future directions by examining human social intelligence and recent efforts on building computational models.

1.1 Unique challenges of context

ASI is distinct and challenging compared to our physical understanding of the world; it is highly context-dependent[63]. This view is shared by Defense Advanced Research Projects Agency (DARPA), which believes that the future generation of AI should include the human-like skill of contextual adaptation[65]—the capacity to reason about and adapt to various contextual inputs. Here, context could be as large as culture and common sense or as little as two friends' shared experiences[58]. This unique challenge prohibits standard algorithms from tackling ASI problems in real-world environments, which are frequently complex, ambiguous, dynamic, stochastic, partially observable, and multi-agent.

ASI is comprised of numerous social signals that are frequently overloaded and ambiguous[66,67]. This difficulty does not even begin

1 National Key Laboratory of General Artificial Intelligence, Beijing Institute for General Artificial Intelligence (BIGAI), Beijing 100080, China
2 School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China
3 Department of Automation, Tsinghua University, Beijing 100084, China
4 Institute for Artificial Intelligence, Peking University, Beijing 100871, China
Address correspondence to Yixin Zhu, yixin.zhu@pku.edu.cn; Song-Chun Zhu, s.c.zhu@pku.edu.cn

at the level of verbal or so-called natural language; rather, it begins with nonverbal communication. Given different contexts, the same gesture (e.g., pointing to a cup) might convey different meanings[68]. Pointing to a cup may indicate its shape, color, capacity to hold water, or a request for assistance in retrieving some water. Consequently, addressing ASI requires a comprehensive approach; improving specific components of an ASI system would not always result in improved performance[9].

1.2 Overview of the article

Multidisciplinary research, including philosophy, cognitive science, neurology, computer science, applied mathematics, statistics, system engineering, and robotics, informs and inspires ASI. In Section 2, which covers social perception, theory of mind, and social interaction, we begin with experimental evidence and theoretical hypotheses of human social intelligence from the standpoint of cognitive science. In Section 3, we present the AI community's computational counterpart, focused on social perception, theory of mind, and social interaction, with an added topic on social robot and cognitive architectures. In Section 4, we explore significant challenges that impede the development of the ASI and recommend potential future trends. Section 5 gives the conclusion.

2 Human Social Intelligence

Evolutionarily, social intelligence development is advantageous for human adaptation to more complex social situations. As a result, studying human social intelligence provides insight into the foundation, curriculum, points of comparison, and benchmarks required to develop ASI with human-like characteristics[63,69].

We concentrate on the three most important aspects of social intelligence: social perception, ToM, and social interaction. We select these themes not just because they are grounded in well-established cognitive science theories but also because there are readily available tools for developing computational models in these areas (to be discussed in Section 3).

Social perception is the basis for ToM and social interaction. It consists primarily of the perception of social features, such as animacy and agency, and provides low-level, automatic, instantaneous, and non-conscious visual perception[70]. ToM, in contrast, is concerned with sophisticated, analytic, and logical cognitive reasoning, involving a general cognitive system with several essential components, including belief, intent, and desire. Social interaction emphasizes more multi-agent interactive activities, such as communication and cooperation, than social perception and ToM.

2.1 Social perception

What factor is the most fundamental and influential in determining social perception? Contrary to intuition, motion cues composed of simple geometry may suffice[71]. According to Michotte[72], "... the specifying factors—gestures, facial expressions, speech—are innumerable and can be differentiated by an infinity of nuances. However, they are all additional refinements compared with the key factors, which are the simple kinetic structures." Heider-Simmel stimuli[73] is perhaps the most seminal work (see the redrawing in Fig. 1). Participants were instructed to watch a film depicting three simple 2D geometric shapes (a large triangle, a small triangle, and a small circle) roaming in the vicinity of a rectangle. Even when told explicitly that these are merely simple shapes, participants still make a rapid, spontaneous, and consistent perception of animate social agents with various

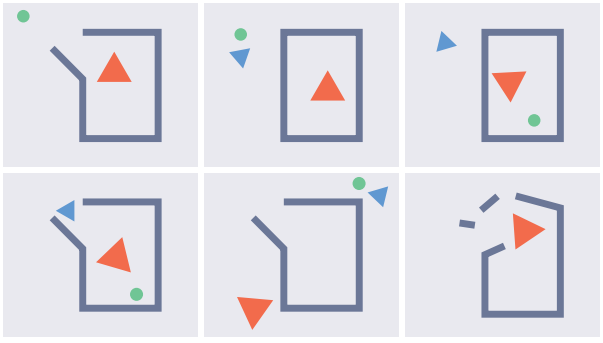

Fig. 1 Heider-Simmel stimuli[73]. Humans can perceive complex mental states and social interactions based solely on the motion of simple geometric shapes.

complex mental states, including desires, goals, emotions, personalities, and coalitions. These mental states combine to form a narrative-like description of the display, such as a hero rescuing a victim from a bully. This interpretation of simple moving shapes as animated agents is a remarkable demonstration of how the human visual system can infer complex social relationships and mental states from simple motion cues with minimal visual characteristics. Even though they involve impressions typically associated with higher-level cognitive processing, such interpretations appear to be predominately perceptual in nature, i.e., relatively rapid, automatic, irresistible, and highly stimulus-driven.

The Heider-Simmel experiment demonstrates two essential aspects of human social perception: the perception of animacy and agency. Animacy denotes that the perceived entities are animate as opposed to inanimate (e.g., physical objects), whereas agency refers to animate beings who are goal-oriented and capable of planning to achieve their goals rationally and efficiently. Below, we concentrate primarily on these two properties (i.e., animacy and agency).

Animacy. Experiments have demonstrated that infants can distinguish between animate and inanimate motion characteristics as early as six months of age[74]. Children ages 3 to 4 can accurately distinguish between mental and physical actions[75]. How can such complex social phenomena be perceived so early?

Michotte[76] describes a seminal experiment that yielded the initial evidence. In this experiment, participants were shown two small squares separated by several inches and arranged in a line. In the first scenario, the first square (A) moves in a straight line until it reaches the second square (B), at which point A stops moving and B begins moving in the same direction (also called the launching effect). In case two, the first square (A) approaches the second square (B). While A approaches, B moves away from A quickly and stops when it is several inches away again. In the first instance, observers observe A physically causing B's motion (also termed as phenomenal causality or the illusion of causality). In contrast, in the second case, A and B are perceived as alive with their own intentions, i.e., A attempting to capture B and B attempting to escape, even though all that is occurring in such films is simple kinematics.

Scholl and Tremoulet[71] provide a comprehensive review of a series of causal perception and animacy experiments conducted by Michotte[76] and Heider and Simmel[73]. Michotte's experiments and subsequent variations reveal that the spatiotemporal parameters mediate causal perceptions, such as relative velocity, speed–mass interaction, path length, and spatial and temporal gap. Minor manipulations, such as a brief spatial or temporal gap, could quickly transform the perceptions from physical causality to

animated interaction[77]. Overwhelming evidence indicates that human perception of animacy appears hardwired into the visual system and is therefore implicit, automatic, and distinct from higher-level cognitive interpretations.

Agency. We now know that humans can automatically perceive complex social phenomena as early as six months of age. A natural follow-up question would be: How can we distinguish between social events and physical phenomena? The solution lies in the notion of agency[78]. An agent is rationally controlled because it has an internal energy source, whereas an object is not.

Similar to animacy, the social perception of agency is primarily associated with motion kinematics as opposed to featural properties. Gergely et al.[79] and Csibra et al.[80] find that relatively simple motion sequences, without self-initiated movement to cue animacy, can elicit an impression of goal-directed behavior in infants aged nine months.

The perception of agency is frequently studied in tandem with animacy for more complex social phenomena. Gao et al.[81] study a particularly salient form of perceived animacy and agency via tasks based on dynamic visual search (the Find-the-Chase task) and a new type of interactive display (the Don't-Get-Caught! task). They used two cues to evaluate the objective accuracy of such perceptions: (1) chasing subtlety—the degree to which the wolf deviates from a perfectly heat-seeking pursuit, and (2) directionality—whether and how the shapes face each other. Gao et al.[82] present the wolfpack effect, a novel social cue to perceived animacy that could effectively, irresistibly, and subtly influence human visual performance and interactive behavior. The study of chasing investigates how the visual system maintains and updates the dynamic social perception of animacy and agency over time and motion[83]; the researchers discovered that temporal dynamics could lead the visual system to either construct or actively reject interpretations of chasing.

What are these perceptions' underlying units? In other words, are these social perceptions identifiable as discrete objects without the necessary movement properties? van Buren et al.[84] depict one disc (the "wolf") pursuing another disc (the "sheep") amidst several distractor discs that are moving. Lines were visible between each pair of discs. In the Unconnected condition, both lines connected distractors in pairs. In the Connected condition, however, one line connected the wolf to a distractor, and the other line connected the sheep to a different distractor. Observers in the Connected condition were markedly less likely to describe these behaviors in terms of mental state. According to the outcomes of their experiments, discrete visual objects are the fundamental units of social perception.

Summary. Does the human visual system have a natural tendency to recognize animacy and agency? The aforementioned experimental findings support the hypothesis that specific bottom-up perceptual processing is specialized and difficult to be "penetrated" by higher-level cognition[71]. This type of social perception may be at the intersection of perceptual and cognitive processing, where basic stimuli are transformed into causal, animate, or even intentional qualities, which are strongly linked to higher-level cognitive processing.

2.2 ToM

ToM is an additional crucial aspect of social intelligence. In their study examining ToM abilities in chimpanzees, Premack and woodruff[85] first establish the term and idea of ToM. The chimpanzee Sarah was shown a brief clip of an experimenter attempting to perform simple tasks. Subsequently, Sarah observed images of several objects, one of which solved the experimenter's dilemma. Sarah could select the correct photograph, demonstrating that she comprehended the task and the problem at hand, i.e., to depict the current scenario and the experimenter's intentions. Their findings highlight two essential components of ToM: a representation of the affair state and a representation of an individual's motivational link to the state, i.e., belief and intention[86].

Formally, ToM entails attributing mental states (such as beliefs, intents, or desires) to oneself and others, as well as acknowledging that people's perspectives and mental constructs may differ from those of the natural world and from one another[87]. Perspective taking in an internal simulation process is one of the defining characteristics of ToM[88], as understanding another agent requires not peering into the agent's brain chemistry or soul, but rather putting onself in the agent's shoes in order to comprehend the agent's copy of world[89] beyond one's own egocentric perspective. The infamous Sally-Anne test[90,91], a classic first-order false belief task, is a well-known experiment on perspective taking.

ToM is replete with noteworthy experimental findings from cognitive development research. ToM formation around age 4 is one of the most important developmental milestones of early childhood[87]. Infants begin to exhibit gaze-following behaviors, identify themselves and others as agents who perform deliberate actions, and are capable of subjectively experiencing the environment by the end of their first year[63,92]. These behaviors are indicative of early development in ToM. Children follow another agent's gaze at approximately 14 months of age, move to acquire visual information, and visually confirm (check back and forth) that the other agent is experiencing the same reality as themselves[92]. By 14–18 months, the infant begins comprehending the mental states of desire, intention, and the causal relationship between emotions and goals through gaze direction[93]. Around the ages of 3–4, children begin to comprehend the differences between their own beliefs and knowledge and those of others, and thus begin to comprehend false beliefs; however, this ability does not become fully stable until the ages of 5–6[94]. Later in the development trajectory[95] is the establishment of second-order ToM, which entails predicting what one person thinks or feels about what another person thinks or feels[94,96].

Intent. Among all the cognitive components of ToM, we concentrate on the intent component and examine the evidence of the development of human intent in greater depth. Since humans can inversely infer the underlying intents of others through social contact and act to fulfill those intents based on their beliefs and desires, intent may be the most crucial component of ToM[9]. In fact, research has shown that humans do not encode the entirety of action details but rather observe and interpret actions in terms of their intentions and store these interpretations for later retrieval[97]. As a fundamental organizing principle that regulates how we comprehend one another and act in the environment, the concept of intent has been awarded a central position within social intelligence and should thus be an essential component of future AI.

The developmental psychology literature indicates that six-month-old infants view human actions as goal-directed behavior[98]. By the age of 10 months, infants segment continuous behavior streams into discrete units that correspond to what adults would perceive as distinct goal-directed acts[9,99,100]. After their first birthday, infants begin to comprehend that an agent may explore multiple plans to achieve a goal and choose one based on environmental conditions[101]. 18-month-old children can deduce and reproduce an action's intended purpose, even if the activity frequently fails to achieve the aim[102]. In addition, infants

can replicate behaviors rationally and effectively based on an evaluation of the environmental restrictions, as opposed to just duplicating movements, indicating that they understand the relationships between the environment, action, and underlying intent[103].

Typically, intentions are hierarchically arranged across extensive spatiotemporal ranges as a sequence of goals[9]. Infants are already capable of perceiving intentions on multiple levels, including concrete action goals, higher order plans, and collaborative goals[104]. Young children can offer assistance based on the inferred intentions of others derived from observing their behaviors (including failed efforts)[105]. Figure 2 depicts a toddler as young as 18 months old who, upon watching an adult with both arms full of books repeatedly knocking into a cabinet with closed doors, infers that the adult intends to store books inside the cabinet and then walks over to open the cabinet for the adult[106].

Categorization. Understanding ToM's categorization may also assist our understanding, given that ToM is a vast topic of a general system. Cognitive ToM emphasizes explicit perspective-taking, representing, and strategic reasoning regarding another person's beliefs, intentions, and generating causal inferences and predictions of the other's behavior. In contrast, affective ToM is more associated with the representation of emotional states and feelings and typically does not emphasize goal states or valuations of possible actions[94]; Roiser and Sahakian[107] employ the words cold cognition (unemotional) and hot cognition (emotion-laden). Cognitive ToM can be further divided into ToM for motivation (i.e., another organism's valuation, intention, purpose, and goal) and ToM for knowledge (i.e., another organism's belief states or taught schemas/scripts)[86].

Individual differences in cognitive strategies are also present[94]. The theory-theory method[108] and simulation-theory approach[109] are examples of these diverse ToM strategies. The theory-theory approach may be based on a set of intrinsic rules or on causal and probabilistic reasoning models, which may be analogous to cold cognition[94] in which mental states are inferred through intellectual processes. The simulation-theory approach relies on the individual's own motivations and deductive reasoning[110].

Challenges. Despite the many approaches used to investigate ToM (such as behavioral analysis, neuroimaging, and neural signal analysis), a coherent picture of what ToM is, how humans and other species engage in it, and what neurological systems contribute to its functioning is still largely unknown[111,112].

2.3 Social interaction

We continue by introducing several concepts and significant studies of social interaction in human social intelligence. Studying social cues, phenomena, rules, and mechanisms in human social interaction could equip ASI with more sophisticated human-like communication and collaboration capabilities.

Social cues. Whiltshire et al.[113] defined a taxonomy of social cues and signals, which includes the following five categories of social cues: paralinguistic (voice prosody and non-language sounds), facial expression (motion and position of facial muscles), gaze (motion and position of the eyes and predicted sight-line), kinematics (motion, position, and posture of the body), and proxemics (use of interpersonal space)[63].

Fig. 2 Altruistic helping in human infants. Human infants as young as 18 months readily help others achieve their goals in a range of contexts, requiring both an understanding of others' goals and an altruistic desire to assist[106].

Gaze communication. Psychological evidence[114] suggests that eyes are stimuli with distinct "hardwired" neural pathways in the brain for their interpretation. Humans have the unique capacity to infer the intentions of another based on gazes. Gaze communication is a primitive form of human communication whose underlying social-cognitive and social-motivational infrastructure serve as a psychological platform upon which diverse linguistic systems might be constructed[58,115]. Thus, gaze communication plays a crucial role in expressing concealed mental states and enhancing verbal communication in social interactions[116].

Joint attention. Fan et al.[115] thoroughly delineated two hierarchical layers of human gaze communication dynamics: atomic-level and event-level. Event-level gaze communication refers to high-level, complex social communication events, such as non-communicative, mutual gaze, gaze aversion, gaze following, and joint attention. Each gaze communication event is a temporal composite of a few gaze communications at the atomic level. Atomic-level gaze communication describes the granular structures of human gaze interactions, including single, mutual, avoid, refer, follow, and share.

Joint attention is the most advanced sort of gaze communication, as it requires two agents (1) to have the same intention to share attention on common stimuli and (2) to be aware that they are sharing a common ground[115]. Typically, joint attention requires a mutual gaze to establish a communication channel, a refer gaze to direct attention to the target, a follow gaze to examine the referred stimuli, and a final mutual gaze to guarantee that the experience is shared[115,117]. In addition to this top-down approach that forms joint attention, there is also a bottom-up approach whereby two agents are drawn to the same stimuli and are familiar with one another. At 48 months of age, infants develop joint attention with mental attribution to represent their own perception, that of an agent, and the object[114]. The formation of shared attention is a vital initial step toward social interaction and imitation, a predecessor to ToM, and the basic foundation of social intelligence[118–120].

Pointing. In social communication, pointing is another essential social cue. According to Tomasello[58], pointing is one of the earliest forms of communication exclusive to the human species (the other is pantomiming). Pointing is also an indicator of particular cognitive abilities, such as being an intentional actor and having ToM[121]. Bates et al.[122] and Brinck[121] are credited with introducing the distinction between imperative pointing and declarative pointing. Declarative pointing is primarily inter-subjective with a signaling function, whereas imperative pointing is based on behaviorally motivated regularities and is used to request the addressee to do something for the subject.

Because the recipient must use context to imagine, discern, and reason about the communicator's communication intentions, the interpretation of pointing is highly context-dependent. Tomasello[58] presented an intriguing example (see Fig. 3): one agent points to a bicycle outside the library to her companion, and depending on the environment, this pointing gesture could have entirely different communication intentions. The common ground between agents is an essential element of social communication and collaboration. All human communication, including linguistic communication, is only possible when the

Context	Intent
A and B mutually know that the bicycle belongs to B's boyfriend C.	"C is in the library. Let's go to find him! "
A and B mutually know that B broke up with C yester-day.	"C is already in the library, so perhaps we should skip it."

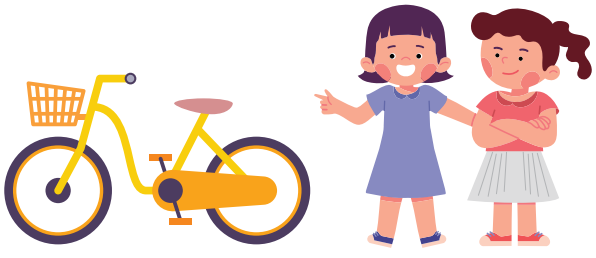

Fig. 3 Importance of context in social signal interpretation[58]. Two girls spot a bicycle outside the library, and one of them, A, points it out to the other, B. Given that A and B are aware that the bicycle belongs to B's boyfriend C, B could take A's statement as "C is at the library; let's go find him!" Given that A and B are aware that B broke up with C yesterday, B could understand A's statement as "C is already in the library, therefore we should probably avoid it."

agents involved have established a common ground composed of shared attention, shared experience, and common cultural knowledge.

Levinson[123] developed the concept of interaction engine, which allows communication intentions to be conveyed and recognized in both linguistic and nonlinguistic encounters. This interactive nature substantially impacts how young children coordinate social interactions with peers[124]. This article does not cover verbal communication studies. Nonetheless, it is essential to note that the basic skills required for effective language communication could be derived from the more rudimentary structures provided here for action control, nonlinguistic communications, and joint actions[125].

Cooperation. Cooperation is a type of social interaction that is more complex than simple communication, as it requires a psychological infrastructure of shared intentionality. This infrastructure is comprised of two crucial factors: (1) social-cognitive skills for creating common conceptual ground with others, such as joint attention and joint intention, and (2) prosocial motivations and norms to help and share with others[58].

Cichocki and Kuleshov[126] examined the precise distinctions between the four notions of communication, coordination, cooperation, and collaboration. By this rigorous definition, com refers to the exchange of information between agents, coordination refers to the alignment of multiple agents towards the achievement of specific common goals through the efforts of individual agents, cooperation means that each individual agent/robot exchanges relevant information and resources in support of each other's goals, rather than a shared common goal, and collaboration requires agents to exchange information and knowledge in support of a shared task.

Tomasello[127] presents a comprehensive analysis and discussion of cooperation. According to his idea of collaboration, "shared cooperative actions" have two essential characteristics: (1) the participants have a joint goal in the sense that we (in mutual knowledge) do X together; and (2) the participants coordinate their interdependent roles—their plans and sub-plans of action, including helping one another as needed in their respective roles. The agents engaging in the cooperation are in We-mode instead of I-mode, i.e., they are imagining a "We".[128–130] Tomasello[127] also proposed a dual-level attentional structure (the shared focus of attention at a higher level, differentiated into perspectives at a lower level) and a dual-level intentional structure (shared goal with individual roles), arguing that the former is directly parallel to the latter and may ultimately derive from it. Fig. 4 illustrates the core idea.

2.4 Summary

This section provides a glimpse into the realm of human social intelligence from the perspective of cognitive science, covering three essential topics: social perception, theory of mind, and social interaction, with growing social interactivity and cognitive complexity. For social perception (Section 2.1), we have explored (1) two most significant concepts (i.e., animacy and agency), (2) what may be the most fundamental, distinguishing, and determining aspect of social perception, and (3) where social perception fits within the human cognitive mechanism. Regarding ToM (Section 2.2), we have discussed its evolution and defining traits. Specifically, we have investigated (1) the findings of one of ToM's most essential components, and (2) the classification of ToM, and (3) its applied cognitive strategies. As for social interaction (Section 2.3), we (1) provided a detailed analysis spanning several most important aspects of social interaction (i.e., gaze communication, joint attention, pointing, cooperation), (2) discussed why these problems are significant, and (3) the theory underlying the social interaction.

It is essential to highlight that these three fundamental aspects of human social intelligence are not isolated but are inextricably linked. Social perception is the foundation for the formation of ToM; they both play crucial roles in human social interaction. Only with well-functioning abilities of social perception and ToM can humans interpret the latent meaning of social cues, understand other agents' mental states (e.g., belief and intent), and cooperate tacitly in a shared task, which are the requirements

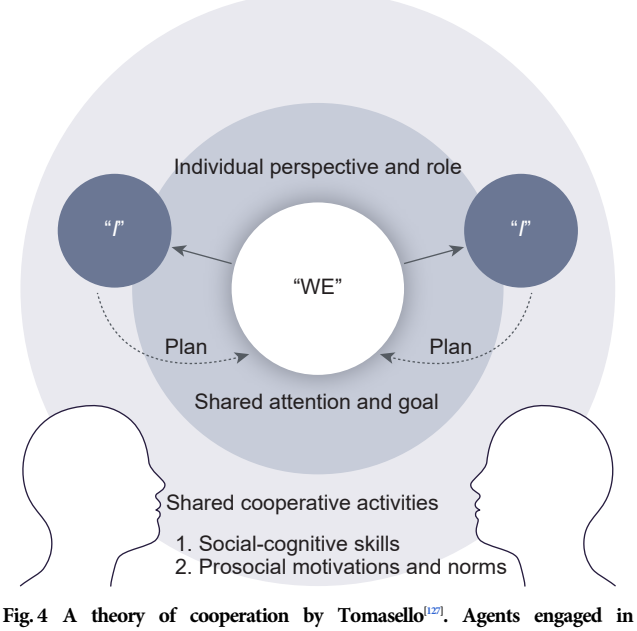

Fig. 4 A theory of cooperation by Tomasello[127]. Agents engaged in cooperation think and act in We-mode rather than I-mode. They have a joint goal and coordinate their roles. For shared cooperative tasks, social-cognitive skills and prosocial incentives and norms are two crucial components[58].

of ASI. In the following section, we describe the computational efforts devoted to these three aspects with a fourth aspect, the social robot and cognitive architectures.

3 Artificial Social Intelligence

In this section, we introduce social intelligence from a computational perspective and highlight some computational works on social perception (in simulated and real-world scenarios), ToM, social interaction (i.e., social communication and cooperation), and social robot. The first three parts are in the same order as in the last section; we add a subsection on social robot and cognitive architectures because this field encompasses the other three aspects of social intelligence and leads to the development of future applications.

3.1 Social perception in simulated scenarios

Since humans possess an innate ability to perceive social cues from extremely simple stimuli, we investigate ways to computationally model social perception in simulated scenarios, akin to the Heider-Simmel stimuli introduced in Section 2.1.

Shu et al.[10] present a unified theory that describes the interrelationships between the perception of physical and social events (see Fig. 5). They employed a simulation-based approach to generate various animations depicting rich behavioral patterns. Through human studies, these animations reveal that the perception of dynamic stimuli transitions gradually from physical to social events and vice versa. In addition, they devise a learning-based computational framework to account for human judgments. Specifically, the model learns to identify latent forces by inferring a family of potential functions capturing physical laws and value functions of agent goals, thereby projecting the animations into a sociophysical space with two psychological dimensions: an intuitive sense of whether physical laws are violated and an impression of whether an agent possesses intentions to perform goal-directed actions.

Tang et al.[131] investigate the problem of simultaneously perceiving physics and mind using a leash-chasing display, in which a disc ("sheep") is being chased by another disc ("wolf") that is physically constrained by a leash tied to a third disc ("master"). They discover that (1) an intuitive physical system, such as a leash, can significantly mitigate the detrimental effects of spatial deviation and the diminishing objecthood on perceived chasing, thereby enhancing its robustness, and (2) a mutual dependency exists between physics and mind, where disrupting one will inevitably impair the perception on the other, supporting a joint perception of physics and mind.

Flatland is a new experimental paradigm introduced by Shu et al.[132] for exploring social inference in physical situations. Results demonstrate that human interpretations of interactive events in Flatland can be accounted for by a computational model that combines inverse hierarchical planning with a physical simulation engine to reason about objects and agents.

Shu et al.[133] examine the perception of social interaction using decontextualized motion trajectories, in which stimuli are extracted from drone-recorded aerial films of a real-world setting. To account for human judgments of interactiveness between two moving dots and the dynamic change of such judgments over time, they construct a hierarchical model that represents interactivity using latent variables and learns the distribution of critical movement features that signal potential interactivity. Intriguingly, the model can generalize to handle the original Heider-Simmel animations[73]. In addition, the generative model can also synthesize decontextualized animations with a controlled degree of interactiveness. The temporal parsing of trajectories and the conditional interactive fields for each sub-interaction are depicted in Fig. 6.

To investigate the cognitive architecture of perceived animacy, Gao et al.[134] devise Bayesian models that integrate domain-specific hypotheses of social agency with domain-general cognitive constraints on sensory, memory, and attentional processing. The proposed model posits that perceived animacy combines a bottom-up, feature-based, parallel search for goal-directed movements with architecturally distinct processes that make perceived animacy fast, flexible, and cognitively efficient. By distinguishing target agents from distractor objects in the "wolf-chasing-sheep" setting, they demonstrate that a Bayesian ideal observer model may explain the efficacy of human perceived animacy with realistic cognitive constraints.

3.2 Social perception in real-world scenarios

In addition to simulations, we further demonstrate computational modeling of social perception in more challenging real-world situations.

Fan et al.[120] investigate the topic of inferring shared attention in their collected third-person social scene video dataset VideoCoAtt by employing a spatiotemporal neural network utilizing human gaze directions and potential target boxes extracted from the context. In their subsequent study[115] (see Fig. 7), the authors systematically investigate the subject of human gaze communication by constructing spatiotemporal graphs for real-world social scenarios in the collected VACATION video dataset.

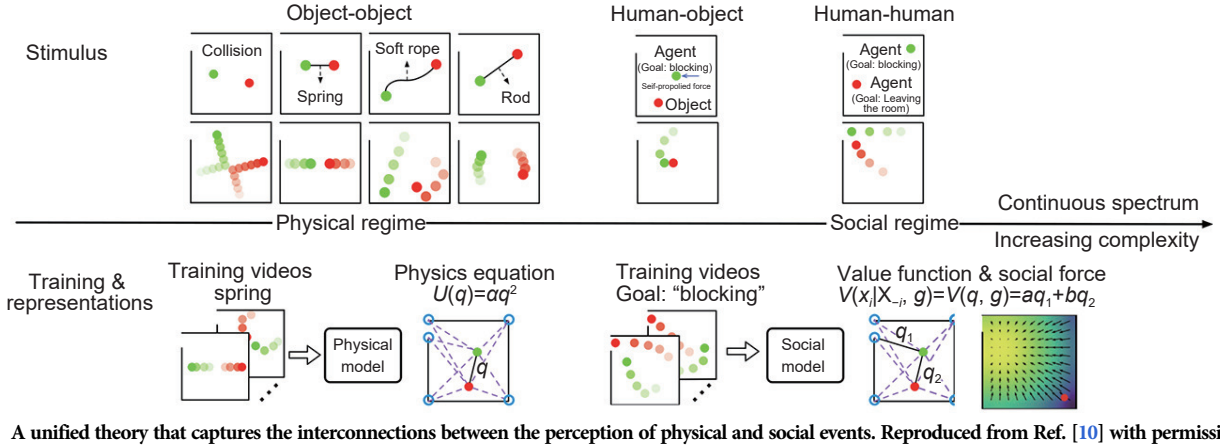

Fig. 5 A unified theory that captures the interconnections between the perception of physical and social events. Reproduced from Ref. [10] with permission of Elsevier Inc., © 2021.

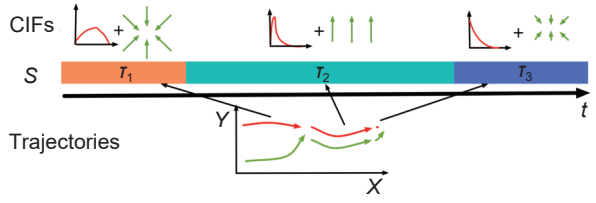

Fig. 6 Perception of human interaction from motion trajectories. The bottom depicts the motion trajectories. The colored bars in the middle represent the temporal parsing of the trajectories in terms of the sub-interaction types (S). The top row depicts the change within a conditional interactive field (CIF) in sub-interactions as the interaction progresses, where the CIF represents the expected relative motion pattern conditioned on the motion of the reference agent. Reproduced from Ref. [133] with permission.

They devise a graph neural network and an event network for the prediction of gaze communication at the atomic and event levels, respectively.

To jointly infer human attention, intention, and task from videos, Wei et al.[135] introduce a hierarchical model of human-attention-object (HAO) and a beam search algorithm. According to their definition, the intention consists of the human pose, attention, and objects, whereas the task is represented as a series of intentions. Xie et al.[136] offer an unsupervised method for localizing functional objects and predicting human intents and trajectories from surveillance footage of public places. Agents are influenced by the attractive or repulsive "fields" of functioning objects, referred to as "dark matter" (see Fig. 8). In addition to estimating the agent's intent, the model can also derive the agent's trajectory via agent-based Lagrangian mechanics.

Holtzen et al.[137] present a method that enables robots to infer a person's hierarchical intent from partially observed RGB-D videos. They represent intent as a novel hierarchical, compositional, and probabilistic And-Or-Graph structure that describes a relationship between actions and plans. Human intent is inferred by reverse-engineering a person's decision-making and action-planning processes under a Bayesian probabilistic programming framework. Experiments conducted in a 3D environment reveal that the inferred human intent (1) corresponds well with human judgment, and (2) provides useful contextual cues for object tracking and action recognition.

3.3 ToM

The computational modeling of ToM may concentrate on different components, such as belief, intent, and desire. Gonzalez and Chang[138] divide computational models of ToM into several broad categories, including Game ToM[139], Observational (RL)[140], Inverse RL[141], and Bayesian ToM[142]. These models contain modules for representing the goals and desires of an agent, inferring the mental states of other agents (e.g., beliefs, goals, desires, intentions, and feelings), and integrating these goals and mentalizing computations to generate optimal policies.

We start this section with some of the most representative studies on different ToM components and modeling methods. Yuan et al.[143] jointly infer object states, robot knowledge, and human beliefs using parse graphs, which accurately identify human (false-)beliefs. Fan et al.[144] (see Fig. 9) incorporate different nonverbal communication cues (e.g., gaze, human poses, and gestures) to infer agents' mental states based solely on visual inputs. By aggregating beliefs and physical-world states, their approach effectively forms five minds during the interactions between two agents. In particular, they construct a common mind to avoid the infinite recursion commonly used in prior works. In addition, they devise a hierarchical energy-based model that

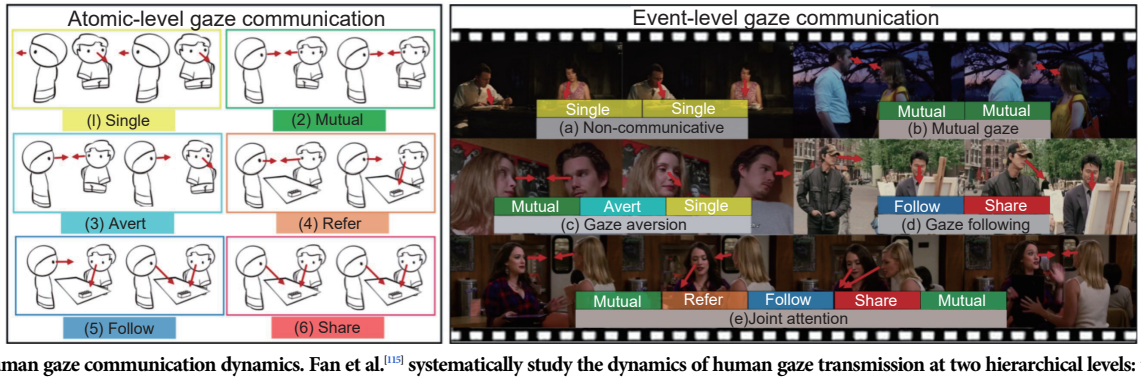

Fig. 7 Human gaze communication dynamics. Fan et al.[115] systematically study the dynamics of human gaze transmission at two hierarchical levels: the atomic level and the event level. Atomic-level gaze communication describes the granular architecture underlying human gaze interactions. Event-level gaze communication refers to complex social communication at the highest level. Each gaze communication event is a temporal composition of several gaze communications at the atomic level. Reproduced from Ref. [115] with permission.

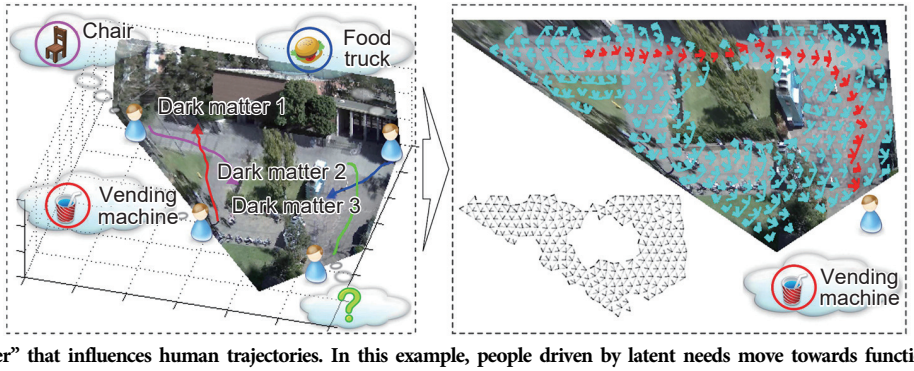

Fig. 8 The "dark matter" that influences human trajectories. In this example, people driven by latent needs move towards functional objects (i.e., "dark matter") that can satisfy their needs. Reproduced from Ref. [136] with permission of IEEE, © 2017.

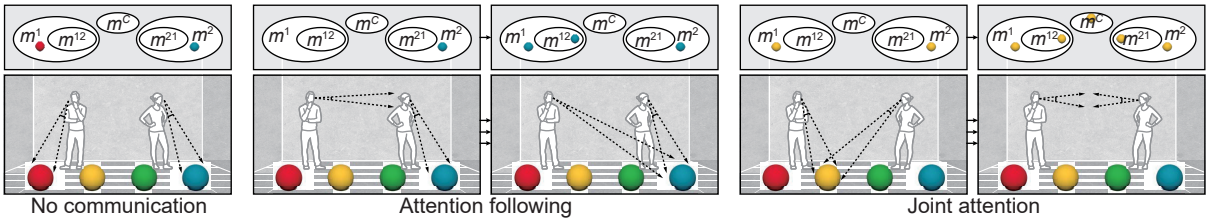

Fig. 9 Triadic belief dynamics in nonverbal communication[144]. In five minds, three sorts of communication events emerge from social interactions (bottom) and causally construct agents' belief dynamics (top). Reproduced from Ref. [144] with permission.

simultaneously tracks and predicts social cues, social communication events, and belief dynamics in five minds. Arslan[145] investigate how 5-year-olds choose and revise reasoning strategies in second-order false belief tasks by constructing two computational cognitive models of this process: an instance-based learning model and a RL model. Oguntola[146] develop an interpretable modular neural framework for modeling the intentions of other observed entities, demonstrating the model's efficacy in a Minecraft search and rescue task. They also demonstrate that, under the right conditions, integrating interpretability can dramatically improve prediction performance. Zeng et al.[147] suggest a brain-inspired model of belief ToM, leveraging high-level knowledge of brain regions' functions relevant to ToM. Although tested on false belief tasks, such cognitive architecture may be difficult to motivate at the computational level[94].

One stream in ToM is based on Bayesian methods. Baker et al.[148] investigate the rational quantitative attribution of beliefs, desires, and percepts in human mentalizing from agents' movement in a local spatial environment (see Fig. 10). They devise a Bayesian theory of mind (BToM) model in a partially observable Markov decision process (POMDP) setting for rational planning and state estimation, which extends classical expected-utility agent models to sequential actions in complex, partially observable domains. In two experiments, their model accurately captures the quantitative mental-state judgments of human participants by

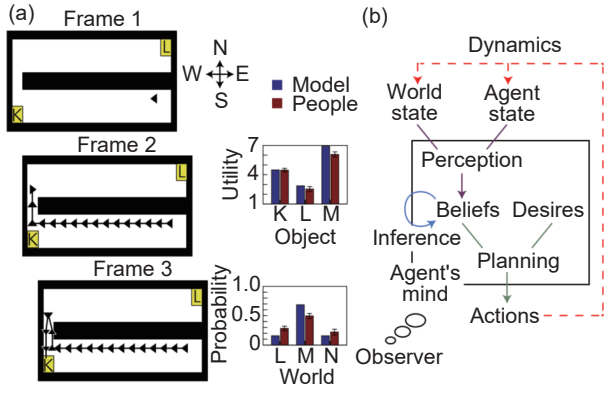

Fig. 10 Experimental scenario and model schema for rational quantitative attribution of beliefs, desires, and percepts in human mentalizing[148]. (a) In the experimental scenario, the agent leaves their office where they can see the K truck (Frame 1). Next, the agent walks past it to the opposite side of the building, where the L truck is parked (Frame 2). Finally, the agent returns to the K truck (Frame 3). The bar charts illustrate the model and human prediction of the agent's utility (i.e., which truck is the agent's preference) and belief (i.e., which truck the agent initially believed to be parked on the other side of the building). (b) The folk-psychological schema for ToM, formulated as a generative action model based on the solution of a POMDP. In this generative model, mentalizing is formulated as Bayesian inference about unseen variables (beliefs, desires, perceptions) conditioned on observed actions. Reproduced from Ref. [148] with permission of Nature Publishing Group, © 2017.

alternating numerous stimulus parameters over a large number of stimuli. A family of simpler non-mentalistic motion features reveals the value contributed by the model's component. BToM appears particularly well-suited to model the inherent uncertainty required to infer unobservable mental states and to capture the judgments of human participants[142]. However, the scalability of BToM is often problematic, only tested in scenarios that are typically simple[94].

RL represents another stream in ToM computational modeling; Wen et al.[149] and Moreno et al.[150] are examples of recursive reasoning models for higher-order ToM in a RL framework. According to Skinner's theory, Hakimzadeh[151] contend that RL plays a crucial role in human intuition and cognition, and theories such as the language of thought hypothesis, script theory, and Piaget's theory of cognitive development offer complementary approaches. They present a computational building block that supports the principles of productivity, systematicity, and inferential coherence for Piaget's schema theory. Reference [152] point out that ToM can indeed be formulated as an inverse reinforcement learning (IRL) problem, where expectations for how mental states produce behavior are represented by a RL model. By simulating the hypothesized beliefs and desires, an RL model predicts the actions of other individuals, and the mental-state inference is accomplished by inverting this model. Overall, RL models, such as IRL and multi-agent reinforcement learning (MARL), are highly scalable but computationally intensive and less interpretable.

Under a POMDP setting, Yuan et al.[153] argue that misalignment of values could impede group performance in cooperation; hence, communication plays a vital role during which a robot needs to serve as an effective listener and an expressive speaker. In the context of value alignment, they investigate how to foster effective bidirectional human-robot communications and propose an explainable artificial intelligence (XAI) system in which a collection of robots anticipates human values by using in-situ feedback while explaining their decision-making processes to users (see Fig. 11). Their XAI system integrates a cooperative communication model to infer human values associated with multiple desirable goals, mimic human mental dynamics, and predict optimal explanations using graphical models.

A related direction is game ToM[139], which leverages concepts like Nash equilibria[138]. de Weerd et al.[154,155] employ a combination of computational agents and Bayesian model selection to determine the extent to which individuals use higher-order ToM reasoning in a particularly competitive game known as matching pennies. Their findings suggest that humans do not primarily employ their high-order ToM abilities. In a case study of the paper-scissors-rock game, Kanwal et al.[156] develop a ToM-based agent, capable of using gestures for non-verbal communication. Tejwani[157] formalize a theory of social interactions, encompassing cooperation, conflict, coercion, competition, and trade, by extending a nested Markov decision process (MDP) where agents

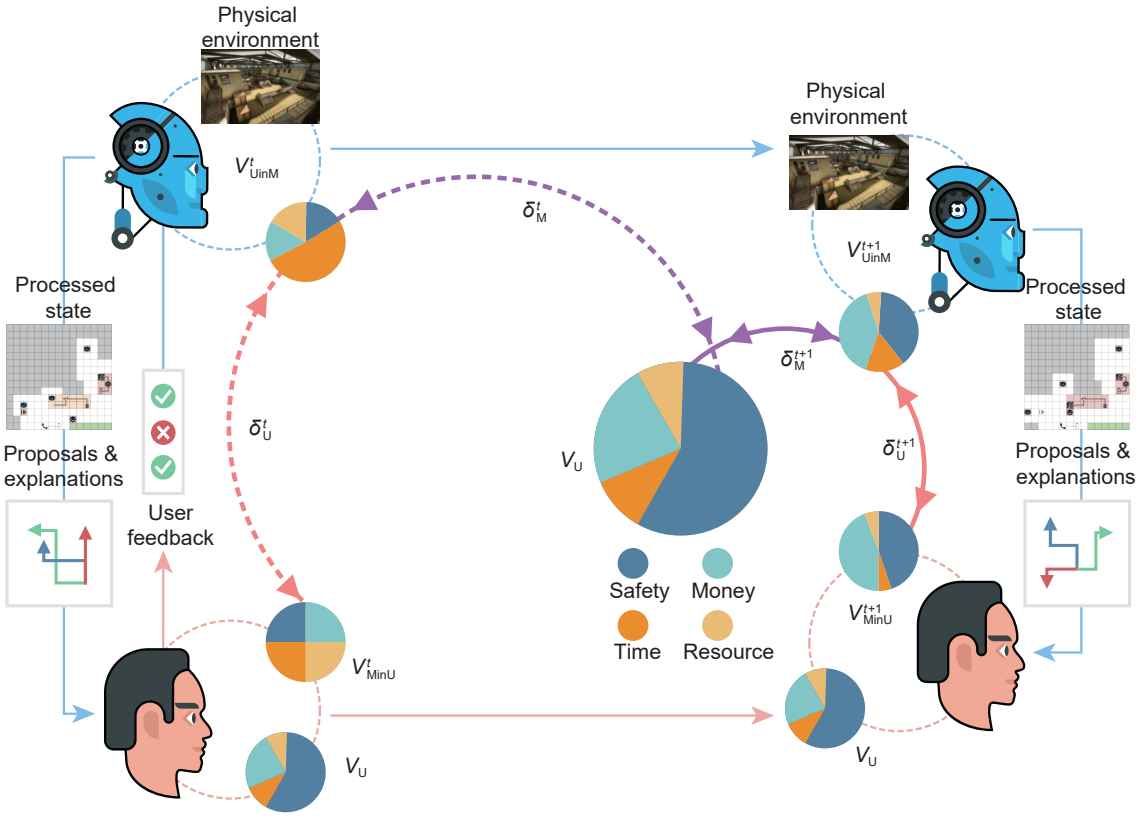

Fig. 11 Bidirectional human-robot value alignment[153]. In a collaboration task, the values—the significance of various goals—are represented by pie charts. In each interaction round, the machine receives signals from the physical environment and processes observations to generate an abstract environment state. Next, the machine offers the processed map together with movement proposals and explanations to human users, who accept or reject the proposals according to the given human values and the current state of the map. Finally, the machine updates its estimation of human values based on the user's feedback and takes action based on the new values. Reproduced from Ref. [153] with permission.

reason about arbitrary functions of each other's hidden rewards. In a follow-up study, Tejwani[158] expand the reward function to incorporate both physical and social goals. Their method permits more complex behaviors, such as politely hindering or aggressively assisting another agent. Panella and Gmytrasiewicz[159] devise a new computational framework, interactive partially observable Markov decision process (I-POMDP), wherein the agent does not explicitly model the beliefs and preferences of other agents but rather represents them as stochastic processes implemented by probabilistic deterministic finite-state controllers (PDFCs). Using Bayesian inference, the agent updates its belief over the PDFCs models of other agents.

Deep learning (DL) is an effective means to approximate complex ToM computations. Aru et al.[69] examine the difficulties associated with applying DL to ToM problems. Although the architectures and learning algorithms are not the ultimate brain-like learning system, they argue that DL remains a solid solution in large-scale tasks and could provide scientific models to aid our comprehension of higher mental functions. They also point out that the problems of existing DL methods are taking shortcuts rather than learning ToM; the system may learn a much simpler decision rule (see Fig. 12). DL for ToM is explored predominantly with deep reinforcement learning (DRL), wherein the agent's experiences and objectives are intertwined. Usually, the task's reward structure determines what the agent accomplishes and learns. However, in the case of ToM, there may not exist a straightforward cost function or reward structure that would necessitate the emergence of ToM. Crucially, Zhao et al.[160] demonstrate in a multi-agent setting that rewards may simply be a byproduct of ToM, not playing a causal role in establishing

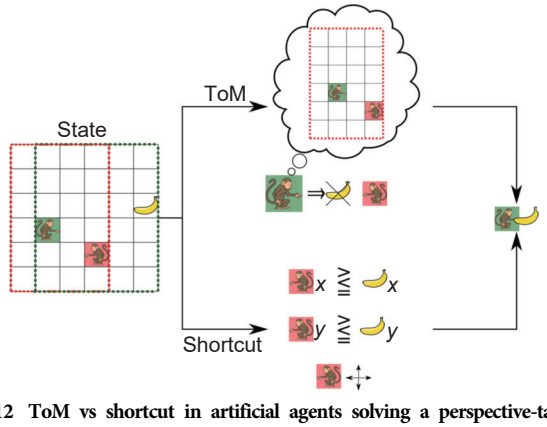

Fig. 12 ToM vs shortcut in artificial agents solving a perspective-taking task. In the left example of an environment state, the dominating agent is denoted by the red square, and the subordinate agent by the green square. The colored dashed lines represent the visual fields of the corresponding agents. If the solution used ToM, the agent in the green square should infer that the dominant agent cannot see the banana, hence pursuing it. In comparison, in the shortcut solution, the agent merely considers the dominating agent's orientation and distance without inferring its perspective. Reproduced from Ref. [69] with permission of Springer, © 2023.

effective coordination.

3.4 Social communication and cooperation

Computational endeavors in modeling social interaction primarily focus on social communication (both nonverbal and verbal) and cooperation.

Nonverbal communication. Jiang et al.[66] model pointing as a

communicative act between agents who have a mutual understanding that the pointed observation must be relevant and interpretable; the act of pointing is an invitation to jointly attend to an object, which elicits mutual inference between agents of each other's minds [67]. The proposed model measures relevance by defining a Smithian value of information (SVI) as the utility gain of a pointing signal. By integrating SVI into rational speech act (RSA), their pragmatic model of pointing permits contextually flexible interpretations. Tang et al.[128] demonstrate that agents can successfully and robustly employ bootstrapping to converge to a joint intention from randomness under an Imagined We framework, leveraging a real-time cooperative hunting task subject to various setting manipulations. Stacy et al.[161] propose a computational account of overloaded signaling from a shared agency perspective, which we refer to as the Imagined We for communication. Within this framework, communication is a means for cooperators to coordinate their perspectives, allowing them to act in concert to achieve shared objectives (see Fig. 13). In a series of simulations, the model performs effectively under growing ambiguity and increasing levels of reasoning, highlighting how shared knowledge and cooperative logic may perform the majority of the heavy lifting in language.

Verbal communication. Studying social communication using natural language in the wild is still challenging. Hence, researchers tend to study verbal communication in a confined domain. Gao et al.[162] devise a novel XAI framework for attaining human-like communication in human-robot collaborations, in which the robot builds a hierarchical mind model of the human user and generates explanations of its own mind as a form of communication based on its online Bayesian inference of the user's mental states. A user study using a real-time human-robot cooking task demonstrates that the generated explanations considerably enhance the collaboration performance and user perception of the robot.

Cooperation. Cooperative tasks demand stronger ToM reasoning in social communication. The notion of ToM-based communication, which chooses information-sharing actions based on relevance and estimation of human beliefs[163], tackles the question of when and what type of information humans require. Wang et al.[164] introduce ToM to build socially intelligent agents, who can communicate and cooperate effectively to accomplish challenging tasks. These agents determine when and with whom to reveal their intentions and sub-goals based on the inferred mental states of others. Pöppel et al.[165] study how efficient, automatic coordination mechanisms at the level of mental states (intentions, objectives), also known as belief resonance, may lead to collaborative situated problem-solving. They describe a model of hierarchical active inference for collaborative agent (HAICA) that blends Bayesian ToM with a perception-action system based on predictive processing and active inference. Belief resonance is realized by allowing the inferred mental states of one agent influence another agent's prediction beliefs regarding its own goals and intentions, hence influencing the agent's task behavior without explicit collaborative reasoning.

3.5 Social robot and cognitive architectures

The social robot is an interdisciplinary research field that requires comprehensive studies of social perception, ToM, and social interaction. We expect a social robot to be endowed with cognitive and affective capabilities, in order to comprehend the feelings, intentions, and beliefs of human agents, which are not only directly expressed by the user but also shaped by bodily cues (e.g., gaze, posture, facial expressions) and vocal cues (e.g., vocal tones and expressions)[166]. A social robot is expected to (1) develop adaptive behavioral models[167], (2) be socially adept, (3) establish a natural, fluent, and effective human-like communication and interaction with humans[168], (4) establish empathetic relationships with humans and be perceived as a teammate or a colleague rather than a tool, (5) offer proactive and parental help based on the observations and understanding of the human situation, and (6) build trust with humans[45]. Understanding robots' decisions promotes the growth of trust and is crucial for facilitating contact between humans and social robots[63].

However, there are still many obstacles to overcome before constructing an ideal social robot[167]. It is difficult to incorporate behavioral adaption techniques, cognitive architectures, persuasive communication strategies, and empathy into a single solution for understanding nonverbal phenomena in social interactions, as contexts are constantly changing. A common limitation of current research is that researchers have focused on a particular aspect of a social robot, such as (1) emphasizing a communication strategy, (2) studying a particular behavior as a response to human action, or (3) conducting experimental studies that include only partial factors.

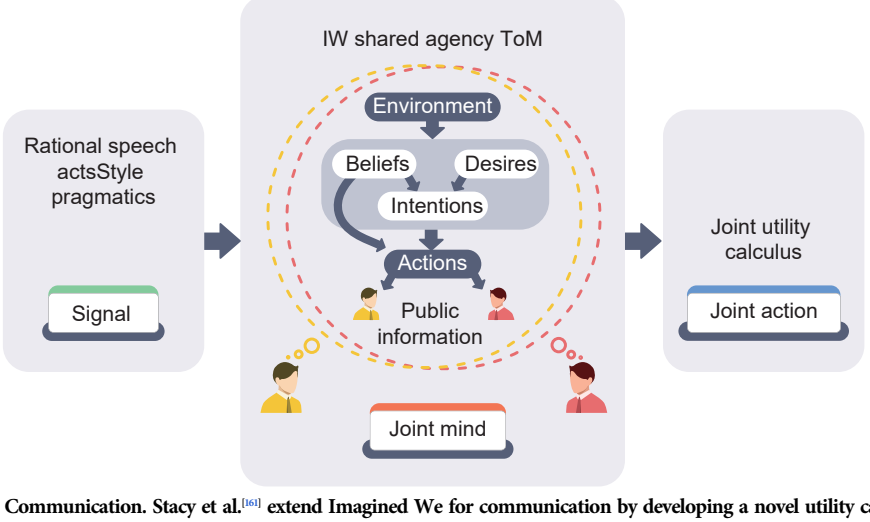

Fig. 13 ***Imagined We*** **for Communication. Stacy et al.[161] extend Imagined We for communication by developing a novel utility calculus of a signal based on shared agency ToM and interactions in the physical world. By integrating with RSA, their model effectively recognizes (1) signals change each Imagined We, (2) minds produce predictable and rational joint actions under ToM reasoning, and (3) actions have well-defined expected utilities, derived through joint planning.**

Cognitive architecture. A cognitive architecture, as a software implementation of a general theory of intelligence, is not a single algorithm or method tackling a particular problem; rather, it is the task-independent infrastructure that learns, encodes, and applies knowledge to produce behavior[169]. One of the challenges in cognitive architecture design is to create a sufficient structure to support coherent and purposeful behavior, while at the same time providing sufficient flexibility to adapt to the specifics of its tasks and environment. ASI in robotic agents relies heavily on the construction of cognitive architecture, which involves both abstract models of cognition and software instantiations of such models[170]. Researchers are working on developing cognitive architectures that approach a fully cognitive state, embedding mechanisms of perception, adaptation, and motivation[171]. Next, we briefly introduce three most common cognitive architectures.

Learning intelligent distribution agent (LIDA) cognitive architecture[172] is an integrated artificial cognitive system that models a broad spectrum of biological cognition, from low-level perception and action to high-level reasoning. Two hypotheses underlie the LIDA architecture and its corresponding conceptual model: (1) Much of human cognition functions through cognitive cycles, which are interactions between conscious contents, memory systems, and action selection, occur frequently (10 Hz). (2) Cognitive cycles serve as the cognitive atoms of which higher-level cognitive processes are composed.

Soar. The Soar cognitive architecture[173] is composed of interacting task-independent modules, including short-term and long-term memories, processing modules, learning mechanisms, and interfaces between them. Since Soar hypothesizes that sufficient regularities exist above the neural level to capture the functionality of the human mind, the majority of knowledge representations in Soar are symbol structures, with architecturally maintained numeric metadata biasing the retrieval and learning of those structures[169]. Soar also facilitates non-symbolic reasoning via the spatial visual system, an interface between perception and working memory.

Adaptive control of thought-rationale architecture (ACT-R)[174,175] includes modules such as (1) a visual module for identifying objects in the visual field, (2) a manual module for controlling the hands, (3) a declarative module for retrieving information from memory, (4) a goal module for tracking current goals and intentions, and (5) a central production system to coordinate these modules. There are buffers within each module that transmit information back and forth to the central production system. The architecture assumes a mixture of serial and parallel processing.

Cognitive architectures in social robots. We now discuss some notable works that implement various cognitive architectures in social robots. Wiltshire et al.[168] discuss the problem of engineering human social-cognitive mechanisms to enable robot social intelligence and provide an integrative perspective of social cognition as a systematic theoretical underpinning for computational instantiations of these mechanisms. They also provide a series of recommendations to facilitate the development of the perceptual, motor, and cognitive architecture. Breazeal et al.[176] provide an integrated socio-cognitive architecture (see Fig. 14) to endow an anthropomorphic robot with the ability to infer mental states such as beliefs, intents, and desires from the observable behavior of its human partner via simulation-theoretic techniques. Kennedy et al.[177] describe an approach known as a like-me simulation, in which the agent uses its own knowledge and capabilities as a model of another agent to predict that agent's actions. They present three examples of a like-me mental simulation in a social context implemented in the embodied version of the adaptive control of thought-rationale architecture (ACT-R) cognitive architecture, ACT-R Embodied (ACT-R/E), including perspective taking, teamwork, and dominant-submissive social behavior. Moulin-Frier et al.[178] suggest the DAC-h3 architecture, which incorporates a reactive interaction engine, a number of state-of-the-art perceptual and

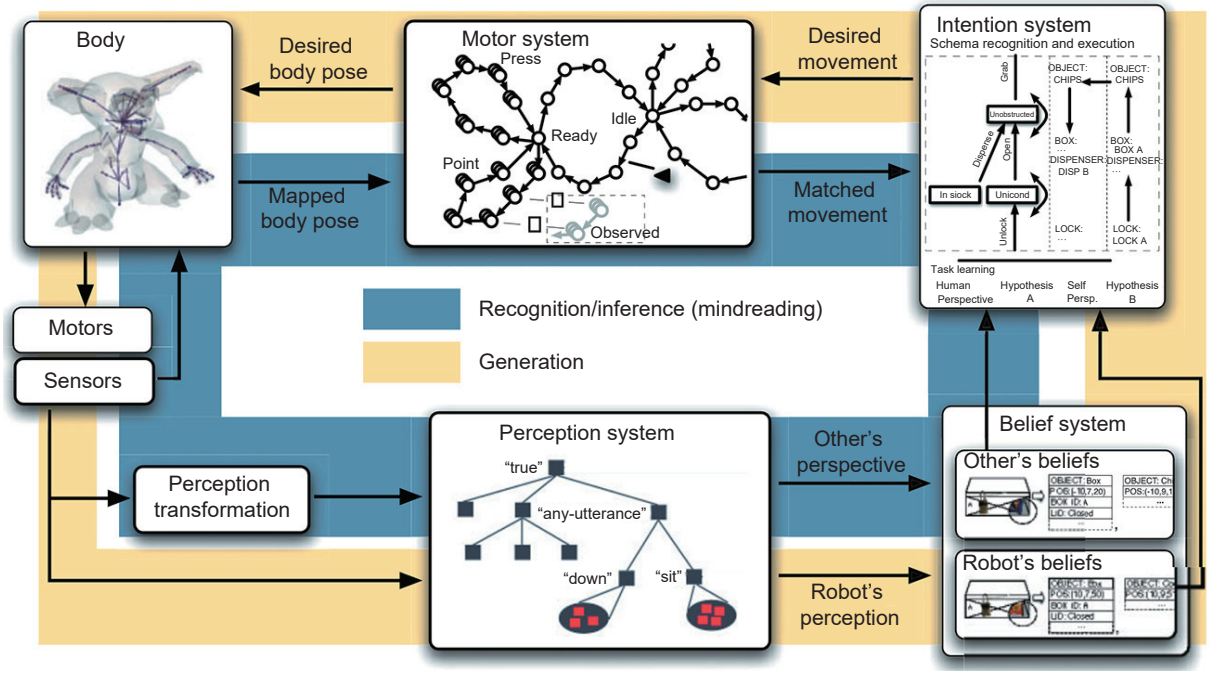

Fig. 14 System architecture incorporating simulation-theoretic mechanisms as a foundational and organizational principle[176]. The two concentric bands represent two distinct operational modes. In generation mode (the light band), the robot builds its own mental states in order to behave intelligently in the environment. In simulation mode (the dark band), the robot constructs and represents its human collaborator's mental states by monitoring their behavior and adopting their mental perspective. Reproduced from Ref. [176] with permission of SAGE Publications, © 2009.

motor learning algorithms, planning capabilities, and an autobiographical memory. The architecture as a whole drives the robot's behavior to solve the symbol grounding problem, acquire language capabilities, perform goal-oriented behavior, and articulate a verbal narrative of its own experience in the world. Franchi et al.[179] present a brain-inspired architecture, the intentional distributed robotic architecture (IDRA), which aims to permit the autonomous development of new goals in situated agents beginning with simple hard-coded instincts.

4 Discussions

4.1 Recent advances in datasets and environments

Datasets and environments are quintessential for developing modern AI. The past few years have witnessed a significant boom of modern treatment. In this section, we provide a brief review of recent notable works.

No previous dataset or benchmark has systematically analyzed physically grounded perception of complex social interactions that extend beyond short actions (e.g., high-five) or simple group tasks (i.e., gathering), until Netanyahu et al.[180]. They resemble a collection of physically-grounded abstract social events (PHASE) that simulates a wide variety of real-world social interactions by incorporating social concepts, such as helping another agent. PHASE is comprised of 2D animations of agent pairs, moving in continuous space with multiple objects and landmarks, generated procedurally by a physics engine and a hierarchical planner.

Inspired by intuitive psychology, Shu et al.[181] present a benchmark consisting of a large dataset of procedurally generated 3D animations, Action, Goal, Efficiency, coNstraint, uTility (AGENT), structured around four scenarios (goal preferences, action efficiency, unobserved constraints, and cost-reward trade-offs) that probe key concepts of core intuitive psychology.

Puig et al.[182] introduced watch-and-help (WAH), a challenge for testing social intelligence in agents, wherein an AI agent is tasked to help a human-like agent perform a complex household task efficiently. They build VirtualHomeSocial, a multi-agent household environment, and provide a benchmark including both planning and learning-based baselines.

Sap et al.[183] proposed a dataset to evaluate language-based commonsense reasoning about social interactions, including reasoning about motivation and about emotional reactions[94].

Bard et al.[184] propose the cooperative and imperfect information card game, Hanabi, as a challenging benchmark. It requires reasoning about the beliefs and the intentions of other players, focusing on the ad-hoc setting where an agent has to coordinate with a team they encounter for the first time.

4.2 Evaluation protocols

The evaluation of social intelligence is arguably the most challenging problem in developing ASI. To answer the question "are we at least making progress towards ASI?", we need an account of how the social intelligence of machines should be measured[185]. The evaluation can aid in testing and training computational models[94].

However, the formation of universally accepted criteria for the design and implementation of ASI benchmarks and the accompanying evaluation protocols is still in its infancy and represents a significant barrier to the field's continued progress. Because human judgments can be ambiguous and difficult to express, many social intelligence tasks do not include requirements that can be easily captured using hand-crafted rules. Hence, a balanced benchmark should likely involve humans evaluating the performance of algorithms. Existing approaches for assessing social intelligence in humans continue to have shortcomings[64].

The Turing test is a test of a machine's ability to exhibit intelligent behavior equivalent to or indistinguishable from that of a human. However, current systems that perform well on these tests typically do so by employing techniques that are not generalizable to other problems. Other approaches for assessing social intelligence competency are often derived from various sources, such as peer-/superior-/self-ratings and observers' behavioral assessments[63, 186]. Notably, the Animal-AI Olympics[187] is initiated by testing artificial agents on tasks derived directly from animal cognition research in an effort to establish common ground.

Typically, the evaluation of ToM in DL is based on the performance of a task; however, this approach is problematic since DL systems may exploit shortcuts—they learn to employ simpler decision rules than recovering the underlying ToM[69]. An important aspect of the ASI is to measure cognitive skills, adaptability, and meta-level learning and reasoning ability rather than specific problem-solving ability[126]. Using more abstract cognitive processes, such as the ability to (1) transfer information from one domain to another, (2) retain information for extended periods, and (3) correct errors in performance, may be future effective strategies for assessing ASI[185].

4.3 Future trends

In this section, we discuss future trends in ASI. We hope these four directions inspire future works in ASI.

A holistic approach. Cognitive and neuroscience research[188] shows that while distinct brain regions are involved in specific tasks, a core network is involved in all ToM tasks, suggesting that humans take a more holistic approach to social intelligence than existing computational models, which often focus on a single aspect of the problem. Through multidisciplinary study spanning psychology, neurology, cognitive science, computer science, statistics, and mathematics, future progress could be accelerated.

Learning methods. Infants develop intelligence gradually[189]. This suggests that learning, and in particular lifelong/continuous learning[190], is a crucial path for developing ASI. The objective of lifelong/continuous learning is to successively learn a model for a large number of activities without forgetting the knowledge acquired from the previous tasks. Other potentially effective learning strategies include multi-task learning[191, 192], one-/few-shot learning, and meta-learning[193].

Open-ended and interactive environment. Infants live in a physical world, full of rich regularities that organize perception, action, and ultimately thought[189]. Infants' intelligence is dispersed across their interactions and experiences with the physical world, which serves to stimulate the development of higher mental functions. In addition, infants behave and learn in a social environment where more experienced partners facilitate learning and provide support. An important aspect of human infants' learning is that they explore; they move and act in extremely unpredictable, random, and non-goal-directed ways. During exploration, they uncover new issues and solutions, and exploration makes intellect open-ended and inventive. Open-endedness departs from the single-task paradigm to an unbounded number of tasks, or even no task at all, simply a world with different possibilities. Open-ended environments could provide a fruitful playground where agents coordinate, cooperate, and compete to solve tasks, and learn similar strategies to social

intelligence in humans, and even more complex behavior[69].

Human biases. The development of social intelligence demands an open-ended setting, yet ToM-like skills would not spontaneously "pop out" from AI agents playing in such contexts[69]. We must also introduce better biases, even structural biases, as a form of built-in common sense, as there may be multiple biases and limits in the human brain that facilitate the acquisition of social intelligence. For instance, there may be innate biases of attention to the human face, speech, hands, eyes, gaze-direction, and biological motion, and these early biases ensure that the infant learns about the components of the world that provide information about the minds of other people. These biases could be hard-coded, evolve from interactions with other agents, or be taught by humans.

5 Conclusion

Although there have been significant advances in AI research, we are still a long way from obtaining human-level intelligence. ASI is a crucial missing component for artificial general intelligence (AGI) on par with humans and symbolizes the future path of AI. Acknowledging ASI as a distinct research area will enhance the field's awareness and encourage academics to discuss and investigate the topic's challenging problems. As one of the most significant promising subfields in AI, ASI requires more theoretical and computational work from the AI community.

Acknowledgment

The authors would like to thank Prof. Tao Gao (UCLA) for brainstorming while the authors were with UCLA, Miss Zhen Chen (BIGAI) and Miss Qing Lei (PKU) for making the nice figures, and two anonymous reviews for constructive feedback. This work was supported in part by the National Key R&D Program of China (No. 2022ZD0114900) and the Beijing Nova Program.

Article History

Received: 21 October 2022; Revised: 3 January 2023; Accepted: 7 January 2023

References

[1] J. McCarthy, What is AI?, http://www-formal.stanford.edu/jmc/whatisai.html, 2004.

[2] T. Ziemke and S. Thellman, Do we really want AI to be human-like? *Sci. Robot.*, vol. 7, no. 68, p. eadd0641, 2022.

[3] A. M. Turing, Computing machinery and intelligence, *Mind*, vol. 59, pp. 443–460, 1950.

[4] M. I. Posner, *Foundations of Cognitive Science*. Cambridge, MA, USA: MIT Press, 1989.

[5] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[6] J. B. Tenenbaum, C. Kemp, T. L. Griffiths, and N. D. Goodman, How to grow a mind: Statistics, structure, and abstraction, *Science*, vol. 331, no. 6022, pp. 1279–1285, 2011.

[7] B. M. Lake, R. Salakhutdinov, and J. B. Tenenbaum, Human-level concept learning through probabilistic program induction, *Science*, vol. 350, no. 6266, pp. 1332–1338, 2015.

[8] B. M. Lake, T. D. Ullman, J. B. Tenenbaum, and S. J. Gershman, Building machines that learn and think like people, *Behav. Brain Sci.*, vol. 40, p. e253, 2017.

[9] Y. Zhu, T. Gao, L. Fan, S. Huang, M. Edmonds, H. Liu, F. Gao, C. Zhang, S. Qi, Y. N. Wu, et al., Dark, beyond deep: A paradigm shift to cognitive AI with humanlike common sense, *Engineering*, vol. 6, no. 3, pp. 310–345, 2020.

[10] T. Shu, Y. Peng, S. C. Zhu, and H. Lu, A unified psychological space for human perception of physical and social events, *Cogn. Psychol.*, vol. 128, p. 101398, 2021.

[11] T. Gerstenberg and J. B. Tenenbaum, Intuitive theories, in *Oxford Handbook of Causal Reasoning*, M. R. Waldmann, Ed. Oxford, UK: Oxford University Press, 2017, pp. 515–547.

[12] E. S. Spelke, *What Babies Know: Core Knowledge and Composition Volume 1*. New York, NY, USA: Oxford University Press, 2022.

[13] E. S. Spelke and K. D. Kinzler, Core knowledge, *Dev. Sci.*, vol. 10, no. 1, pp. 89–96, 2007.

[14] J. R. Kubricht, K. J. Holyoak, and H. Lu, Intuitive physics: Current research and controversies, *Trends Cogn. Sci.*, vol. 21, no. 10, pp. 749–759, 2017.

[15] A. Newell, Physical symbol systems, *Cogn. Sci.*, vol. 4, no. 2, pp. 135–183, 1980.

[16] I. Biederman, R. J. Mezzanotte, and J. C. Rabinowitz, Scene perception: Detecting and judging objects undergoing relational violations, *Cogn. Psychol.*, vol. 14, no. 2, pp. 143–177, 1982.

[17] P. W. Battaglia, J. B. Hamrick, and J. B. Tenenbaum, Simulation as an engine of physical scene understanding, *Proc. Natl. Acad. Sci.*, vol. 110, no. 45, pp. 18327–18332, 2013.

[18] C. Bates, P. W. Battaglia, I. Yildirim, and J. B. Tenenbaum, Humans predict liquid dynamics using probabilistic simulation, in *Proc. 37th Annu. Meeting of the Cognitive Science Society*, Pasadena, CA, USA, 2015, pp. 172–176.

[19] W. Liang, Y. Zhao, Y. Zhu, and S. C. Zhu, Evaluating human cognition of containing relations with physical simulation, in *Proc. 37th Annu. Meeting of the Cognitive Science Society*, Pasadena, CA, USA, 2015, pp. 1356–1361.

[20] J. Kubricht, C. Jiang, Y. Zhu, S. C. Zhu, D. Terzopoulos, and H. Lu, Probabilistic simulation predicts human performance on viscous fluid-pouring problem, in *Proc. 38th Annu. Meeting of the Cognitive Science Society*, Philadelphia, PA, USA, 2016, pp. 1805–1810.

[21] J. Kubricht, Y. Zhu, C. Jiang, D. Terzopoulos, S. C. Zhu, and H. Lu, Consistent probabilistic simulation underlying human judgment in substance dynamics, in *Proc. 39th Annu. Meeting of the Cognitive Science Society*, London, UK, 2017, pp. 700–705.

[22] T. D. Ullman, E. Spelke, P. Battaglia, and J. B. Tenenbaum, Mind games: Game engines as an architecture for intuitive physics, *Trends Cogn. Sci.*, vol. 21, no. 9, pp. 649–665, 2017.

[23] T. Ye, S. Qi, J. Kubricht, Y. Zhu, H. Lu, and S. C. Zhu, The Martian: Examining human physical judgments across virtual gravity fields, *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 4, pp. 1399–1408, 2017.

[24] K. Smith, L. Mei, S. Yao, J. Wu, E. S. Spelke, J. Tenenbaum, and T. D. Ullman, The fine structure of surprise in intuitive physics: When, why, and how much?, in *Proc. 42nd Annu. Meeting of the Cognitive Science Society*, virtual, 2020, pp. 3048–3054.

[25] L. S. Piloto, A. Weinstein, P. Battaglia, and M. Botvinick, Intuitive physics learning in a deep-learning model inspired by developmental psychology, *Nat. Hum. Behav.*, vol. 6, no. 9, pp. 1257–1267, 2022.

[26] S. Li, K. Wu, C. Zhang, and Y. Zhu, On the learning mechanisms in physical reasoning, presented at the 36th Conf. Neural Information Processing Systems, New Orleans, LA, USA, 2022.

[27] J. Wu, I. Yildirim, J. J. Lim, W. T. Freeman, and J. B. Tenenbaum, Galileo: Perceiving physical object properties by integrating a physics engine with deep learning, in *Proc. 28th Int. Conf. Neural Information Processing Systems*, Montreal, Canada, 2015, pp. 127–135.

[28] Y. Zhu, C. Jiang, Y. Zhao, D. Terzopoulos, and S. C. Zhu, Inferring forces and learning human utilities from videos, in *Proc. 2016*

Conf. Computer Vision and Pattern Recognition (*CVPR*), Las Vegas, NV, USA, 2016, pp. 3823–3833.

[29] J. Wu, E. Lu, P. Kohli, W. T. Freeman, and J. B. Tenenbaum, Learning to see physics via visual de-animation, in *Proc. 31st Int. Conf. Neural Information Processing Systems*, Long Beach, CA, USA, 2017, pp. 152–163.

[30] S. Qi, Y. Zhu, S. Huang, C. Jiang, and S. C. Zhu, Human-centric indoor scene synthesis using stochastic grammar, in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 5899–5908.

[31] C. Jiang, S. Qi, Y. Zhu, S. Huang, J. Lin, L. F. Yu, D. Terzopoulos, and S. C. Zhu, Configurable 3D scene synthesis and 2D image rendering with per-pixel ground truth using stochastic grammars, *Int. J. Comput. Vis.*, vol. 126, no. 9, pp. 920–941, 2018.

[32] W. Liang, Y. Zhu, and S. C. Zhu, Tracking occluded objects and recovering incomplete trajectories by reasoning about containment relations and human actions, in *Proc. 32nd AAAI Conf. Artificial Intelligence and 30th Innovative Applications of Artificial Intelligence Conf. and 8th AAAI Symp. Educational Advances in Artificial Intelligence*, New Orleans, LA, USA, 2018, pp. 7106–7113.

[33] S. Huang, S. Qi, Y. Zhu, Y. Xiao, Y. Xu, and S. C. Zhu, Holistic 3D scene parsing and reconstruction from a single RGB image, in *Proc. 15th European Conf. Computer Vision* (*ECCV*), Munich, Germany, 2018, pp. 194–211.

[34] S. Huang, S. Qi, Y. Xiao, Y. Zhu, Y. N. Wu, and S. C. Zhu, Cooperative holistic scene understanding: Unifying 3D object, layout, and camera pose estimation, in *Proc. 32nd Conf. Neural Information Processing Systems*, Montréal, Canada, 2018, pp. 206–217.

[35] Y. Chen, S. Huang, T. Yuan, Y. Zhu, S. Qi, and S. C. Zhu, Holistic++ scene understanding: Single-view 3D holistic scene parsing and human pose estimation with human-object interaction and physical commonsense, in *Proc. 2019 IEEE/CVF Int. Conf. Computer Vision* (*ICCV*), Seoul, Republic of Korea, 2019, pp. 8647–8656.

[36] B. Zheng, Y. Zhao, J. C. Yu, K. Ikeuchi, and S. C. Zhu, Beyond point clouds: Scene understanding by reasoning geometry and physics, in *Proc. Conf. Computer Vision and Pattern Recognition*, Portland, OR, USA, 2013, pp. 3127–3134.

[37] B. Zheng, Y. Zhao, J. Yu, K. Ikeuchi, and S. C. Zhu, Scene understanding by reasoning stability and safety, *Int. J. Comput. Vis.*, vol. 112, no. 2, pp. 221–238, 2015.

[38] Y. Zhu, Y. Zhao, and S. C. Zhu, Understanding tools: Task-oriented object modeling, learning and recognition, in *Proc. 2015 IEEE Conf. Computer Vision and Pattern Recognition* (*CVPR*), Boston, MA, USA, 2015, pp. 2855–2864.

[39] M. Han, Z. Zhang, Z. Jiao, X. Xie, Y. Zhu, S. C. Zhu, and H. Liu, Reconstructing interactive 3D scenes by panoptic mapping and CAD model alignments, in *Proc. 2021 IEEE Int'l Conf. Robotics and Automation* (*ICRA*), Xi'an, China, 2021, pp. 12199–12206.

[40] Z. Zhang, Z. Jiao, W. Wang, Y. Zhu, S. C. Zhu, and H. Liu, Understanding physical effects for effective tool-use, *IEEE Robot. Automat. Lett.*, vol. 7, no. 4, pp. 9469–9476, 2022.

[41] M. Han, Z. Zhang, Z. Jiao, X. Xie, Y. Zhu, S. C. Zhu, and H. Liu, Scene reconstruction with functional objects for robot autonomy, *Int. J. Comput. Vis.*, vol. 130, no. 12, pp. 2940–2961, 2022.

[42] T. L. Griffiths and J. B. Tenenbaum, Theory-based causal induction, *Psychological Review*, vol. 116, no. 4, pp. 661–716, 2009.

[43] M. Edmonds, J. Kubricht, C. Summers, Y. Zhu, B. Rothrock, S. C. Zhu, and H. Lu, Human causal transfer: Challenges for deep reinforcement learning, in *Proc. 40th Annu. Meeting of the Cognitive Science Society*, Madison, WI, USA, 2018, pp. 324–329.

[44] M. Edmonds, S. Qi, Y. Zhu, J. Kubricht, S. C. Zhu, and H. Lu, Decomposing human causal learning: Bottom-up associative learning and top-down schema reasoning, in *Proc. 41st Annu. Meeting of the Cognitive Science Society*, Montreal, Canada, 2019, pp. 1696–1702.

[45] M. Edmonds, F. Gao, H. Liu, X. Xie, S. Qi, B. Rothrock, Y. Zhu, Y. N. Wu, H. Lu, and S. C. Zhu, A tale of two explanations: Enhancing human trust by explaining robot behavior, *Sci. Robot.*, vol. 4, no. 37, p. eaay4663, 2019.

[46] M. Edmonds, X. Ma, S. Qi, Y. Zhu, H. Lu, and S. C. Zhu, Theory-based causal transfer: Integrating instance-level induction and abstract-level structure learning, in *Proc. AAAI Conference on Artificial Intelligence*, New York, NY, USA, 2020, pp. 1283–1291.

[47] C. Zhang, B. Jia, M. Edmonds, S. C. Zhu, and Y. Zhu, Acre: Abstract causal reasoning beyond covariation, in *Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), Nashville, TN, USA, 2021, pp. 10638–10648.

[48] M. Xu, G. Jiang, C. Zhang, S. C. Zhu, and Y. Zhu, EST: Evaluating scientific thinking in artificial agents, arXiv preprint arXiv: 2206.09203, 2022.

[49] C. Zhang, S. Xie, B. Jia, Y. N. Wu, S. C. Zhu, and Y. Zhu, Learning algebraic representation for systematic generalization in abstract reasoning, in *Proc. 17th European Conf. Computer Vision*, Tel Aviv, Israel, 2022, pp. 692–709.

[50] B. Falkenhainer, K. D. Forbus, and D. Gentner, The structure-mapping engine: Algorithm and examples, *Artif. Intell.*, vol. 41, no. 1, pp. 1–63, 1989.

[51] P. N. Johnson-Laird, Mental models and human reasoning, *Proc. Natl. Acad. Sci.*, vol. 107, no. 43, pp. 18243–18250, 2010.

[52] C. Zhang, F. Gao, B. Jia, Y. Zhu, and S. C. Zhu, Raven: A dataset for relational and analogical visual reasoning, in *Proc. 2019 IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), Long Beach, CA, USA, 2019, pp. 5312–5322.

[53] C. Zhang, B. Jia, F. Gao, Y. Zhu, H. Lu, and S. C. Zhu, Learning perceptual inference by contrasting, in *Proc. 33rd Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2019, pp. 1075–1087.

[54] W. Zhang, C. Zhang, Y. Zhu, and S. C. Zhu, Machine number sense: A dataset of visual arithmetic problems for abstract and relational reasoning, in *Proc. AAAI Conf. Artificial Intelligence*, New York, NY, USA, 2020, pp. 1332–1340.

[55] C. Zhang, B. Jia, S. C. Zhu, and Y. Zhu, Abstract spatial-temporal reasoning via probabilistic abduction and execution, in *Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), Nashville, TN, USA, 2021, pp. 9731–9741.

[56] A. Hafri and C. Firestone, The perception of relations, *Trends Cogn. Sci.*, vol. 25, no. 6, pp. 475–492, 2021.

[57] M. Tomasello, Do apes ape, in *Social Learning in Animals: The Roots of Culture*, C. M. Heyes and B. G. Galef, Jr. , Eds. San Diego, CA, USA: Academic Press, 1996, pp. 319–346.

[58] M. Tomasello, *Origins of Human Communication*. Cambridge, MA, USA: MIT Press, 2010.

[59] S. Kita, *Pointing: Where Language, Culture, and Cognition Meet*. New York, NY, USA: Psychology Press, 2003.

[60] R. M. Scott, E. Roby, and R. Baillargeon, How sophisticated is infants' theory of mind?, in *The Cambridge Handbook of Cognitive Development*, O. Houdé and G. Borst, Eds. Cambridge, UK: Cambridge University Press, 2022, pp. 242–268.

[61] E. Herrmann, J. Call, M. V. Hernández-Lloreda, B. Hare, and M. Tomasello, Humans have evolved specialized skills of social cognition: The cultural intelligence hypothesis, *Science*, vol. 317, no. 5843, pp. 1360–1366, 2007.

[62] E. L. Thorndike, Intelligence and its uses, *Harper's Magazine*, vol. 140, pp. 227–235, 1920.

[63] J. Williams, S. M. Fiore, and F. Jentsch, Supporting artificial social intelligence with theory of mind, *Front. Artif. Intell.*, vol. 5, p. 750763, 2022.

[64] D. Silvera, M. Martinussen, and T. I. Dahl, The tromsø social intelligence scale, a self-report measure of social intelligence, *Scand. J. Psychol.*, vol. 42, no. 4, pp. 313–319, 2001.

[65] J. Launchbury, A DARPA perspective on artificial intelligence,https://www.darpa.mil/about-us/darpa-perspective-on-

ai, 2017.
[66] K. Jiang, S. Stacy, A. Chan, C. Wei, F. Rossano, Y. Zhu, and T. Gao, Individual vs. joint perception: A pragmatic model of pointing as smithian helping, in *Proc. 43rd Annu. Meeting of the Cognitive Science Society*, Vienna, Austria, 2021, pp. 1781–1787.
[67] K. Jiang, A. Dahmani, S. Stacy, B. Jiang, F. Rossano, Y. Zhu, and T. Gao, What is the point? A theory of mind model of relevance, in *Proc. 44th Annu. Meeting of the Cognitive Science Society*, Toronto, Ontario, Canada, 2022, pp. 3369–3375.
[68] L. Wittgenstein, *The Big Typescript: TS 213*, Hoboken, NJ, USA: Wiley-Blackwell, 2012.
[69] J. Aru, A. Labash, O. Corcoll, and R. Vicente, Mind the gap: Challenges of deep learning approaches to theory of mind, *Artif. Intell. Rev.*,doi: 10.1007/s10462-023-10401-x.
[70] K. Frankish, Dual-process and dual-system theories of reasoning, *Philosophy Compass*, vol. 5, no. 10, pp. 914–926, 2010.
[71] B. J. Scholl and P. D. Tremoulet, Perceptual causality and animacy, *Trends Cogn. Sci.*, vol. 4, no. 8, pp. 299–309, 2000.
[72] A. Michotte, The emotions regarded as functional connections, in *Michottes Experimental Phenomenology of Perception*, G. Thinés, A. Costall, and G. Butterworth, Eds. Abingdon, UK: Routledge, 1991, pp. 103–116.
[73] F. Heider and M. Simmel, An experimental study of apparent behavior, *Amer. J. Psychol.*, vol. 57, no. 2, pp. 243–259, 1944.
[74] D. H. Rakison and D. Poulin-Dubois, Developmental origin of the animate–inanimate distinction, *Psychol. Bull.*, vol. 127, no. 2, pp. 209–228, 2001.
[75] H. M. Wellman and D. Estes, Early understanding of mental entities: A reexamination of childhood realism, *Child Dev.*, vol. 57, no. 4, pp. 910–923, 1986.
[76] A. Michotte, *The Perception of Causality*. London: Routledge, 2017.
[77] D. S. Berry, S. J. Misovich, K. J. Kean, and R. M. Baron, Effects of disruption of structure and motion on perceptions of social causality, *Pers. Soc. Psychol. Bull.*, vol. 18, no. 2, pp. 237–244, 1992.
[78] Y. Luo, L. Kaufman, and R. Baillargeon, Young infants' reasoning about physical events involving inert and self-propelled objects, *Cogn. Psychol.*, vol. 58, no. 4, pp. 441–486, 2009.
[79] G. Gergely, Z. Nádasdy, G. Csibra, and S. Bíró, Taking the intentional stance at 12 months of age, *Cognition*, vol. 56, no. 2, pp. 165–193, 1995.
[80] G. Csibra, G. Gergely, S. Bíró, O. Koós, and M. Brockbank, Goal attribution without agency cues: The perception of 'pure reason' in infancy, *Cognition*, vol. 72, no. 3, pp. 237–267, 1999.
[81] T. Gao, G. E. Newman, and B. J. Scholl, The psychophysics of chasing: A case study in the perception of animacy, *Cogn. Psychol.*, vol. 59, no. 2, pp. 154–179, 2009.
[82] T. Gao, G. McCarthy, and B. J. Scholl, The wolfpack effect: Perception of animacy irresistibly influences interactive behavior, *Psychol. Sci.*, vol. 21, no. 12, pp. 1845–1853, 2010.
[83] T. Gao and B. J. Scholl, Chasing vs. stalking: Interrupting the perception of animacy, *J. Exp. Psychol. :Hum. Percept. Perform.*, vol. 37, no. 3, pp. 669–684, 2011.
[84] B. van Buren, T. Gao, and B. J. Scholl, What are the underlying units of perceived animacy? Chasing detection is intrinsically object-based, *Psychon. Bull. Rev.*, vol. 24, no. 5, pp. 1604–1610, 2017.
[85] D. Premack and G. Woodruff, Does the chimpanzee have a theory of mind? *Behav. Brain Sci.*, vol. 1, no. 4, pp. 515–526, 1978.
[86] T. Rusch, S. Steixner-Kumar, P. Doshi, M. Spezio, and J. Gläscher, Theory of mind and decision science: Towards a typology of tasks and computational models, *Neuropsychologia*, vol. 146, p. 107488, 2020.
[87] N. Chevalier and A. Blaye, False-belief representation and attribution in preschoolers: Testing a graded-representation hypothesis, *Curr. Psychol. Lett.*, vol. 18, no. 1, 2006.
[88] Y. Barnes-Holmes, L. McHugh, and D. Barnes-Holmes, Perspective-taking and theory of mind: A relational frame account, *Behav. Anal. Today*, vol. 5, no. 1, pp. 15–25, 2004.
[89] R. Fjelland, Why general artificial intelligence will not be realized, *Humanit. Soc. Sci. Commun.*, vol. 7, p. 10, 2020.
[90] H. Wimmer and J. Perner, Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children' s understanding of deception, *Cognition*, vol. 13, no. 1, pp. 103–128, 1983.
[91] S. Baron-Cohen, A. M. Leslie, and U. Frith, Does the autistic child have a "theory of mind"? *Cognition*, vol. 21, no. 1, pp. 37–46, 1985.
[92] H. M. Wellman, Developing a theory of mind, in *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, U. Goswami, Ed. Chichester, West Sussex: John Wiley & Sons, 2011, pp. 258–284.
[93] R. Saxe, S. Carey, and N. Kanwisher, Understanding other minds: Linking developmental psychology and functional neuroimaging, *Annu. Rev. Psychol.*, vol. 55, pp. 87–124, 2004.
[94] C. Langley, B. I. Cirstea, F. Cuzzolin, and B. J. Sahakian, Theory of mind and preference learning at the interface of cognitive science, neuroscience, and AI: A review, *Front. Artif. Intell.*, vol. 5, p. 778852, 2022.
[95] C. Westby and L. Robinson, A developmental perspective for promoting theory of mind, *Top. Lang. Disord.*, vol. 34, no. 4, pp. 362–382, 2014.
[96] J. Perner and B. Lang, Development of theory of mind and executive control, *Trends Cogn. Sci.*, vol. 3, no. 9, pp. 337–344, 1999.
[97] D. A. Baldwin and J. A. Baird, Discerning intentions in dynamic human action, *Trends Cogn. Sci.*, vol. 5, no. 4, pp. 171–178, 2001.
[98] A. L. Woodward, Infants selectively encode the goal object of an actor's reach, *Cognition*, vol. 69, no. 1, pp. 1–34, 1998.
[99] A. N. Meltzoff and R. Brooks, "Like me" as a building block for understanding other minds: Bodily acts, attention, and intention, in *Intentions and Intentionality: Foundations of Social Cognition*, B. F. Malle, L. J. Moses, and D. A. Baldwin, Eds. Cambridge, MA, USA: The MIT Press, 2001, pp. 171–191.
[100] D. A. Baldwin, J. A. Baird, M. M. Saylor, and M. A. Clark, Infants parse dynamic action, *Child Dev.*, vol. 72, no. 3, pp. 708–717, 2001.
[101] M. Tomasello, M. Carpenter, J. Call, T. Behne, and H. Moll, Understanding and sharing intentions: The origins of cultural cognition, *Behav. Brain Sci.*, vol. 28, no. 5, pp. 675–691, 2005.
[102] A. N. Meltzoff, Understanding the intentions of others: Re-enactment of intended acts by 18-month-old children, *Dev. Psychol.*, vol. 31, no. 5, pp. 838–850, 1995.
[103] G. Gergely, H. Bekkering, and I. Király, Rational imitation in preverbal infants, *Nature*, vol. 415, no. 6873, pp. 755–755, 2002.
[104] A. L. Woodward, J. A. Sommerville, S. Gerson, A. M. E. Henderson, and J. Buresh, The emergence of intention attribution in infancy, *Psychol. Learn. Motiv.*, vol. 51, pp. 187–222, 2009.
[105] Z. X. Tan, J. L. Mann, T. Silver, J. B. Tenenbaum, and V. K. Mansinghka, Online Bayesian goal inference for boundedly-rational planning agents, in *Proc. 34th Int. Conf. Neural Information Processing Systems*, Vancouver, Canada, 2020, pp. 19238–19250.
[106] F. Warneken and M. Tomasello, Altruistic helping in human infants and young chimpanzees, *Science*, vol. 311, no. 5765, pp. 1301–1303, 2006.
[107] J. P. Roiser and B. J. Sahakian, Hot and cold cognition in depression, *CNS Spectr.*, vol. 18, no. 3, pp. 139–149, 2013.
[108] A. Gopnik and H. M. Wellman, Why the child's theory of mind really is a theory, *Mind & Language*, vol. 7, no. 1-2, pp. 145–171, 1992.
[109] R. M. Gordon, Folk psychology as simulation, *Mind & Language*, vol. 1, no. 2, pp. 158–171, 1986.
[110] R. M. Gordon, 'Radical' simulationism, in *Theories of Theories of Mind*, P. Carruthers and P. K. Smith, Eds. Cambridge, UK: Cambridge University Press, 1996, pp. 11–21.

[111] N. J. Emery and N. S. Clayton, Comparative social cognition, *Annu. Rev. Psychol.*, vol. 60, pp. 87–113, 2009.
[112] S. M. Schaafsma, D. W. Pfaff, R. P. Spunt, and R. Adolphs, Deconstructing and reconstructing theory of mind, *Trends Cogn. Sci.*, vol. 19, no. 2, pp. 65–72, 2015.
[113] T. J. Wiltshire, E. J. Lobato, J. Velez, F. Jentsch, and S. M. Fiore, An interdisciplinary taxonomy of social cues and signals in the service of engineering robotic social intelligence, in *Proc. SPIE 9084*, *Unmanned Systems Technology XVI*, Baltimore, Maryland, United States, 2014, p. 90840F.
[114] N. J. Emery, The eyes have it: The neuroethology, function and evolution of social gaze, *Neurosci. Biobehav. Rev.*, vol. 24, no. 6, pp. 581–604, 2000.
[115] L. Fan, W. Wang, S. C. Zhu, X. Tang, and S. Huang, Understanding human gaze communication by spatio-temporal graph reasoning, in Proc, *2019 IEEE/CVF International Conference on Computer Vision*, no. ICCV, p. 5723, 5732.
[116] H. Admoni and B. Scassellati, Social eye gaze in human-robot interaction: A review, *J. Hum. -Robot Int.*, vol. 6, no. 1, pp. 25–63, 2017.
[117] C. Moore, P. J. Dunham, and P. Dunham, *Joint Attention: Its Origins and Role in Development*, London, UK, Psychology Press, 2014.
[118] C. Moore and V. Corkum, Social understanding at the end of the first year of life, *Dev. Rev.*, vol. 14, no. 4, pp. 349–372, 1994.
[119] Y. Nagai, Understanding the development of joint attention from a viewpoint of cognitive developmental robotics, Ph. D. dissertation, Osaka University, Osaka, Japan, 2004.
[120] L. Fan, Y. Chen, P. Wei, W. Wang, and S. C. Zhu, Inferring shared attention in social scene videos, in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 6460–6468.
[121] I. Brinck, The pragmatics of imperative and declarative pointing, *Cogn. Sci. Quart.*, vol. 3, no. 4, pp. 429–446, 2004.
[122] E. Bates, L. Camaioni, and V. Volterra, The acquisition of performatives prior to speech, *Merrill-Palmer Quart.*, vol. 21, no. 3, pp. 205–226, 1975.
[123] S. C. Levinson, On the human "interaction engine", in *Roots of Human Sociality*, S. C. Levinson and N. J. Enfield, Eds. London: Routledge, 2020, pp. 39–69.
[124] F. Rossano, J. Terwilliger, A. Bangerter, E. Genty, R. Heesen, and K. Zuberbühler, How 2- and 4-year-old children coordinate social interactions with peers, *Phil. Trans. Roy. Soc. B*, vol. 377, no. 1859, p. 20210100, 2022.
[125] G. Pezzulo, The "interaction engine" : A common pragmatic competence across linguistic and nonlinguistic interactions, *IEEE Trans. Autonom. Mental Dev.*, vol. 4, no. 2, pp. 105–123, 2012.
[126] A. Cichocki and A. P. Kuleshov, Future trends for human-AI collaboration: A comprehensive taxonomy of AI/AGI using multiple intelligences and learning styles, *Comput. Intell. Neurosci.*, vol. 2021, p. 8893795, 2021.
[127] M. Tomasello, *Why We Cooperate*, Cambridge, MA, USA: MIT Press, 2009.
[128] N. Tang, S. Stacy, M. Zhao, G. Marquez, and T. Gao, Bootstrapping an imagined we for cooperation, in *Proc. 42nd Annu. Meeting of the Cognitive Science Society*, virtual, 2020, pp. 2453–2456.
[129] S. Stacy, Q. Zhao, M. Zhao, M. Kleiman-Weiner, and T. Gao, Intuitive signaling through an "imagined we", in *Proc. 42nd Annu. Meeting of the Cognitive Science Society*, virtual, 2020, p. 1880.
[130] S. E. T. Stacy, The imagined we: Shared Bayesian theory of mind for modeling communication, Ph. D. dissertation, University of California, Los Angeles, CA, USA, 2022.
[131] N. Tang, S. Gong, Z. Liao, H. Xu, J. Zhou, M. Shen, and T. Gao, Jointly perceiving physics and mind: Motion, force and intention, in *Proc. 43rd Annu. Meeting of the Cognitive Science Society*, Vienna, Austria, 2021, pp. 735–741.
[132] T. Shu, M. Kryven, T. D. Ullman, and J. Tenenbaum, Adventures in flatland: Perceiving social interactions under physical dynamics, in *Proc. 42nd Annu. Meeting of the Cognitive Science Society*, virtual, 2020, pp. 2901–2907.
[133] T. Shu, Y. Peng, L. Fan, H. Lu, and S. C. Zhu, Perception of human interaction based on motion trajectories: From aerial videos to decontextualized animations, *Top. Cogn. Sci.*, vol. 10, no. 1, pp. 225–241, 2018.
[134] T. Gao, C. L. Baker, N. Tang, H. Xu, and J. B. Tenenbaum, The cognitive architecture of perceived animacy: Intention, attention, and memory, *Cogn. Sci.*, vol. 43, no. 8, p. e12775, 2019.
[135] P. Wei, Y. Liu, T. Shu, N. Zheng, and S. C. Zhu, Where and why are they looking? Jointly inferring human attention and intentions in complex tasks, in *Proc. 2018 IEEE/CVF Conf. Computer Vision and Pattern Recognition*, Salt Lake City, UT, USA, 2018, pp. 6801–6809.
[136] D. Xie, T. Shu, S. Todorovic, and S. C. Zhu, Learning and inferring "dark matter" and predicting human intents and trajectories in videos, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 7, pp. 1639–1652, 2018.
[137] S. Holtzen, Y. Zhao, T. Gao, J. B. Tenenbaum, and S. C. Zhu, Inferring human intent from video by sampling hierarchical plans, in *Proc. 2016 IEEE/RSJ Int. Conf. Intelligent Robots and Systems* (*IROS*), Daejeon, Republic of Korea, 2016, pp. 1489–1496.
[138] B. González and L. J. Chang, Computational models of mentalizing, in *The Neural Basis of Mentalizing*, M. Gilead and K. N. Ochsner, Eds. Cham, Switzerland: Springer, 2021, pp. 299–315.
[139] W. Yoshida, R. J. Dolan, and K. J. Friston, Game theory of mind, *PLoS Comput. Biol.*, vol. 4, no. 12, p. e1000254, 2008.
[140] S. V. Albrecht and P. Stone, Autonomous agents modelling other agents: A comprehensive survey and open problems, *Artif. Intell.*, vol. 258, pp. 66–95, 2018.
[141] S. Arora and P. Doshi, A survey of inverse reinforcement learning: Challenges, methods and progress, *Artif. Intell.*, vol. 297, p. 103500, 2021.
[142] C. L. Baker, R. Saxe, and J. B. Tenenbaum, Bayesian theory of mind: Modeling joint belief-desire attribution, in *Proc. 33rd Annu. Meeting of the Cognitive Science Society*, Boston, MA, USA, 2011, pp. 2469–2474.
[143] T. Yuan, H. Liu, L. Fan, Z. Zheng, T. Gao, Y. Zhu, and S. C. Zhu, Joint inference of states, robot knowledge, and human (false-) beliefs, in *Proc. 2020 IEEE Int. Conf. Robotics and Automation* (*ICRA*), Paris, France, 2020, pp. 5972–5978.
[144] L. Fan, S. Qiu, Z. Zheng, T. Gao, S. C. Zhu, and Y. Zhu, Learning triadic belief dynamics in nonverbal communication from videos, in *Proc. 2021 IEEE/CVF Conf. Computer Vision and Pattern Recognition* (*CVPR*), Nashville, TN, USA, 2021, pp. 7308–7317.
[145] B. Arslan, N. A. Taatgen, and R. Verbrugge, Five-year-olds' systematic errors in second-order false belief tasks are due to first-order theory of mind strategy selection: A computational modeling study, *Front. Psychol.*, vol. 8, p. 275, 2017.
[146] I. Oguntola, D. Hughes, and K. Sycara, Deep interpretable models of theory of mind, in *Proc. 2021 30th Int. Conf. Robot and Human Interactive Communication* (*RO-MAN*), Vancouver, Canada, 2021, pp. 657–664.
[147] Y. Zeng, Y. Zhao, T. Zhang, D. Zhao, F. Zhao, and E. Lu, A brain-inspired model of theory of mind, *Front. Neurorobot.*, vol. 14, p. 60, 2020.
[148] C. L. Baker, J. Jara-Ettinger, R. Saxe, and J. B. Tenenbaum, Rational quantitative attribution of beliefs, desires and percepts in human mentalizing, *Nat. Hum. Behav.*, vol. 1, no. 4, p. 0064, 2017.
[149] Y. Wen, Y. Yang, R. Luo, J. Wang, and W. Pan, Probabilistic recursive reasoning for multi-agent reinforcement learning, arXiv: 1901.09207, 2019.
[150] P. Moreno, E. Hughes, K. R. McKee, B. A. Pires, and T. Weber, Neural recursive belief states in multi-agent reinforcement learning, arXiv preprint arXiv: 2102.02274, 2021.
[151] A. Hakimzadeh, Y. Xue, and P. Setoodeh, Interpretable reinforcement learning inspired by piaget' s theory of cognitive

development, arXiv preprint arXiv: 2102.00572, 2021.
[152] J. Jara-Ettinger, Theory of mind as inverse reinforcement learning, *Curr. Opin. Behav. Sci.*, vol. 29, pp. 105–110, 2019.
[153] L. Yuan, X. Gao, Z. Zheng, M. Edmonds, Y. N. Wu, F. Rossano, H. Lu, Y. Zhu, and S. C. Zhu, In situ bidirectional human-robot value alignment, *Sci. Robot.*, vol. 7, no. 68, p. eabm4183, 2022.
[154] H. de Weerd, R. Verbrugge, and B. Verheij, Negotiating with other minds: The role of recursive theory of mind in negotiation with incomplete information, *Auton. Agents Multi-Agent Syst.*, vol. 31, no. 2, pp. 250–287, 2017.
[155] H. de Weerd, D. Diepgrond, and R. Verbrugge, Estimating the use of higher-order theory of mind using computational agents, *B. E. J. Theor. Econ.*, vol. 18, no. 2, p. 20160184, 2018.
[156] A. Kanwal, W. M. Qazi, M. A. Altaf, A. Athar, M. Hussain, S. T. S. Bukhari, and A. T. Apasiba, A step towards the development of socio-cognitive agent, *Lahore Garrison Univ. Res. J. Comput. Sci. Informat. Technol.*, vol. 4, no. 3, pp. 23–38, 2020.
[157] R. Tejwani, Y. L. Kuo, T. Shu, B. Stankovits, D. Gutfreund, J. B. Tenenbaum, B. Katz, and A. Barbu, Incorporating rich social interactions into MDPs, in *Proc. 2022 Int. Conf. Robotics and Automation* (*ICRA*), Philadelphia, PA, USA, 2022, pp. 7395–7401.
[158] R. Tejwani, Y. L. Kuo, T. Shu, B. Katz, and A. Barbu, Social interactions as recursive MDPs, in *Proc. 5th Conf. Robot Learning*, London, UK, 2022, pp. 949–958.
[159] A. Panella and P. Gmytrasiewicz, Interactive POMDPs with finite-state models of other agents, *Auton. Agents Multi-Agent Syst.*, vol. 31, no. 4, pp. 861–904, 2017.
[160] M. Zhao, N. Tang, A. L. Dahmani, Y. Zhu, F. Rossano, and T. Gao, Sharing rewards undermines coordinated hunting, *J. Comput. Biol.*, vol. 29, no. 9, pp. 1022–1030, 2022.
[161] S. Stacy, C. Li, M. Zhao, Y. Yun, Q. Zhao, M. Kleiman-Weiner, and T. Gao, Modeling communication to coordinate perspectives in cooperation, in *Proc. 43rd Annu. Meeting of the Cognitive Science Society*, Vienna, Austria, 2021, pp. 1851–1857.
[162] X. Gao, R. Gong, Y. Zhao, S. Wang, T. Shu, and S. Chun. Zhu, Joint mind modeling for explanation generation in complex human-robot collaborative tasks, in *Proc. 2020 29th IEEE Int. Symp. Robot and Human Interactive Communication* (*RO-MAN*), Naples, Italy, 2020, pp. 1119–1126.
[163] M. C. Buehler, J. Adamy, and T. H. Weisswange, Theory of mind based assistive communication in complex human robot cooperation, arXiv preprint arXiv: 2109.01355, 2021.
[164] Y. Wang, F. Zhong, J. Xu, and Y. Wang, ToM2C: Target-oriented multi-agent communication and cooperation with theory of mind, arXiv: 2111.09189, 2022.
[165] J. Pöppel, S. Kahl, and S. Kopp, Resonating minds-emergent collaboration through hierarchical active inference, *Cogn. Comput.*, vol. 14, no. 2, pp. 581–601, 2022.
[166] V. Chidambaram, Y. H. Chiang, and B. Mutlu, Designing persuasive robots: How robots might persuade people using vocal and nonverbal cues, in *Proc. 2012 7th ACM/IEEE Int. Conf. Human-Robot Interaction* (*HRI*), Boston, MA, USA, 2012, pp. 293–300.
[167] O. Nocentini, L. Fiorini, G. Acerbi, A. Sorrentino, G. Mancioppi, and F. Cavallo, A survey of behavioral models for social robots, *Robotics*, vol. 8, no. 3, p. 54, 2019.
[168] T. J. Wiltshire, S. F. Warta, D. Barber, and S. M. Fiore, Enabling robotic social intelligence by engineering human social-cognitive mechanisms, *Cogn. Syst. Res.*, vol. 43, pp. 190–207, 2017.
[169] J. E. Laird, Introduction to soar, arXiv preprint arXiv: 2205.03854, 2022.
[170] A. Lieto, M. Bhatt, A. Oltramari, and D. Vernon, The role of cognitive architectures in general artificial intelligence, *Cogn. Syst. Res.*, vol. 48, pp. 1–3, 2018.
[171] M. Vircikova, G. Magyar, and P. Sincak, The affective loop: A tool for autonomous and adaptive emotional human-robot interaction, in *Robot Intelligence Technology and Applications 3*, J. H. Kim, W. Yang, J. Jo, P. Sincak, and H. Myung, Eds. Cham, Switzerland: Springer, 2015, pp. 247–254.
[172] J. Snaider, R. McCall, and S. Franklin, The LIDA framework as a general tool for AGI, in *Proc. 4th Int. Conf. Artificial General Intelligence*, Mountain View, CA, USA, 2011, pp. 133–142.
[173] J. E. Laird, *The Soar Cognitive Architecture*, Cambridge, MA, USA: MIT Press, 2019.
[174] J. R. Anderson, C. Lebiere, M. Lovett, and L. Reder, ACT-R: A higher-level account of processing capacity, *Behav. Brain Sci.*, vol. 21, no. 6, pp. 831–832, 1998.
[175] J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin, An integrated theory of the mind, *Psychol. Rev.*, vol. 111, no. 4, pp. 1036–1060, 2004.
[176] C. Breazeal, J. Gray, and M. Berlin, An embodied cognition approach to mindreading skills for socially intelligent robots, *Int. J. Robot. Res.*, vol. 28, no. 5, pp. 656–680, 2009.
[177] W. G. Kennedy, M. D. Bugajska, A. M. Harrison, and J. G. Trafton, "Like-me" simulation as an effective and cognitively plausible basis for social robotics, *Int. J. Soc. Robot.*, vol. 1, no. 2, pp. 181–194, 2009.
[178] C. Moulin-Frier, T. Fischer, M. Petit, G. Pointeau, J. Y. Puigbo, U. Pattacini, S. C. Low, D. Camilleri, P. Nguyen, M. Hoffmann, et al., Dac-H3: A proactive robot cognitive architecture to acquire and express knowledge about the world and the self, *IEEE Trans. Cogn. Dev. Syst.*, vol. 10, no. 4, pp. 1005–1022, 2018.
[179] A. M. Franchi, F. Mutti, and G. Gini, From learning to new goal generation in a bioinspired robotic setup, *Adv. Robot.*, vol. 30, no. 11-12, pp. 795–805, 2016.
[180] A. Netanyahu, T. Shu, B. Katz, A. Barbu, and J. B. Tenenbaum, PHASE: Physically-grounded abstract social events for machine social perception, arXiv: 2103.01933, 2021.
[181] T. Shu, A. Bhandwaldar, C. Gan, K. A. Smith, S. Liu, D. Gutfreund, E. Spelke, J. B. Tenenbaum, and T. D. Ullman, AGENT: A benchmark for core psychological reasoning, arXiv: 2102.12321, 2021.
[182] X. Puig, T. Shu, S. Li, Z. Wang, Y. H. Liao, J. B. Tenenbaum, S. Fidler, and A. Torralba, Watch-and-help: A challenge for social perception and human-AI collaboration, arXiv: 2010.09890, 2021.
[183] M. Sap, H. Rashkin, D. Chen, R. LeBras, and Y. Choi, SocialiQA: Commonsense reasoning about social interactions, arXiv preprint arXiv: 1904.09728, 2019.
[184] N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, et al., The Hanabi challenge: A new frontier for AI research, *Artif. Intell.*, vol. 280, p. 103216, 2020.
[185] H. Shevlin, K. Vold, M. Crosby, and M. Halina, The limits of machine intelligence, *EMBO Rep.*, vol. 20, no. 10, p. e49177, 2019.
[186] F. Lievens and D. Chan, Practical intelligence, emotional intelligence, and social intelligence, in *Handbook of Employee Selection*, J. L. Farr and N. T. Tippins, Eds. New York, NY, USA: Routledge, 2017, pp. 342–364.
[187] M. Crosby, B. Beyret, and M. Halina, The animal-AI Olympics, *Nat. Mach. Intell.*, vol. 1, no. 5, pp. 257–257, 2019.
[188] M. Schurz, J. Radua, M. Aichhorn, F. Richlan, and J. Perner, Fractionating theory of mind: A meta-analysis of functional brain imaging studies, *Neurosci. Biobehav. Rev.*, vol. 42, pp. 9–34, 2014.
[189] L. Smith and M. Gasser, The development of embodied cognition: Six lessons from babies, *Artif. Life*, vol. 11, nos. 1–2, pp. 13–29, 2005.
[190] G. M. van de Ven and A. S. Tolias, Three scenarios for continual learning, arXiv preprint arXiv: 1904.07734, 2019.
[191] R. Caruana, Multitask learning, *Mach. Learn.*, vol. 28, no. 1, pp. 41–75, 1997.
[192] Y. Zhang and Q. Yang, A survey on multi-task learning, *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 12, pp. 5586–5609, 2021.
[193] J. Vanschoren, Meta-learning: A survey. arXiv preprint arXiv: 1810.03548, 2018.