Article scientifique | **Article** | **2023** | | **Published version** | **Open Access**

Combining the Projective Consciousness Model and Virtual Humans for Immersive Psychological Research: A Proof-of-concept Simulating a ToM Assessment

---

Rudrauf, David; Sergeant-Perthuis, Grégoire; Tisserand, Yvain; Monnor, Teerawat; Durand De Gevigney, Valentin; Belli, Olivier

This publication URL: https://archive-ouverte.unige.ch/unige:168784
Publication DOI: 10.1145/3583886

Combining the Projective Consciousness Model and Virtual Humans for Immersive Psychological Research: A Proof-of-concept Simulating a ToM Assessment

D. RUDRAUF, FPSE, Section of Psychology, Swiss Center for Affective Sciences, and Computer Science University Center, University of Geneva, Switzerland

G. SERGEANT-PERHTUIS, Swiss Center for Affective Sciences, University of Geneva, Switzerland

Y. TISSERAND and T. MONNOR, FPSE, Section of Psychology and Swiss Center for Affective Sciences, University of Geneva, Switzerland

V. DE GEVIGNEY, Independent Researcher, France

O. BELLI, Evolutio, Switzerland

8

Relating explicit psychological mechanisms and observable behaviours is a central aim of psychological and behavioural science. One of the challenges is to understand and model the role of consciousness and, in particular, its subjective perspective as an internal level of representation (including for social cognition) in the governance of behaviour. Toward this aim, we implemented the principles of the Projective Consciousness Model (PCM) into artificial agents embodied as virtual humans, extending a previous implementation of the model. Our goal was to offer a proof-of-concept, based purely on simulations, as a basis for a future methodological framework. Its overarching aim is to be able to assess hidden psychological parameters in human participants, based on a model relevant to consciousness research, in the context of experiments in virtual reality. As an illustration of the approach, we focused on simulating the role of Theory of Mind (ToM) in the choice of strategic behaviours of approach and avoidance to optimise the satisfaction of agents' preferences. We designed a main experiment in a virtual environment that could be used with real humans, allowing us to classify behaviours as a function of order of ToM, up to the second order. We show that agents using the PCM demonstrated expected behaviours with consistent parameters of ToM in this experiment. We also show that the agents could be used to estimate correctly each other's order of ToM. Furthermore, in a supplementary experiment, we demonstrated how the agents could simultaneously estimate order of ToM and preferences attributed to others to optimize behavioural outcomes. Future studies will empirically assess and fine tune the framework with real humans in virtual reality experiments.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; User models; • **Computing methodologies** → **Multi-agent systems**; *Multi-agent planning*; **Theory of mind**; **Intelligent agents**; **Cooperation and coordination**;

---

1 INTRODUCTION

Human psychology entails highly complex information processing. This processing plays a causal role in the generation of behaviours. Modelling such complexity to simulate human experience and behaviours, and predict outcomes of experimental research, is important for the development of psychological and behavioural science. An outstanding issue, and theoretical, methodological and technical challenge, is to understand and model how consciousness, and in particular its subjective perspective, may contribute to this process.

The approach developed in this report stems from the general rationale that computational models of human psychology should, among other requirements, strive to be:

(1) Integrative, i.e., targeting a comprehensive model of the human mind, including simulations of mechanisms related to consciousness and its subjective perspective, and their relations to perceptual, affective and social cognitive processes,
(2) Generative of embodied states and behaviours that could be measured in human participants in well-controlled and effective experimental and observation contexts, such as **Virtual Reality (VR)**;
(3) Capable of making inferences to assess internal psychological parameters in others based on their behaviours and through interactions with them, as a function of model parameters;

Furthermore, our approach leverages the possibility of using computational models embodied as Virtual Humans:

(1) to serve as artificial confederates interacting with other virtual and/or real humans in the context of social psychology experiments;
(2) to explore and test hypotheses about the mechanisms underlying observable behaviours (the question addressed with the approach would be, for instance: Which parameters can make my model behave and perform in a task as human would?);
(3) to serve as a tool to generate model-based psychological profiles and assessments (the question addressed with the approach would be, for instance: Can my model predict and explain observed interindividual differences in behaviour and performance in a given task?).

One of our motivations is to work toward a modeling framework that could claim cognitive plausibility, based on sound psychological and behavioural principles, in addition to demonstrate predictive power (see also [1] about this issue).

Here, we present, in a preliminary manner, the approach we are developing for this purpose based on the **Projective Consciousness Model (PCM)**. The model of consciousness is used as a level of processing and control for the generation of meaningful behaviours. Our goal in this report is to present a proof-of-concept of the approach. We used a limited example targeting the assessment of **Theory of Mind (ToM)** in a simple entry game, based on simulations of virtual humans in three-dimensional environments that could be used to run experiments in VR. The model implementation is based on a previously published implementation [2], which we extended to incorporate new mechanisms of inference of others' order of ToM, and combined inference

of ToM order and preferences (see Section 3.2). We used a mock-up experiment designed for the purpose of demonstrating: (1) that our model can generate behaviours that would be expected from humans in a ToM task that could run in VR, and (2) that it is able to estimate the ToM parameters driving the behaviours of another artificial agent in the same task, which could be replaced by a real human in the context of an actual experiment. We also illustrates how the approach can be extended to more challenging tasks in which both preferences of another agents and its ToM parameters have to be inferred. Likewise, we show how parameters of preferences influence behavioural outcomes, notably to illustrate the robustness of the model to variations in parameters.

2 BACKGROUND AND RATIONALE

2.1 Immersive Environments and Virtual Agent Modeling for Psychological Research: General Considerations

Immersive virtual environment technologies have been proposed as a promising tool for social psychological research, capable of mitigating issues with experimental control-mundane realism trade-off, lack of replication, and non-representative samples [3]. In this perspective, Virtual Confederates, e.g., real humans embodied as virtual avatars, have been used as a research tool for overcoming limitations of real interactions with human confederates and paper-and-pencil designs [4]. Likewise, virtual humans in gamified environments have been applied, sometimes in combination with machine learning analysis of human participants' responses, to the screening of PTSD and other psychiatric disorders, through verbal interviews and the analysis of both verbal and non-verbal cues [5, 6], as well as to practicing negotiations [7].

While certain approaches depart from game-theoretic frameworks [6], others have proposed to leverage such frameworks combined with computational modeling of agents [8]. The resulting models integrate social utility functions, reasoning about iterative thinking limits, statistical approaches, and consider situations in which a player choice affects the payoff of other players. This payoff entails a complex process of mutual influence and inference. Among different challenges for the development of models, a central one is the integration of simulations of mental representations that capture internal processes as they operate in human psychology: "Theorists analyze games in the form of matrices or trees but players presumably construct internal representations that might barely resemble matrices or trees" [8]. We hold that this issue also entails to understand and model consciousness.

2.2 Consciousness Theories: The Problem of the Subjective Perspective

Much cognitive processing is unconscious and consciousness is only the tip of the iceberg [9–11]. Nevertheless, it remains a central component of human information processing, and it is thus important to integrate models of consciousness and its impact on decision-making in models that wish to mimic human processing.

Theories and models of consciousness developed over the past three decades encompass five broad, non-mutually exclusive conceptual frameworks [12]: integrated information theories [13], global workspace theories [14, 15], internal self-model theories [16], higher-level representations, and attention mechanisms; the two first frameworks being the most prominent.

Overall, it can be said that consciousness operates as a global workspace [14, 15]. Such workspace features limited capacity. It accesses multimodal information and integrates it with information from memory. It integrates mechanisms of uncertainty monitoring and reduction, and error corrections. It is used for non-social and social imaginary simulations and appraisal of outcomes. Its overall function is to perform planning, decision-making and action programming, in a serial manner. Along these lines, five "axioms" have been proposed for artificial models of

consciousness by Aleksander [17]: (1) presence, including mechanisms for representing the situated individual within the world; (2) imagination, or internal simulations of action without sensory input; (3) attention, to guide perception and modulate imagination; (4) planning, through the imaginary exploration of possible actions; (5) emotion as part of a mechanism of appraisal of plans and behavioural outcomes (which could be related to the states of an agent or to states it infers in others). Furthermore, consciousness integrates a representation of the body in space in relation to its environment, playing a role in homeostasis, survival, and well-being, and relying on embodied appraisal and emotion [16, 18, 19].

It remains largely unaddressed however, in particular from a modeling standpoint, how information is accessed, shaped and exploited through the global workspace of consciousness to accomplish the functions ascribed to consciousness. The issue directly relates to another essential axiom about models of consciousness that could be added to Aleksander's list: its qualitative experience or subjective character [20, 21]. For long, consciousness research has emphasized the phenomenologically pervasive and central role of a "subjective perspective," conceived of as a nontrivial, viewpoint-dependent, unified, embodied, internal representation of the world in perspective [13, 22–26]. In such representation, contents appears in a three-dimensional non-Euclidean perspectival manner that could play a role in appraisal and departs from the more Euclidean objective environment in which consciousness is embodied (see Reference [2]). One of its functional roles would be to enable conscious systems to take different perspectives through imagination or action, to evaluate affordances and maximize utility in a context-dependent manner [26–28]. The subjective structure in question would entail the combination of cognitive (spatial) and affective representations, for action programming [29, 30]. In complex social animals, perspective taking is also pivotal to perform ToM [31]. Understanding and operationalizing how such subjective perspective may participate in the process of information integration and behavioural control carried out by consciousness is an important challenge for consciousness modeling [20, 21, 32–35]. While acknowledging the importance of the issue, many have decided to set it aside [15, 36]. Others have proposed to address the issue based purely on information theoretic concepts, but largely fail to capture the phenomenon explicitly and in a specific manner as a result [13, 37].

2.3 The Projective Consciousness Model and Active Inference

The PCM [2, 26, 38–40] aims at explicitly tackling the problem of the subjective perspective of consciousness and its role in active inference.

Active inference conceptualizes the operation of the mind as a recursive cycle, including two main steps: (1) the inference of the causes of sensory information, (2) the planning of action. Resulting action outcomes provide a new context for the next cycle [41]. For instance, an agent S (the subject) perceiving anger on the face of another agent O while being looked at by O, might infer that it is disliked by O. It might then consider that approaching O could be dangerous, while avoiding it could be safer, and choose to move away. After moving away, S might realize that O is actually smiling in a friendly manner while looking at the agent. In the next cycle, S might then infer that the expression of anger was intended as a joke, and decide that it would actually be safe and pleasant to approach O.

Active inference has been formulated within variational optimization approaches, such as the **Free-energy Principle (FEP)**, which approximates Bayesian inference based on the minimization of a free energy acting as a cost function [41–45]. Free energy is an upper-bound on surprise as a deviation between prior expectations and sensory evidence. Importantly, prior beliefs in an agent performing active inference can include models of preferences and desires, encoded as expectations. In keeping with the example above, the expectation that O would not be pleased if S approached it, would raise the expected free energy of S if it were to approach O. However, it would

lower its free energy if S envisioned to move away instead. The approach has shown promise to understand the emergence of affective, affiliative, and communicative behaviours [40, 46–50].

The PCM is based on two main principles. (1) Consciousness is central to active inference in humans. (2) Consciousness integrates information in a viewpoint-dependent manner, within a **Field of Consciousness (FoC)** in perspective. The concept is based on the fact that our integrative conscious experience of space, as informed by multimodal sensory information and memory, is that of a three-dimensional space in perspective, which depends on the adopted point of view, both in perception and imagination (through imaginary perspective taking). The FoC is governed by three-dimensional projective geometry [26], which is the geometry of perspective. It acts on this basis as a global workspace [14, 15] for the integration of information and the planning of action.

We recently showed how the PCM could explain and predict perceptual illusions such as the Moon Illusion, based on the calibration of a three-dimensional projective chart under **free energy (FE)** minimisation, combining simulations and VR [39] (see also Reference [26]). But the FoC is thought to play a much broader role beyond perceptual experience [2]. It corresponds to a three-dimensional projective space, representing, within a subjective frame in perspective, an internal world model. One of its function according to the theory is to assess the distribution of affective and epistemic values ascribed to entities and actions in that world, as a function of perspectives being taken. It orders entities according to a point of view, modulating their apparent size and thus relative importance. For instance, back to the example above, after moving away, S would perceive O as smaller from the distance, and thus O would occupy less of S's FoC than if it were closer. That difference would make the perceived negative attitude of O less important and impactful in terms of affective value. (Of course, since relative apparent size depends on relative distances between agents, if O would have moved toward S, then its apparent size from the perspective of S would not have decreased as much, and it would even have increased if it had gotten closer to S. In this case, S would have to revise its understanding of the situation and its choice of behaviour, e.g., running further away, or revising the preferences and intentions it attributes to O). At the same time, with the increased distance of S from O, there would be more sensory uncertainty for S about the state of O, reducing the epistemic value of the current perspective (see Reference [2] for details). The FoC can take multiple perspectives on the world model using projective transformations. For instance, S could imagine that if it were closer to O than it actually is and could better see its face, it might turn out that O's face was not expressing anger but a state of concentration. Free energy can be expressed as a function of the FoC and thus of perspective taking, and reflect perceived or expected deviations from preferred values, as well as uncertainty with respect to sensory evidence. The idea is that its minimization through recursive imaginary perspective taking should make agents search for perspectives on the world that maximize their preferences and minimize uncertainty. This process would provide the agents with possible paths of action, as perspective changes relate to parameters of motion. We recently applied these principles to simulate complex adaptive and maladaptive behaviours among artificial agents, in a robotic context [2]. The agents embedded multi-agent models to infer the preferences of other agents based on their emotion expression and orientation, and to simulate other agents' FoC, according to projective transformations that respected known psychophysical laws. The process enabled the agents to appraise and predict other agents' behaviours, to generate affective (valence-related) and epsitemic (curiosity-related) drives toward preferred states, and plan their action accordingly.

2.4 Theory of Mind

One interesting issue is to undersand how the subjective perspective of consciousness could play a role in ToM. ToM is the ability to infer others' mental states, beliefs, and desires and to predict

their behaviours, for instance, for strategic planning, and it relies on the integration and imaginary manipulation of cognitive and affective information [30, 51–53].

ToM is often conceptualized within simulation theory, which entails that humans use their own cognitive and affective appartus to imagine themselves in the position of others and simulate their subjective experience and likely behaviours; a process that would underline empathy [54, 55]. ToM through perspective taking is considered as important for emotion regulation and social-affective development [56–62]. It entails a balance between reward-expectation and the cost of executive function [63].

Different levels of ToM have been distinguished [64]. *Level*-1 and *level*-2, respectively, correspond to what is also described in the literature as **Visual perspective taking 1 (VPT1)** and **Visual perspective taking 2 (VPT2)** [65]: the ability to infer, respectively, whether an object can be seen from a given point of view, and whether the object would look different from different points of view. *Level*-3 concerns the understanding that knowing requires verification through direct sensory evidence (or uncertainty reduction). *Level*-4 and *level*-5 concern the ability to understand, respectively, true and false beliefs in others, to predict their behaviour. *Level*-6 concerns the ability to understand that others can themselves perform ToM. Levels 1–5 correspond to so-called *first-order* ToM, and *Level*-6 to *second-order* ToM. The notion of order of ToM can be generalized recursively to *third-order* ToM, i.e., the ability to understand that others can perform ToM about the ToM of others, and so on to *n-order* ToM, up to the maximal capacity of an individual.

ToM is often assessed through verbal tasks [66, 67], which may be susceptible to different biases, and it is important to develop non-verbal tasks assessing ToM based on outcome behaviours [68].

Following these principles, we implemented ToM in PCM-agents using the FoC to simulate others' subjective states and demonstrated that a variety of meaningful adaptive and maladaptive behaviours would ensue as a function of psychologically relevant parameters (see Reference [2]; see also Section 3 in this report).

2.5 Theory of Mind in Models of Agents and Game Theory: Perspectives for Immersive Approaches

Concepts of ToM can be found in classical **Belief-Desire-Intention (BDI)** models of agents [69], which themselves entail embedded appraisal models [70].

ToM has become of interest in game theory [71] to understand and model rationale and irrational strategic planning in human and non-human competitive contexts [72, 73]. It relates to strategic uncertainty in the face of social situations with dependence on others' choices. It entails assessing subjective probabilities based on others' behaviours, and the interaction between processes of decision-making under risk and higher order beliefs about others [74].

Models of rationale, multi-agent coordination in unpredictable environments, applied to two-dimensional strategic games, have been introduced [75, 76]. These models incorporated inferences about ToM, and took into account decision under uncertainty about others' beliefs, based on recursive nesting of models (or Recursive Modeling Methods). The aim was to maximize expected utility, with mechanisms of belief update maximizing predictive power about observed behaviours of other agents. Likewise, simple Bayesian models of inferences of agents' intentions have been introduced to predict navigation in mazes, based on probabilistic inverse planning for action understanding [77]. Recursive modeling of multi-agent interactions under uncertainty have been investigated, with the overarching aim of exploring possible outcomes of intervention strategies through simulations, e.g., in the context of bullying [78]. The approach used internal models of other agents that integrated models of their preferences, and a variety of social influence factors and biases (such as consistency, self-interest, speaker's self-interest, trust, likability, and affinity).

Of note, in practice, Bayesian models as such may be limited by the curse of dimensionality that makes it difficult to compute exact posterior distributions. Variational methods exist to mitigate this problem, which make it more tractable.

Models based on adaptive control theory and probabilistic learning of agents have been investigated, using classical game theory paradigms and game-theoretic metrics, as an alternative to approaches such as Deep Learning [73]. The latter approaches may yield good predictive power, but are based on distributed models that may be difficult to interpret, as they operate in practice as a black-box. Models with parameters that can be interpreted along psychological dimensions are warranted for many applications and psychological science. For instance, Yoshida et al. [79] proposed an approach, confronting simulations and empirical data, to assess orders of theory of mind implied by behaviours, in a two-dimensional competitive digital board game between two agents. The simulations were based on a model inspired from **Reinforcement Learning (RL)**, recursively evaluating strategic predictions as a function of orders of ToM by maximizing expected future reward. The approach was then used to assess ToM capacities in Autism Spectrum Disorder participants [80].

One general issue is to devise innovative methods capable of adapting and performing in a variety of situations and contexts, which are not always specifically designed for a given game theoretical paradigm and associated metrics. Such methods should be able to simulate, predict and assess behaviours, such as approach and avoidance, joint attention, emotion expressions, or more generally, navigation in a three-dimensional world. They should do so as a function of cognitive and affective processes such as ToM, in a more ecological manner than non-immersive approaches, i.e., in a manner that is closer to real-world human behaviours as observed in the field. A promising approach is to combine computational models of agents and virtual reality, for instance, to assess ToM capacities through simulations and embodied interactions.

As hinted in Section 2.1 above, VR offers a promising framework to study social interactions through immersive technologies [81]. Virtual environments may be shared by human participants and virtual humans [82, 83]. Narang et al. [84] developed an approach combining a Bayesian model of ToM applied to artificial agents and VR. This approach was used to infer the intention of action of human participants in VR, and to control the navigation and approach-avoidance behaviours of virtual humans in social crowds. The model was a simple model, making inferences based on observed proxemics and gaze-based cues.

Importantly, none of the approaches reviewed above did aim at modeling internal mechanisms of representation that would integrate a model of the subjective perspective of consciousness. Such modeling is important for consciousness research and, more generally, for psychological and behavioural research in humans.

3 MODEL

3.1 Presentation of the Model

The model we present, as a proof-of-concept, is an implementation of the PCM principles close to the implementation we introduced in Reference [2]. The approach has similarities with Reference [79], but instead of considering the expectation of a reward function for multiple agents, we consider a mean of free-energy quantities within each agent that takes into consideration the simulation of active inference in other agents. In other words, each agent embeds a multi-agent model or system [85] to simulate others. Agents must thus be able to reverse infer preferences and the order of ToM of other agents, while in Reference [79], it was the policy of other agents that was inferred. More generally, our approach is quite close to Recursive Modeling Methods that have been proposed in similar contexts, and include mechanisms of inferences about preferences of others

and factors of social influences [75, 76, 78]. The innovation is that we integrate an explicit model of the three-dimensional subjective perspective of consciousness in the process, which performs functions ascribed to consciousness based on view-point dependent subjective parameters (see Section 2.2 above). The model entails affective and epistemic (curiosity) drives based on projective geometrical mechanisms, and is applied to control virtual humans in virtual environments.

We extend the previous version of the model [2] with more advanced capacities of inferences, so that agents can infer preferences and ToM capacities in others, based on retrospective or prospective simulations of their behaviours, in a recursive manner. Predictions yielding best predictive power are used to update beliefs, in a manner that considers not only the emotion expression and orientation of other agents but also their relative behaviours of approach and avoidance, as indicators of interest labeled with affective valence (see Figure 1).

Each agent A_i computes projections about itself as subject S, and about other agents A_j, using the same basic processing pipeline. For a given state or move m_t, evaluated by the agent, the agent computes a projective chart $\psi(m_t)$, corresponding to the FoC it attributes to a given agent, including itself. Perceived value μ and uncertainty with respect to sensory evidence σ (given the current state of the agent) are computed based on $\psi(m_t)$ and the preferences attributed by A_i to the agent under consideration. These parameters are used to define a parametric probability distribution $P(\mu, \sigma)$, which is compared to an ideal distribution $P(\mu_0, \sigma_0)$ through the **Divergence of Kullback-Leibler (DKL)**. This yields a cost function that is sensitive to divergence from both preferences and uncertainty. Emotions are also expressed by the agents accordingly (not indicated in Figure 1). The process is repeated recursively to assess successive moves, according to the depth of processing used by the agent (large round arrow, top right in Figure 1). The algorithm entails a **Multi-agent System (MAS)** embedded within each agent. Multiple alternate sequences of moves M are computed, to define a series of anticipated states. The sequence of moves that the agent retains corresponds to that which minimizes its overall FE, taking or not into account anticipations about other agents states. The first move of the sequence is chosen by the agent as its actual move $m(S)$. That actual move controls the state of the associated virtual human $VH(S)$. The agent then takes as inputs the observed states in the world, including of other agents (locations, orientations, emotion expressions). If those states diverge above a certain threshold θ from the anticipated states, then a mechanism of reverse inference is triggered. Otherwise, the agent keep computing projections based on its current beliefs, including preferences, ToM parameters, and more generally, states (locations, orientations and emotion expressions of others). The mechanism of reverse inference tests different hypotheses about parameters such as preferences attributed to others and order of ToM used by others. It runs the same recursive algorithm used by the agent to simulate new projections, and retains the parameters that best explain the observed states to update its beliefs.

When referring to active inference, we mean the process of inferring and acting according to inference recursively, which can be summarized as follows. Let S be the space of sensory inputs and Γ the space of states that the agent can be in, and let M be the set of action the agent can perform. In the inference step, a state $\gamma \in \Gamma$ is induced by sensory input h by minimizing a cost function $c : S \times \Gamma \to \mathbb{R}$,

$$\gamma^* = \underset{\gamma \in \Gamma}{\operatorname{argmin}}\, c(h, \gamma), \tag{1}$$

and during the action selection step, the subject chooses the action according to a second cost function $c_1 : \Gamma \times M \to \mathbb{R}$,

$$m^* = \underset{m \in M}{\operatorname{argmin}}\, c_1(m, \gamma^*), \tag{2}$$

which in turn induces a change at the level of the sensory input, since the environment reacts to this action.
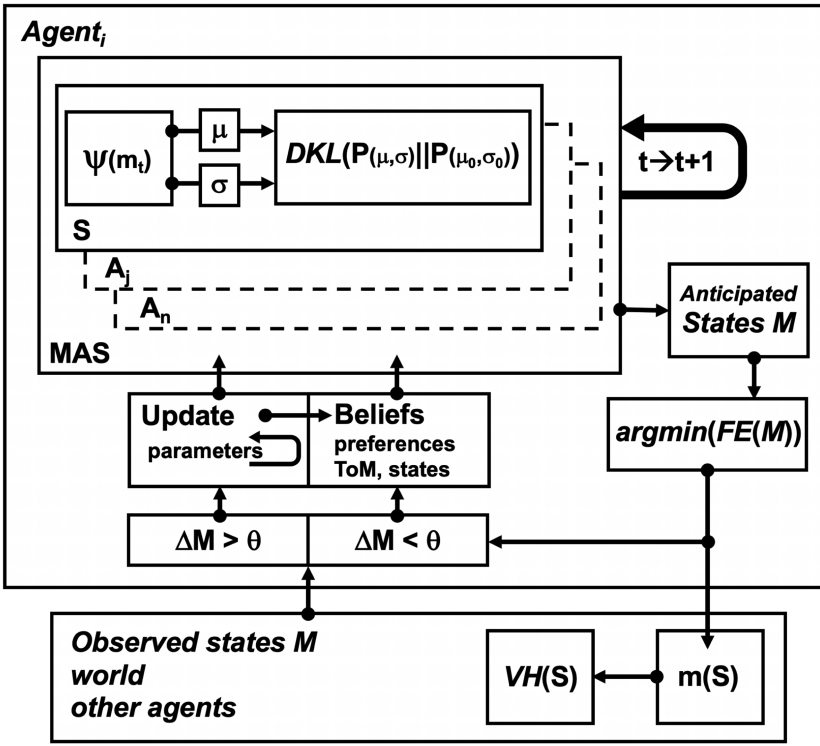
Fig. 1. Summary of the model architecture. Each agent computes representations of itself as subject *S*, and of other agents, using the same basic processing pipeline. It entails perspective taking based on a projective chart as a function of envisioned actions, and the computation of a Divergence of Kullback-Leibler (DKL), based on estimated perceived value and uncertainty. The process is repeated recursively to assess successive moves, according to the depth of processing used by the agent (large round arrow, top right). The sequence of moves that the agent retains corresponds to that which minimizes its overall free energy (FE), taking or not into account anticipations about other agents states. If the anticipated states and actual outcomes of action diverge above a certain threshold, then a mechanism of reverse inference is triggered. It runs the same recursive algorithm used by the agent to simulate new projections, and retains the parameters that best explain the observed states to update its beliefs. Otherwise, the agent keeps computing projections based on its current beliefs, including preferences, ToM parameters, and more generally, states (locations, orientations, and emotion expressions of others). (See text for details.)

In our setting, we consider a collection of entities, *E*, constituted of objects and agents. Agents express emotions and can infer and act according to their preferences and those ascribed to others, with respect to a situation, while objects cannot act. When singling out an agent, for example, when making explicit how active inference works for this agent, we will call it a subject. The space of agents will be denoted *A*. An agent $a \in A$ can express a positive emotion $e_+ \in [0, 1]$ and a negative emotion $e_- \in [0, 1]$. The space of sensory inputs of a subject is constituted of the configurations of other entities in the ambient space and the emotions that agents express. The space of states is the preferences it can have for other entities when the subject does only ToM-0 (that is no ToM in our context), and preferences attributed to other agents for higher order of ToM. Subjects act in two ways, they can move and express emotions.

The details of the following model we use are presented in Reference [2].

The preference for an entity is a real number in $[0, 1]$ denoted as p. Every subject, s, has an embodied perspective on the Euclidean ambient space that corresponds to a choice of a projective transformation that we denote as ψ_s; we will call it the projective chart associated to the agent. The quantity that links perspective taking and pleasantness of a situation is the perceived value μ that is computed for each entity $e \in E$ as

$$\mu = p\gamma \frac{v_p^{1/4}}{v_{tot}^{1/4}} + q_n \left(1 - \gamma \frac{v_p^{1/4}}{v_{tot}^{1/4}}\right), \tag{3}$$

where v_p is the perceived volume of the entity in the total FoC of the subject of volume v_{tot}. The perceived value μ is an average of the preference for the entity and a reference preference q_n weighted by the relative perceived volume of the entity; the power $1/4$ on the volume is taken to match documented psychophysical laws (see Reference [2] for a psychophysical and computational justification of this variable).

The subject also computes an uncertainty with respect to sensory evidence, denoted σ, that is greater with larger eccentricity with respect to the point of view of the subject, and the distance of the entity. In other words, there is more certainty about entities that appear actually or would be expected by imagination to be in front of and close to the subject.

The subject is driven toward an ideal with high perceived value and low uncertainty. To compute the divergence from this ideal, the perceived value and uncertainty are associated with a probability distribution, $Q(.|\mu, \sigma) \in \mathbb{P}([0, 1])$, centered in μ and of "width" σ. This divergence is computed with the Kullback-Leibler divergence of Q from the ideal distribution P narrowly centered on values close to 1. Let us recall that for any two probability distributions $P, Q \in \mathbb{P}(\Omega)$, over a space Ω, with $dQ = f\,dP$,

$$\text{DKL}(Q\|P) = \int f \ln f\,dP. \tag{4}$$

Let us now detail the active inference cycles of subjects with **ToM of order** 0 **(ToM-0)** to **ToM of order** 2 **(ToM-2)**. Here, we shall not focus on the inference part of the process nor on emotion expression but rather on how agents select their moves, one can refer to Reference [2] for a detailed presentation on how preferences are updated and emotion expressed.

The preferences of a subject for the other entities with ToM-0 is encoded in a vector $(q_e, e \in E)$. The configuration of an entity, e, is a subset of $\mathbb{R}^3$ denoted as $X_e \subseteq \mathbb{R}^3$ and the collection of configurations will be denoted as X. The subject chooses its move m from a set of moves M by minimizing the following average of Kullback-Leibler divergences,

$$C_0(m, X, q) = \sum_{\substack{e \in E \\ e \neq s}} \frac{1}{|E| - 1} \text{DKL}(Q(.|\mu_{\psi(m),q}(X_e), \sigma_{\psi(m),q}(X_e)\|P). \tag{5}$$

When the subject performs T0M-1 (ToM of order 1), it has a preference matrix $(p_{sae} \in [0, 1], a \in A, e \in E)$ that encodes preferences that agents have with respect to other entities according to the subject. The true preferences of the subject, i.e., the preference vector of s is $p_{ss.}$. Agents may be influenced by other agents in the way they infer preferences and the way they act, this is encoded by the influence vector on preferences $(J_{se}^p, e \in E)$ and on moves $(J_{se}^m, e \in E)$, respectively. Subjects with ToM-1 can predict the move of the other agents assuming that they have order 0 ToM; in fact they cannot assume that the other agents have a higher order of ToM or else it would contradict the fact that the subject has ToM-1. The number of steps in the future up to which the subject can predict the moves of the other agents is called the depth of processing and denoted as dp. At step 0 of the prediction, the subject attributes to another agent, a, the preference vector $\tilde{q}_e^0 = p_{sae}$

for entities $e \in E$; and the position of the entities is X^0. At step $k < n$ the predicted position, X^k, expressed emotions e^k and preference vectors $\tilde{q}^k$ are used to predict the displacement, preference update and emotion expression of the others agents, by applying active inference for ToM-0 as described in the previous paragraph. Furthermore, the subject has also an updated version of its preference matrix p^k. The subject then chooses its move at step m^{k+1} by minimizing the following cost function, for $m \in M$,

$$C_1(m, p^k, J, Y^k) = \sum_{\substack{a \in A}} \sum_{\substack{e \in E \\ e \neq b}} \omega_{a,e} \, \mathrm{DKL}(Q(.|\mu_{a, \psi_a(m), p_a^k}(Y_{e,m}^k), \sigma_{a, \psi_a(m)}(Y_{e,m}^k) \| P), \qquad (6)$$

where Y^k is the configuration of the entities that are not the subject at step $k + 1$ and Y_s^k is X_s^k. Y_m^k is a mean to recall that the configuration of the entities depend on the move the subject decides to make, through $Y_{s,m}^k$. Here, for any agent $a \in A$ and entity $e \in E$,

$$\omega_{a,e} = J_a^m \frac{1}{|E| - 1}. \qquad (7)$$

One can remark that C_1 is in fact a weighted mean of several C_0.

From this prediction n steps in the future, the subject chooses the best set of moves that we assimilate to paths, $(m_s^{k*}, k \in [0, n])$, in a set of paths, $\mathcal{P}$, by minimizing

$$\mathrm{FE}(m, p, J) = \sum_{k \in [1,n]} a_k C_1(m^k, p^k, J, X^k), \qquad (\mathrm{FE})$$

where $\sum_{k=1...n} a_k = 1$ and a_k are chosen here to be $a_k = \frac{1}{n}$. The best move to make for the subject is the first move of the best path.

For a subject that has ToM-2, the same procedure as for a subject with ToM-1 holds. The subject can simulate the behaviours of the other agents with respect to the degree of ToM it attributes to them, $(d_a, e \in A)$. From these simulations, it can decide what best sequence of moves to make. To do so, one should consider that the subject has a preference tensor $(h_{sabe}, a \in A, b \in A, e \in E)$ and influence matrices $(I_{sab}^p, a \in A, b \in A)$, $(I_{sab}^m, a \in A, b \in A)$. The case we consider is simpler, as we restrict the preference tensor h to a preference matrix p, such as in ToM-1, by posing that $h_{sabe} = p_{sbe}$. When the subject starts its prediction of the behaviour of the other agents, i.e., at step 0 of the prediction, the influence vectors of an agent a believed to have ToM-1 by the subject are defined as $\tilde{J}_{a.}^0 = I_{sa.}$. The cost function C_2 for the choice of the action of the subject at step k of the prediction is a mean of the cost functions of the other agents depending on the degree of ToM that is attributed to them. We do not enter into more details on how C_2 is computed nor on the cost function for higher dimensions of ToM; they are computed recursively. In the experiment we consider, we assumed that the agents are not influenced in their action by how they believe other agents would feel as a result; what makes the difference between a subject of ToM-1 and ToM-2 is how it predicts the behaviour of the agents, respectively, attributing to them order 0 or 0 to 1 of ToM. In the main experiment of this report, we only focus on inference about ToM order with fixed preferences. In supplementary simulations, we illustrate (see Section 5.6) how our model can tackle situations in which agents simultaneously perform inferences about others' ToM order and preferences.

3.2 Inverse Inference for ToM and Preferences

A subject with ToM-2 can attribute to another agent ToM-0 or 1 and for the subject to truly be able to perform ToM-2, it must be able to attribute correctly to the other agent the order of ToM it truly operates at. To do so, the subject must inverse infer the degree of ToM by analysing the behaviour of the agent.

When the prediction of the subject with respect to the actions of an agent diverges too much from its observed actions, it can start doubting its beliefs on the parameter it previously used to model the other agent's behaviour. It can then find better suited parameters. Here the parameters being considered are the preferences and the order of ToM attributed to the other agents. To do so, the subject uses a measure of divergence from the predictions it made about the action of the other agent, a^p, with respect to the real action it has observed and memorized, a.

Let us consider the following example, at time t, the subject predicts the action $a^p(t)$ with respect to the information it holds about the preferences, $p(t)$, and order of ToM, $d(t)$, of the other agent. If the divergence, $f(a^p(t), a(t))$, is too large, then the subject will update the parameters of its model of the agent, i.e., p and d, to increase predictive power by minimizing a divergence,

$$(p^*, d^*) = \underset{p,d}{\operatorname{argmin}} \, f(a^p(p, d), a(t)). \tag{8}$$

The experiment we use below as a proof-of-concept is a two-choice simulation scenario. In this scenario, both the subject and another agent try to reach a vending machine among two, one being intrinsically more attractive than the other. In the main experiment, both agents try at the same time to avoid running into each other (they have negative preferences toward each other). In the supplementary experiment, the other agent may have negative, neutral or positive preferences toward the subject. The subject cannot see the other agent before the near end of the experimental trial (except at the very beginning). The experiment is divided into two trials. At the first trial, in both experiments, a subject with ToM-2 assumes by default that the other agent is also trying to avoid the subject. It is expected that if the subject encounters the other agent, it will learn from its mistake. In the main experiment, it will revise the order of ToM attributed to the other agent (the only parameter it tries to infer in this experiment). In the supplementary experiment, it will revise both the order of ToM attributed to the other agent, and the preference attributed to the other agent toward the subject. Let us now explain how our agents revise their beliefs when confronted with evidence of misprediction.

A subject s models the order of ToM of an agent by a random variable, that we denote as D, that takes two values 0 and 1. It has as prior law, $(p_1, p_0) \in \mathbb{P}(D \in \{0, 1\})$ for D, that can be parameterized by p_1. The probability distribution plays the role of the belief the subject has on the order of ToM of the other agent. In the simulation scenario, the subject knows where the agent is at time 0, but until the end of the trial, it does not have confirmation of its position. The subject speculates on the position of the agent at each time until it sees it (or not), and gathers new information on its position if it sees it eventually. To do so, it keeps in memory, the predicted position (x_t^0, x_t^1) of the agent at each times t, respectively, assuming that it performs ToM of order 0 or 1. At time $t+1$, if it does not see the agent, then it predicts x_{t+1}^0 using the predicted position of the agent x_t^0 assuming that the agent acts as an agent with ToM of order 0, and it predicts x_{t+1}^1 from x_t^1 assuming the agent acts as an agent with order 1 ToM.

Therefore, at each time t the subject predicts two positions x_t^0 and x_t^1 for the agent. When the subject can attest the real position of the agent, it can confront it to the predicted positions. For example, if this assessment occurs at time t_0, then the subject can confront its predictions with the true position of the agent, by considering $|x_{t_0}^0 - x_{t_0}|$ and $|x_{t_0}^1 - x_{t_0}|$, respectively, the distance between the predicted position and true position of the other agent when the subject assumes that the agent has an order of ToM of 0 versus 1.

The subject computes

$$\mathbb{E}_{p_1}[|X_{t_0} - x_{t_0}|] = (1 - p_1)|x_{t_0}^0 - x_{t_0}| + p_1|x_{t_0}^1 - x_{t_0}| \tag{9}$$

as a metric for consistency of predictions; This metric can naturally be extended to the case where theory of mind and preferences are inferred. If this value is too high, then the subject starts to
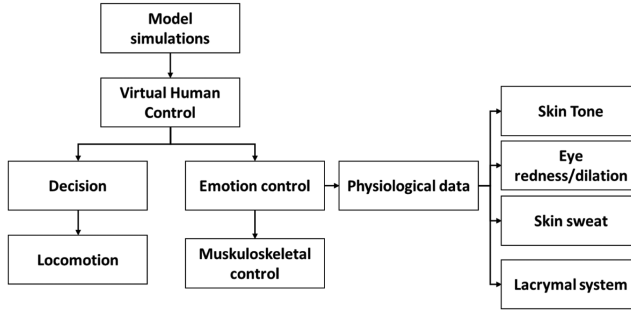
Fig. 2. Virtual Human control components.

doubt its priors and will look for p_1 that minimizes the previous quantity,

$$p_1^*_{t_0} = \operatorname*{argmin}_{p_1} \mathbb{E}_{p_1}[|X_{t_0} - x_{t_0}|]. \tag{10}$$

One shows that this problem is the same as finding the minimum d^*,

$$d^* = \operatorname*{argmin}_{d \in D} \left(|x_{t_0}^d - x_{t_0}| \right) \tag{11}$$

as $p_1^*_{t_0} = \delta(d^*)$, where $\delta(d^*)$ equals 1 on d^* and 0 on the complementary.

In the supplementary experiment, the subject also attempts to infer the preference the other agent may have about it, by comparing predicted and observed outcomes in terms of position, orientation, and emotion expression. For instance, if the subject encounters the other agent and the other agent expresses more positive emotion than expected along with a different pattern of approach versus avoidance, then the subject may infer that the other agent might actually have positive preferences toward it and adjust its strategic behaviour accordingly. Note that in the supplementary experiment, for the interest of the simulation, the subject was capable of up to ToM-3 and the other agent of up to ToM-2.

4 VIRTUAL HUMANS

Recently, we developed real-time simulations of virtual humans with emotional facial expressions, combining physiological and musculoskeletal features [86]. The virtual human control system is processing the data of our model simulations, to manage several facial expressions, including musculoskeletal and physiological state (Figure 2). More generally, the virtual human control system takes as inputs various information from simulations generated by the model, including the 3D world position of the agents, their successive positions in the virtual environment, their emotional states, the expected emotion during the next steps of the simulations, and their beliefs about other agents (positions, perceived values). Based on this information, the virtual human system can procedurally generate a realistic visualization in which agents are moving and acting according to the simulations produced by the model. A pathfinding system is used to apply smooth displacement to the virtual human during its navigation in the virtual environment, based on the sequences of positions and orientations outputted by the model. This system is based on a state-of-the-art pathfinding method using Astar algorithm [87]. The body movements are animated accordingly to be consistent with the overall movement of the virtual human. More precisely, the locomotion animations are blended based on the given direction and speed, which are passed as an input into a 2D Cartesian mapping that controls animations' transition. Regarding facial expressions, the virtual human control system uses the emotion state of the agent, based on three main components: the agent's positive emotion intensity and negative emotion intensity (which when subtracted yield
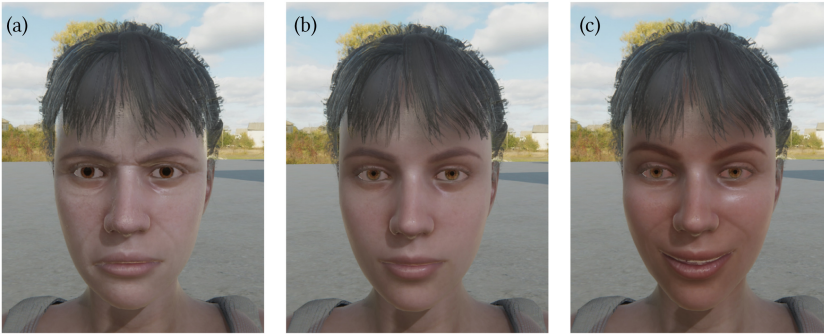
Fig. 3. Example of VH facial expressions including muskuloskeletal and physiological control. (a) "High sympathetic" parasympathetic tone = 0, sympathetic tone = 100 (max), negative emotion expression; (b) "Neutral": parasympathetic tone = 0, sympathetic tone = 0, neutral facial expression; (c) "High parasympathetic," parasympathetic tone = 100 (max), sympathetic tone = 0, positive facial expression.

a parameter of valence), and a surprise coefficient. Using these parameters, realistic facial expressions are generated and applied on the virtual human, including musculoskeletal control based on the **Facial Action Coding System (FACS)** [88], and physiological control (skin tone, sweat, pupil dilation, eye redness) (Figure 3) (see Reference [86] for theoretical and technical details). The musculoskeletal system is based on vertex displacement (blendshapes) and joint animation, while the physiological system is based on dedicated shaders developed for our virtual human system. The mapping between emotional states and virtual human parameters are summarized in Table 1. Of note, the virtual humans' motion along trajectories outputted by the model simulations also provided clues about the emotional state of the agents for inference, as they could be decoded by the subject as behaviours of approaches or avoidance.

Figure 3 shows examples of facial expressions, generated using our virtual human control system, including muskuloskeletal and physiological features.

5   SIMULATION-BASED EXPERIMENTS

We designed a social experiment using simulations to prove our concept. The experimental design was chosen so that behavioural outcomes would imply specific orders of ToM in participants performing the task. We implemented a main experiment focusing on ToM order estimation as a test bed. We also implemented a supplementary experiment based on the first one, in which preferences of another agent toward self had to be inferred. Additionally, we assessed the robustness of generated behaviours as a function of the preference values assumed by the agents.

5.1   Requirements for the Main Experiment

We considered the following design and simulation requirements for a preliminary validation. First, the task's behavioural outcomes should be unambiguous, and allow an experimenter to determine the order of ToM used by a participant who would follow the instructions of the task. Second, the agents simulating participants in the experiment should demonstrate clear adaptive behaviours of navigation and emotion expression in a manner that is consistent with the task expectations, reflecting their attempts to optimise task outcome as a function of the order of ToM used by the agents to predict each other's actions and act accordingly. Here, we considered agents that could perform ToM of order 0 (no ToM), order 1 (ToM at the 1st order), and order 2 (ToM at 2nd order) at most. Third, one agent in the experiment, the subject S, could be used as a virtual "psychologist"

Table 1. Virtual Human Control Mapping

	Physiological	Musculoskeletal	Value
Neutral			
Skin tone	✓		0.5
Pupil dilation	✓		0.5
Eye sclera and cornea redness	✓		0.5
Skin sweat	✓		0
Facial expression		✓	0
Positive emotion with parasympathetic tone			
Skin tone	✓		1
Pupil dilation	✓		0
Eye sclera and cornea redness	✓		1
Skin sweat	✓		0
Facial expression (Action Units)		✓	6,12
Negative emotion with sympathetic tone			
Skin tone	✓		0
Pupil dilation	✓		1
Eye sclera and cornea redness	✓		0
Skin sweat	✓		1
Facial expression (Action Units)		✓	1,4,15
Surprise			
Facial expression (Action Units)		✓	1,2,5,26

agent, and should demonstrate its capability to estimate the hidden ToM parameters of another agent O, based on its predictions and task outcomes. O could then be conceived of, in this proof-of-concept, as a potential participant in an interactive experiment, designed to assess ToM capacities in the participant, and which could be implemented in VR.

We reasoned that if PCM-driven virtual human S could assess another PCM-driven virtual human O correctly according to the experiment expectations, then they would be able to assess real humans in an experiment with real human participants, either by interacting with them or by observing them. Our motivation was to render the simulation sufficient to prove the concept without actually running the experiment in real human participants, which was beyond the scope of this report.

5.2 Experiments: Rationale and Design

We chose to design a rather simple and well-controlled entry-game, inspired by the common situations of having to choose between two stores or, more generally, destinations with different cost-benefits, in which one destination is more attractive than the other but also more likely to be crowded with other people, which can hinder the free exploitation of the destination. The experiment was aimed to create a conflict of approach and avoidance based on non-social (intrinsic attractiveness of a destination) and social (social distancing and competition) preferences. Two agents, the subject S and another agent O, had to compete for reaching one of two **vending machines (VM)** selling coffee, on opposite sides of a building on a parking lot in a gas station. One of them, VM1 was better with higher quality products, and thus was the most attractive one for both agents. The other one, VM2, more mainstream, was not stocked with products of high quality, and thus was less attractive for both agents.
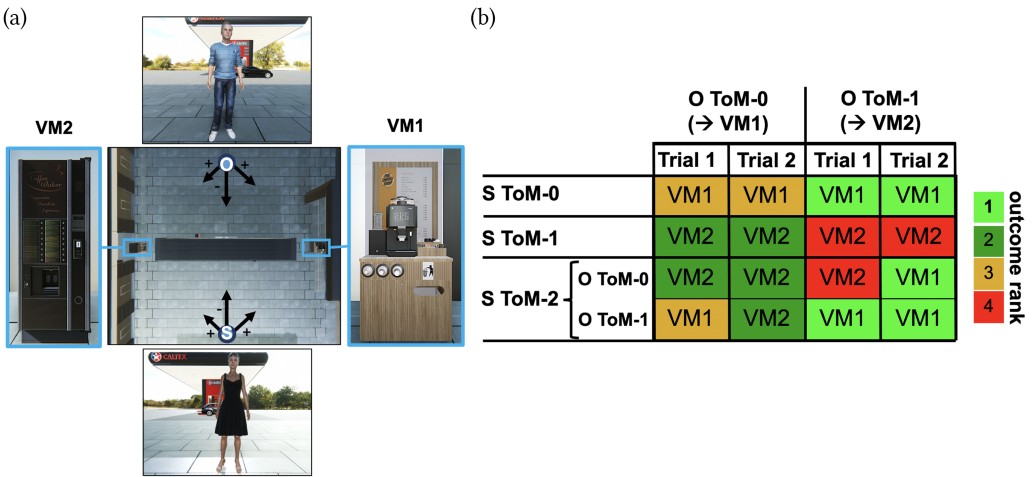
Fig. 4. Main experiment: setup and design. (a) Illustration of experimental setup. Two agents, a subject S and another agent O are on opposite sides of a small building (middle rectangle). On each side of the building, there is a vending machine, VM1 (right) and VM2 (left). Arrows arising from circles marking the initial position of the agents O and S indicate fixed prior preferences towards entities: both S and O like (positive signs) VM1 and VM2 but prefer VM1 (longer arrow), and dislike each other (negative signs). (b) The table shows the expected behavioural outcomes for the different combinations of orders of ToM at $Trial$: 1 and $Trial$: 2 for the subject S. O can either perform ToM-0 (no ToM) or ToM-1 (first order ToM). The table is color coded to indicate the rank of the outcomes in terms of optimisation of perceived value as a function of preferences (scale on the right), from best (rank 1) to worst (rank 4)(see text).

In the main experiment, another assumption was that agents would both prefer avoiding running into each other, which was operationalized as agents having negative preferences about each other, and assuming such negative prior in each other. In the supplementary experiment, the subject had negative preferences toward the other agent, but the other agent could have negative, neutral or positive preferences toward the subject, which always initially assumed that the other agent had negative preferences against it. In all cases, both agents assumed that VM1 was most attractive to the other agent and VM2 a secondary choice (see Figure 4(a)). The situation was designed so that agents would have limited access to sensory evidence about the actual behaviour of the others. Each agent could tell that the other was on the other side of the building at the beginning through sliding doors, first opened and then closed, and also assumed that the other agent was looking for a vending machine. However, a given agent would not be able to observe the behaviour of the other after the doors closed, except when reaching the areas of the vending machines on the side of the building. If both agents chose to go to the same vending machine, then they could observe the other directly; if they chose to go to different vending machines, then they could infer that the other was on the other side.

The task for the agents was to choose which vending machine to go to (approach versus avoid), with the aim of maximising the satisfaction of their preferences. They had to use ToM to plan their actions. Only when reaching their destination could they use sensory evidence to attempt to infer the actual order of ToM used by the other agent. In the supplementary experiment, they also had to infer preferences of the other agent toward self, depending on the outcome and the agents' capacity for ToM. The distance of the two vending machines from the agents at initial condition was equivalent to avoid a bias of distance on appraisal and the motivation of action, which would be entailed by the model, as in real situations.

The main experiment always involved the same initial setup with fixed preferences. Section 5.4 summarizes the fixed parameters of the simulation. The possible outcomes of the experiment were designed to correspond to decreasing levels of optimality in terms of satisfaction of preferences, e.g., level of free energy or emotion expressed, for the agent under consideration, from best to worst optimality rank. Thus, for $rank1$, S finds itself alone at VM1 (thus O went to VM2); for $rank2$, S finds itself alone at VM2 (thus O went to VM1); for $rank3$, S finds itself in the presence of O at VM1; and for $rank4$, S finds itself in the presence of O at VM2.

Figure 4(b) presents the combination of experimental conditions and outcomes for the main experiment. The experiment combined three factors. The first factor, the Subject S ToM capacity, corresponded to the maximum order of ToM for S, with three levels: ToM-0, ToM-1, ToM-2. When an agent was capable of performing ToM-2, it was also capable of performing ToM-1. The second factor, the other O ToM capacity, corresponded to the actual order of ToM used by O, with two levels: ToM-0, ToM-1. This number of possible levels for factor two was prescribed by the fact that the experiment assumed the subjects' maximum order of ToM to be ToM-2. This is a logical consequence of the very definition of ToM: an agent with a maximum order of ToM n (ToM-n) can at most make inferences about another agent with maximum order of ToM of $(n-1)$ (ToM-$(n-1)$) (see Reference [79]). Since the experiment was designed to assess a subject's maximum order of ToM up to ToM-2, the experiment had two possible actual orders of ToM for agent O: ToM-0 and ToM-1. This could be generalized to higher orders of ToM, which our model is capable of performing, but we wanted to limit the number of conditions of the experiment for the sake of the clarity of presentation. The third factor was the trial number, as the task was repeated twice for each condition, with two levels: $Trial:1$, the initial action, and $Trial:2$, the second attempt, so that subjects with sufficiently high capacity for ToM (ToM-2) could infer the ToM order of the other agent O, and adapt their behaviour for the second trial, to reach a more optimal outcome.

This design guaranteed that we could estimate the order of ToM of an agent following the task instructions by comparing outcomes as a function of condition. It is important to note that during both trials, the order of ToM of O is assumed to be fixed, in both the main and supplementary experiment.

The expected outcomes of the main experiment are derived from the behaviour expected from a real human conforming to the task and are presented in Figure 4(b).

If O performs ToM-0, then it should necessarily go to the preferred VW1, and if it performs ToM-1, then it should necessarily go to the less preferred VM2 (to avoid S that it should expect at VM1). There are three possibilities of capacity for ToM for agent S: ToM-0, ToM-1, or ToM-2 (second order ToM). If S is able to perform ToM-2, then it can also perform ToM-1, and adjust the order of ToM it uses to optimize outcomes. Outcomes at $Trial:1$ will depend on whether S assumes initially that O performs ToM-0 or ToM-1. At $Trial:2$ agent S should be able to revise its priors about the order of ToM of O, and optimise outcome on this basis. The experimental design is such that each row in the table is different from the other so that, when considering the different conditions of actual ToM order for O (ToM-0 and ToM-1), and the two trials $Trial:1$ and $Trial:2$, the order of ToM which S is capable of, as well as the order of ToM that S attributes to O, can be determined.

Let us explain how these outcomes are derived when S can have a ToM of order 0, 1, 2, and O of order 0, 1. If an agent performs ToM-0, then it goes toward its preferred machine $VM1$. If an agent performs ToM-1, then it is expected to go to the VM opposite to the one the predicted agent with ToM-0 would go to. The left-tier of the table corresponds to the case in which O performs ToM-0 and therefore always goes to VM1. The right-tier corresponds to the case in which it performs ToM-1 and goes to VM2. The experiment is divided into two trials $Trial:1$ and $Trial:2$. If S does not perform ToM-2, then it cannot learn from its mistakes and the outcome of both trials is the same;

it is reported in the first two lines of the table. If S performs ToM-2, then it may initially assume that O has ToM-0 (the less greedy hypothesis on O). Then, we expect that after one trial of the experiment, S would confront its expectations to the outcome of the trial. If it predicted correctly the outcome and does not meet O, then it does not have to change its beliefs and therefore may assume that O had ToM-0 and is located around $VM1$, for both $Trial : 1$ and $Trial : 2$. Thus, S should go to VM2 in both trials. However, if the beliefs of S were false during the first trial, then both agents should arrive at $VM2$, and S is expected to change its belief about the ToM order of O, from ToM-0 to ToM-1. In the next trial, S should go to $VM1$ and O to $VM2$. A second possible case is if S initially assumes that O has ToM-1, in which case, similarly, S would have to update its beliefs to account for the false prediction. The expected results of the experiment for a subject with ToM-2 are reported in the last two lines of the table, the first one corresponding to the situation in which the subject believes at first that O has ToM-0 and the second one ToM-1.

The supplementary experiment derives from the main one. The general setup is the same, but only a subset of relevant conditions is explored to demonstrate how the subject can simultaneously make inferences about both ToM order and preferences about itself in the other agent, to optimize outcomes. The specific details of this experiment are presented in the corresponding result section.

5.3 Virtual Humans Simulation Procedure

At the beginning of each simulated case, the positions of the agents, the divider between the two virtual humans and the coffee machine were procedurally set based on the given simulation. The visualization was created using Unity (v2021.1), with the **High-definition Rendering Pipeline (HDRP)**. The virtual humans' model was designed using Character Creator. The virtual human control was handled by the Geneva Virtual Humans toolkit [86], managing the locomotion, facial musculoskeletal and physiological parameters, as described in Section 4. A preview window allows the real-time visualization of simulations, and the offline rendering of the simulation with a multi-view system. The system allows the creation of 4K videos from various point of views, a radar view that shows the path of the agents, a general viewpoint of the simulation, a camera focusing on the facial expressions of each virtual human, and the first-person views for each virtual human. The scene is designed to run both in real-time desktop and VR applications and offline with up to 4K rendering. The scene is VR ready, for future real-time experiments in which a real human participant could be part of the scenario. The scene can already be observed by an experimenter within VR (see Figure 5).

5.4 Main Experiment Simulation Parameters

The subject is labelled 1, the agent 2, VM1 is 3, and VM2 is 4. In the experiment, the preference matrix for the subject, p_1, is the following:

$$\begin{pmatrix} 0.5 & 0.01 & 0.6 & 0.55 \\ 0.01 & 0.5 & 0.6 & 0.55 \end{pmatrix}. \tag{12}$$

The preference matrix of the other agent is

$$\begin{pmatrix} 0.5 & 0.01 & 0.6 & 0.55 \\ 0.01 & 0.5 & 0.6 & 0.55 \end{pmatrix}. \tag{13}$$

These matrices contain more information than necessary when the subject or agent performs ToM-0 and the preference vectors in these cases are the true preferences extracted from the previous matrices.

The influence matrix of the subject for preferences and action (see above and Reference [2]) are reduced to self influence, only $I_{saa} \neq 0$ and similarly for the influence vector of the other agent.

Fig. 5. Overview of the scene. Screenshot of the simulation viewer.

5.5 Results of Main Experiment

We focus on the results regarding S. The results are consistent with the design. Agents behave as expected as a function of their relative ToM order. Their end state FE and expressed valence are consistent with the corresponding outcome ranks. Figure 6 shows results of basic behaviours of the agents as a function of ToM contingencies, without S performing inverse inference to learn the order of ToM of O.

S succeeds in inferring O ToM order after the first trial ($Trial$: 1), and optimises outcome accordingly. Figure 7 shows results of the behaviour of the agents as a function of ToM contingencies, when S performs inverse inference to learn the order of ToM of O at the second trial ($Trial$: 2).

5.6 Results of Supplementary Experiment

In this supplementary experiment, our aim was to illustrate complex interactions between parameters, and demonstrate how the subject could correctly or wrongly infer both the other agent's ToM order and its preferences about the subject, by leveraging both behavioural outcomes and emotion expressions to perform ToM; the correctness of the perception of the subject is defined by the compatibility of its inference with the setting of the simulation. We considered five illustrative scenarios, which would illustrate, on the one hand, cases in which the situation was intrinsically ambiguous and thus the problem undecidable, but yielded interesting results, and, on the other hand, cases in which the conditions were compatible with partial or complete successful inference.

In all these scenarios, the subject S had negative preferences toward the other agent O, and thus would try to avoid it. It also initially assumed that the other agent had negative preferences about the subject, and thus that the other agent would also tend to avoid it. In fact, the other agent always had positive preferences toward the subject, and thus would tend to go to places that would offer the best trade-off between reward at the level of the VM, and social reward through
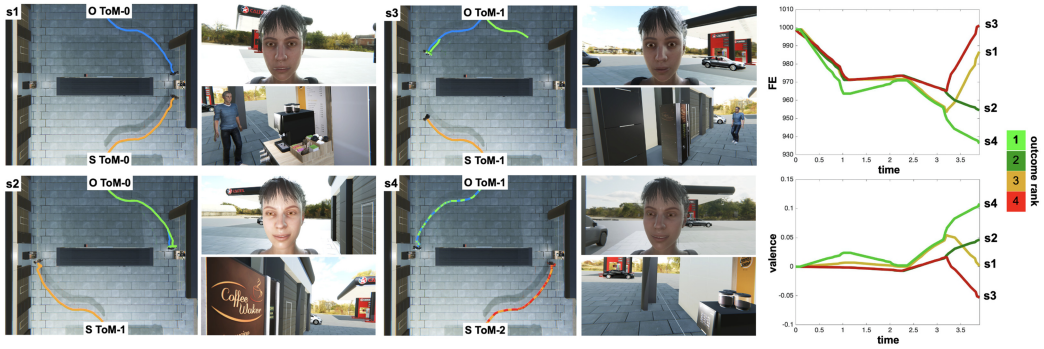
Fig. 6. Basic behaviours of the agents as a function of ToM. For each situation, s1, s2, s3, s4, images of virtual environment. *Left*: view from above. Orange traces are *S*, blue traces are *O*, green traces are predictions about *O* according to *S*, red traces are predictions about *S* attributed to *O* by *S* (if *S* uses ToM-2). *Upper-right*: female virtual human face close-up. *Lower-right*: first person perspective of *S* on *O* (male virtual human). (s1) *S* and *O* perform ToM-0. Agents run into each other at VM1. *S* emotion expression is rather unhappy. (s2) *S* performs ToM-1 and *O* ToM-0. Agents do not run into each other. *S* goes to VM2. *S* emotion expression is rather happy. (s3) Both *S* and *O* perform ToM-1. Agents run into each other at VM2. *S* emotion expression is unhappy. (s4) *S* performs ToM-2 and *O* ToM-1. Agents do not run into each other. *S* goes to VM1. *S* emotion expression is happy. See free energy (FE), valence, and outcome rank in the charts on the right, corresponding to each situation as labeled.

the proximity of the subject. In situations in which *S* would run into *O*, *S* could try to infer whether *O* had negative ($p = 0.1$), neutral ($p = 0.5$) or positive ($p = 0.8$) preferences toward it. For these simulations, *S* was able to perform up to ToM-3, and *O* up to ToM-2.

In scenario 1, the subject *S* initially assumed correctly that the other agent *O* was performing ToM-0. In this case, even though the negative preference attributed by *S* to *O* was incorrect, the two agents ended up at opposite VM. Thus, *S*, lacking sensory evidence, had no reason to revise its beliefs and continued wrongly to attribute negative preferences to *O*, with little impact on outcome.

In scenario 2, *S* initially assumed wrongly (with respect to the task) that *O* was performing ToM-0, whereas it was actually performing ToM-1. In this case, even though the negative preferences and ToM order attributed by *S* to *O* were incorrect, the two agents also ended up at opposite VM. Indeed, since *S* was performing ToM-1, and thus assumed *O* was performing ToM-0, *S* went to VM2 to avoid *O*, which it predicted would go to VM1. But in fact, since *O* was performing ToM-1, and thus assumed *S* was performing ToM-0, *O* predicted that *S* would go to VM1. Moreover, since *O* actually had positive preferences toward *S*, it was all the more driven to go to VM1. Thus, here again, *S*, lacking sensory evidence, had no reason to revise its beliefs and continued wrongly to attribute negative preferences to *O*, with little impact on outcome.

In scenario 3 (see Figure 8(S3)), *S* initially assumed wrongly that *O* was performing ToM-0, whereas it was actually performing ToM-2. In this case, since *O* correctly predicted that *S* would go to VM2 to avoid it, and since *O* had positive preferences toward *S*, *O* went to VM2 as the trade-off ended up being better in terms of free-energy minimization. As a result, both *S* and *O* found themselves at VM2, *S* could use sensory evidence (both in terms of overt approach-avoidance behaviours and emotion expressions) to revise its priors, and was motivated to do so because of its level of surprise. On this basis, *S* used ToM-3 to revise its priors, and correctly attributed ToM-2 and positive preferences of $p = 0.8$ toward it to *O*. With these updated parameters, in a second trial, *S* then chose to go to VM1, both maximizing reward in terms of VM and avoiding *O*, which resulted in minimal free energy.
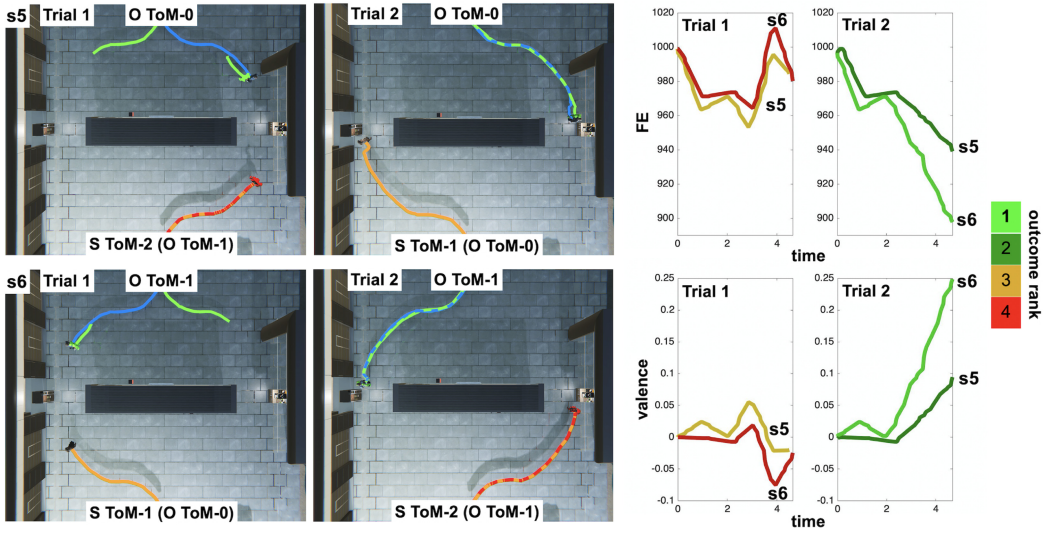
Fig. 7. *S* performs inverse inference about *O* ToM order. For each situation, s5 and s6, images of virtual environment at *Trial* : 1 and *Trial* : 2. Orange traces are *S*, blue traces are *O*, green traces are predictions about *O* according to *S*, red traces are predictions about *S* attributed to *O* by *S* (if *S* uses ToM-2). *Left*: view from above at *Trial* : 1. *right*: view from above at *Trial* : 2. (s5) *Trial* : 1: *S* performs ToM-2 (and attributes ToM-1 to *O*) while *O* actually performs ToM-0. Agents run into each other at VM1. *Trial* : 2: *S* has inferred *O* ToM order, and obtains a better outcome, finding itself alone at VM2. (s6) *Trial* : 1: *S* performs ToM-1 (and attributes ToM-0 to *O*) while *O* actually performs ToM-1. Agents run into each other at VM2. *Trial* : 2: *S* has inferred *O* ToM order, and obtains the best outcome, finding itself alone at VM1. See free energy (FE), valence, and outcome rank in the charts on the right, corresponding to each situation as labeled, for *Trial* : 1 and *Trial* : 2.

In scenario 4 (see Figure 8(S4)), *S* initially assumed correctly that *O* was performing ToM-1. In this case, *S* predicted that *O* would go to VM2 to avoid running into it, and *S* thus chose to go to VM1. However, *S* was still wrong about the negative preference of *O* toward it. In fact, *O* being attracted to *S*, *O* also chose to go to VM1. As a result, both *S* and *O* found themselves at VM1, *S* could use sensory evidence to revise its priors, and was motivated to do so because of its level of surprise. *S* ended up correctly inferring that *O* had in fact positive preferences of $p = 0.8$ toward it. However, the choice between attributing ToM-1 and ToM-0 to *O* was intrinsically ambiguous, as both attributions lead to the same behavioural outcome. In practice, because of the ambiguity of sensory evidence, *S* chose to attribute ToM-1 to *O* because of its initial prior that *O* was performing ToM-1. However, if *S* had been wrong regarding the prior, it could have wrongly inferred that *O* performed ToM-0, but the predicted outcome would have been the same. The problem is undecidable in this condition. In a second trial, *S* then chose to go to VM2, to avoid *O*.

In scenario 5, *S* initially assumed wrongly that *O* was performing ToM-1. In fact, as *S*, *O* was performing ToM-2. *S* expected that *O* would try to avoid it by going to VM2. Thus, *S* went to VM1. The outcome was optimal for *S* but based on the wrong rationale. In fact, *O*, which actually performed ToM-2, also expected that *S* would go to VM2 to avoid running into *O*. Since *O*, contrary to what *S* believed, was attracted to *S*, it chose to go to VM2 to be near *S*. In this case, the situation was intrinsically ambiguous, and the wrong parameters of ToM order and preference attributed to *O* by *S* compensated each other in generating an outcome for *S* that was optimal. Since the two
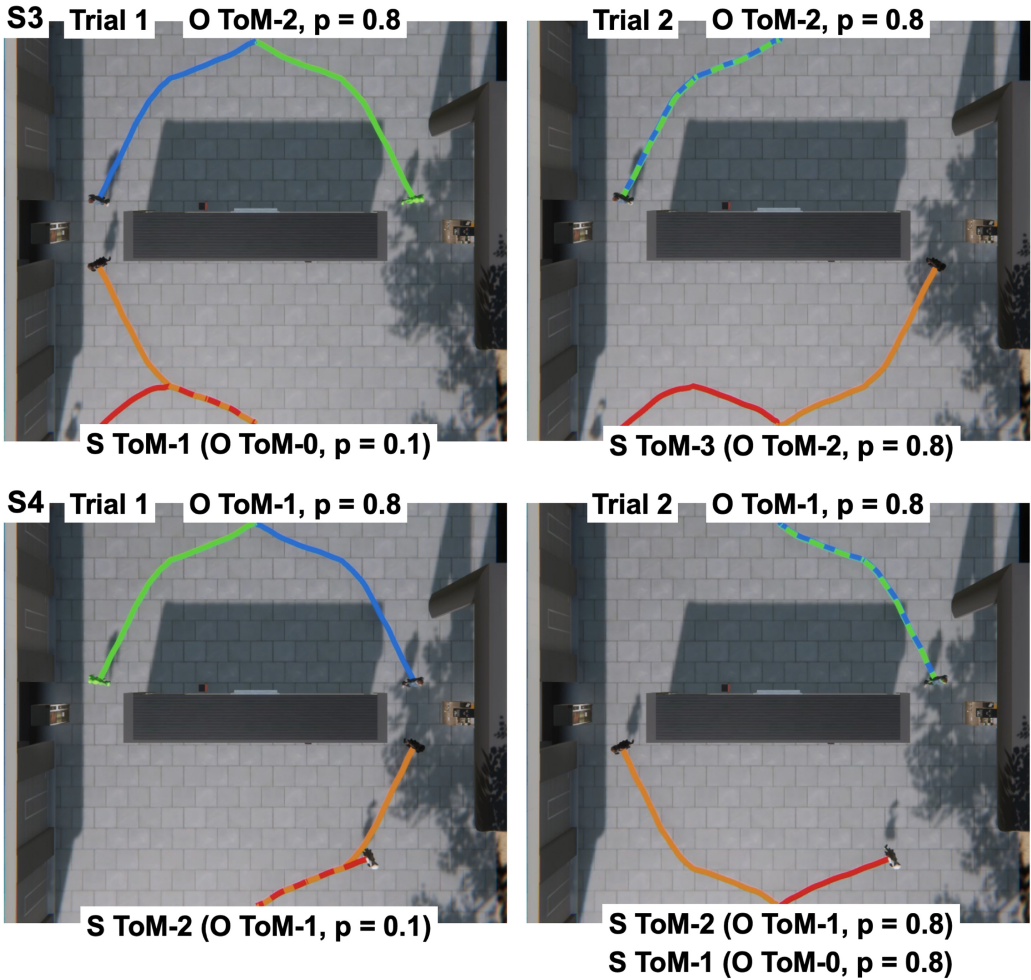
Fig. 8. S performs inverse inference about both O's ToM order and preference about S. Results for scenarios 3 (S3) and For scenarios 4 (S4). Views from above of virtual environment for $Trial : 1$ (*left*) and $Trial : 2$ (*right*). Orange traces are S, blue traces are O, green traces are predictions about O according to S, red traces are predictions about S according to O. See text.

agents did not run into each other, S was not in a position to make further inference to revise its beliefs.

5.7 Results on Robustness of Outcome Behaviours

We investigated, in an exploratory manner, how behavioural outcome between going to VM1 and VM2 depended on values of preference to assess the robustness of the behavioural outcome, and the existence of sharp bifurcations between outcomes. We used the following setup. S performed ToM-1 and O ToM-0. Thus, S correctly predicted that O would go to VM1. We varied:

(1) the preference of S toward VM1 (red trace on Figure 9), from 0.5 (neutral) to 0.8 (quite positive), fixing its preference toward VM2 at 0.58 (slightly positive), and toward O at 0.1 (quite negative);
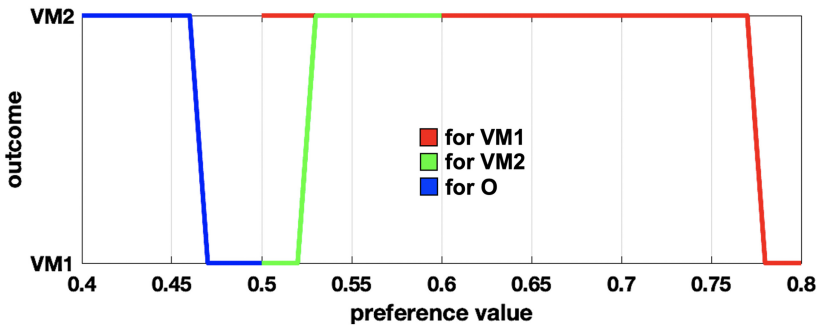
Fig. 9. Outcome robustness to preference values. Outcomes of the experiments with respect to the variation of the preference of S toward VM1 (red curve), VM2 (the green curve), and O (the blue curve); for more details see the first paragraph of Section 5.7.

(2) the preference of S toward VM2 (green trace on Figure 9), from 0.5 (neutral) to 0.6 (moderately positive), fixing its preference toward VM1 at 0.6 (moderately positive), and toward O at 0.1 (quite negative);

(3) the preference of S toward O (blue trace on Figure 9), from 0.4 (moderately negative) to 0.5 (neutral), fixing its preference toward VM1 at 0.6 (moderately positive), toward VM2 at 0.58 (slightly positive).

Figure 9 shows that it took a strong increase in preference toward VM1 from neutral for S to overcome its negative preference toward O and approach VM1 instead of VM2 in spite of that negative preference (red trace). Conversely, it did not take a strong increase in preference toward VM2 from neutral for S to switch from VM1 to VM2, given its negative preference toward O (green trace). Finally, as the preference of S toward O moved from negative toward neutral, S switched from VM2 to VM1 (blue trace).

6 DISCUSSION

Our main objectives in this report were:

(1) to demonstrate that we could integrate an explicit model of the subjective perspective of consciousness in a recursive, multi-agent model (PCM-driven agents) that could be combined with virtual humans;

(2) to simulate behaviours of approach-avoidance, which could be expected to result from configurations of preferences and orders of ToM in real humans;

(3) in a main experiment that could classify these behaviours in terms of underlying generative models as a function of orders of ToM, based on clear behavioural outcomes;

(4) that could be implemented in VR in the future for actual experiments with real humans;

(5) and to demonstrate that our PCM-driven agents could correctly infer parameters such as ToM order used by other agents, as a basis for the future use of such approach to assess real humans in immersive experimental contexts such as VR.

In a supplementary experiment, we showed how a PCM-driven agent could infer simultaneously order of ToM and preference parameters, based on behavioural outcomes and emotion expression, when the problem was decidable. Furthermore, we showed that the overall behaviour of the agents was robust to changes in preferences up to a certain threshold beyond which agents would switch their choices of approach-avoidance.

We emphasize that the aim of this contribution was to offer a proof-of-concept, which still needs developments and empirical validations. One originality of the approach is to leverage active

inference and projective mechanisms to enable artificial agents to interact with each other in a three-dimensional virtual environment.

6.1 Killing three Birds with One Stone

There are three complementary points of view that can be taken on the main experiment we presented as a proof-of-concept of an approach to the problem of integrating computational models, artificial agents, and virtual humans, in a unified framework for psychological science.

First, the main experiment was a way of assessing whether PCM-driven agents would be capable of demonstrating behaviours that would be expected in the task according to their order of ToM (as could be expected from a real human performing the same task and following its instructions). Indeed, we have built an experiment that was adapted *a priori* for distinguishing real human behaviours. We have shown that our virtual humans show rational behavior with respect to conditions on preferences and ToM considered in this experiment. Even though we did not explicitly compare our results to empirical data in this contribution, generally speaking, our results are consistent with results from comparable approaches [75, 76, 78]. However, these approaches did not aim at modeling consciousness explicitly. They were not particularly constrained by the aim of capturing internal processes as they operate in human psychology [8].

Second, we have also shown that the experiment is a good classifier for the behaviours of the virtual humans. Different parameters of the model generating the behaviours of the virtual humans could be distinguished by the experiment.

Third, the approach used agents that could make inferences about psychological parameters driving others, as a human experimenter or psychologist might do to assess the determinants of observed behaviours in humans. The results of the main experiment demonstrated that the target parameters of the model, the order of ToM, could be well discriminated by the virtual human that we called subject S in the experiment. Indeed, S was able to discriminate the order of ToM of the other agent (ToM-0 or ToM-1) consistently, based on its own internal mechanics of inverse inference, as embedded within its recursive multi-agent architecture. Importantly, S would be able to do so, even if the agent was a real human in a VR experiment leveraging motion tracking information from the human participant.

In other words, the same model that could be used to control artificial agents embodied as virtual humans in a VR task, could be used in the future to make inferences about the behaviours of real humans that would participate in the same task. The rationale is that such agents could then assist investigators with assessing and reporting hidden psychological parameters and profiles of different groups and individuals based on the PCM. Although very preliminary, this is in line with a research program with the overarching goal of building synthetic psychology, i.e., the rebuilding of human psychology into artificial systems. Indeed, if successful, such a program should yield artificial agents that can interact with, and interpret each other as well as real humans, as one would expect from real humans toward each other, because artificial and human agents would share similar internal representations and processing mechanisms.

For simplifying the presentation and discussion, we limited the simulations in the main experiment to agents with at most ToM of order 2. However, the framework can be used to simulate agents with higher-order ToM capacities. In the supplementary experiment, we illustrated how agents could use ToM of order 3. We could envision in the future the incorporation of agents capable of ToM of order n. Results in this direction are promising but beyond the scope of this report.

We used simulations of a simple entry-game, i.e., a very controlled and somewhat narrow task. In the main experiment, it was used to demonstrate our proof-of-concept, i.e., by showing that the model could infer the order of ToM in others through interactions in a three-dimensional virtual

environment. In the supplementary experiment, we extended the demonstration by having the model infer both order of ToM and preferences in others. We chose to focus on this particular illustrative example using simulations only, as a more systematic application to a variety tasks, as well as systematic comparison with empirical data, were beyond the scope of this theoretical and technical, preliminary contribution.

But the same model could be used to perform a variety of tasks in which agents would have to infer preferences in other agents based on their behaviours, which currently include their orientation or move towards or away from objects and others, and corresponding emotion expressions. Likewise, assuming adequate experimental and task design, the model could be used to infer a combination of parameters (preferences and order of ToM in others, or any other parameters that the model include). We showed preliminary results in that direction with the supplementary experiment. More generally, PCM-driven agents have the potential to be used to simulate behaviours in a variety of contexts (see, for instance, our recent work on simulating adaptive and maladaptive behaviours relevant to developmental and clinical psychology [2]).

6.2 Comparison of the Model with Other Models Leveraging Behavioral Game Approaches

In a multi-agent setting, agents do not have direct access to the way other agents act (policies) [89]. To predict other's action they must estimate their policies based on parameters that characterize the behavioural profile of the agent. One of these parameters is the order of ToM of the agent. Multi-agent reinforcement learning [90, 91] and in particular recursive reasoning for multi-agent reinforcement learning [92] offer a framework to model how ToM influences strategy elaboration through beliefs about others' beliefs (and entails Interactive Partially Observable Markov Decision Process). In particular, such approach yields good results when fitting human behaviour in low complexity controlled games [1], and generate strategies under fixed order of ToM that are compatible with behavioral patterns related to that order. We used a simple game to show that our model also generates strategies compatible with given orders of ToM, following the same core ideas present in (inverse) multi-agent reinforcement learning, which are recursive reasoning and inference of behaviourally relevant parameters.

The inverse inference of preferences and ToM with our model is an improvement with respect to previous work [2], but we acknowledge that the types of agent personalities modeled are limited, and that we lack empirical data from experiments for further fitting our model to human behaviours. However, we believe that the strength and novelty of our model is that strategic planning is made directly on a space of desires (what we call preferences) that inherits geometrical structure from the space of representation the agent has on its environment (a projective space). When changing from one context to another, one game to another, the underlying model of motivation and decision making does not have to change. What changes are the initial conditions (initial preferences) and the representation of the environment in the inner world of the agent. We hope that doing so will allow for a robust generation of behaviours irrespective of the context. The actions being driven by the (intrinsic) dynamics of the desires (motivations), one does not need to define an explicit reward function. Furthermore, the geometric structure of the representation space captures perspective taking, which allows an agent to make the points of view of other agents comparable to its own. These constructions are psychologically motivated, and aim to model key aspects of consciousness, which are absent from standard approach to multi-agent dynamics.

6.3 Limitations and Perspectives

As this point, without tailored restrictions limiting computational load, the PCM algorithm does not run in real-time. It cannot be used in practice for many real-time interactive VR experiments.

Optimisation approaches both at the software and hardware levels are currently being pursued to mitigate the problem. To close the loop of interactions between real humans and virtual humans in VR, parameters need to be extracted from VR participants to serve as inputs to the model, so that PCM-driven agents will interpret VR participants as they interpret other PCM-driven agents. This can be achieved by leveraging VR technology, including motion capture, eye-tracking and smart interfaces for emotion expression capture, but this important issue is beyond the scope of this contribution.

Another interest of the approach we develop is that it could be used to assess and optimize the capacity of different immersive experimental designs to discriminate psychological parameters and mechanisms based on behaviours (see also Reference [78]). Different designs could be explored through mass simulations of PCM-driven agents along a variety of parameters, independently assessed or in combinations, e.g., preferences and orders of ToM. If target parameters can be well distinguished based on behavioural outcomes, then the design would appear sufficient as a classifier. On the contrary, if different sets of parameters would predict the same outcome behaviours in an ambiguous manner, that would mean that the design of the experiment is not fully discriminant. Designs could then be revised and retested by adding conditions. For instance, if subsets of parameters could not be reliably inferred simultaneously by agents due to the ambiguity of behaviours, as illustrated in the supplementary experiment, then the experiment could introduce different phases of estimations in which the parameters could be assessed independently, e.g., by observing first the behaviours of agents without interacting with them to infer their mutual preferences, and then, based on this prior, attempting to infer their order of ToM through interactions with them.

In this proof-of-concept, we used fixed preferences that were identical between the two agents. This lead to perfect predictions of the trajectories of the other agent by a given agent, when the order of ToM was adequate. Using different values of preferences between those attributed to an agent and those actually used by the agent may induce some error of prediction. Manipulating such differences and studying their impact on fine grained behaviours is interesting but beyond the scope of this contribution. However, we showed preliminary robustness results in Section 5.7 that suggest that the algorithm is robust. Slight randomizations of the preferences did not lead to qualitatively different behaviours that would have impacted the results of the experiments.

Let us note that the way a subject predicts the action of the other agent several steps in the future, and uses this prediction to choose its best move, can be seen as an optimal control problem with a cost function depending on time. It might be possible to consider that the whole simulated situation is itself a control problem. Optimal policies might give a more optimal update rule for simulating the action of the other agent by the subject. However, the approach we chose focused on modeling agents as subjective systems, who model and predict others' actions based on their own internal mechanisms of action selection, as prescribed by simulation theory [54, 55]. The approach assumes that agents may not always have the possibility of developing a full representation of the problem for more globally optimal solutions.

We wish to comment on how non-verbal behaviours express affective states in our current implementation of the agents and virtual humans. Here our virtual humans were expressing affective cues through: facial expressions (across both musculoskeletal and physiological channels; see Reference [86] for motivation and details), gaze/body direction, and approach-avoidance behaviours. In the future, other non-verbal behaviours such as gestures and posture could be integrated and recognized by the model just like facial expressions. They could be expressed by the virtual humans, for instance, by using animations corresponding to typical affective gestures and postures. Likewise, motion speed could be modulated by affective parameters, for instance, through the expected increase or decrease of free energy as a function of time. Such modulation could then be

used by the model for inference as further indication of affective states, e.g., arousal. Furthermore, affective vocalizations could also be integrated in the emotion expression repertory of the model. These developments are in the pipeline but beyond the scope of this report.

Finally, considering longer term perspectives, such framework could be adapted for applications to multiple use-cases. For instance, it could be applied to fully automated, model-based assessments and training of social cognition and soft-skills, leveraging non-verbal behaviours in VR. This would be relevant in clinical settings, e.g., for diagnostic, prognostic, and treatment response monitoring in neurological and psychiatric patients, including kids with Autism Spectrum Disorders. This would also be relevant for the general population, e.g., in relation to ageing, and for professional training, in any trade that would benefit from assessing and training social cognition and soft skills.

6.4 Conclusion

The approach we developed combines PCM-driven agents, which integrate a three-dimensional model of the subjective perspective of consciousness, and virtual humans. Its overarching aim is to study human behaviours and their psychological determinants in virtual reality experiments. At this point, our implementation remains preliminary and calls for applications in more tasks and contexts than the narrow illustrative entry-game example we chose, as well as empirical validations. We believe, however, that the proof-of-concept we presented offers a promising ground toward that aim.

AUTHORS' CONTRIBUTIONS

David Rudrauf: conceived and developed the principles of the model and algorithm, co-designed the proof-of-concept experiment, performed analyses of the simulation results, and co-wrote the article.

Grégoire Sergeant-Perthuis: developed the mathematical formulation of the model and algorithm, co-developed aspects of the model, co-designed the proof-of-concept experiment, and co-wrote the article.

Yvain Tisserand: helped with the conceptual and technical rationale, implemented the virtual humans and virtual environment, performed analyses of the simulation results, and co-wrote the article.

Valentin de Gevigney: contributed to the revision of the manuscript and its additional simulations.

Teerawat Monnor: contributed to discussions about the design of the proof-of-concept, and to editing the initial version of the article.

Olivier Belli: developed and implemented the code used for the simulations, ran the simulations, and co-wrote the article.

REFERENCES

[1] Prashant Doshi, Xia Qu, Adam S. Goodie, and Diana L. Young. 2012. Modeling human recursive reasoning using empirically informed interactive partially observable markov decision processes. *IEEE Trans. Syst. Man Cybernet. Part A: Syst. Hum.* 42, 6 (2012), 1529–1542. DOI: http://dx.doi.org/10.1109/TSMCA.2012.2199484

[2] D. Rudrauf, G. Sergeant-Perthuis, O. Belli, Y. Tisserand, and G. Di Marzo Serugendo. 2022. Modeling the subjective perspective of consciousness and its role in the control of behaviours. *J. Theor. Biol.* 534 (2022), 110957. DOI: http://dx.doi.org/10.1016/j.jtbi.2021.110957

[3] Jim Blascovich, Jack Loomis, Andrew C. Beall, Kimberly R. Swinth, Crystal L. Hoyt, and Jeremy N. Bailenson. 2002. Immersive virtual environment technology as a methodological tool for social psychology. *Psychol. Inq.* 13, 2 (2002), 103–124.

[4] Celso M. de Melo, Peter J. Carnevale, and Jonathan Gratch. 2014. Using virtual confederates to research intergroup bias and conflict. In *Academy of Management Proceedings*, Vol. 2014. Academy of Management Briarcliff Manor, NY, 11226.

[5] Torsten Wörtwein and Stefan Scherer. 2017. What really matters-an information gain analysis of questions and reactions in automated PTSD screenings. In *Proceedings of the 7th International Conference on Affective Computing and Intelligent Interaction (ACII'17)*. IEEE, 15–20.

[6] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014. SimSensei kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the International Conference on Autonomous Agents and Multi-agent Systems*. 1061–1068.

[7] Julia M. Kim, Randall W. Hill Jr., Paula J. Durlach, H. Chad Lane, Eric Forbell, Mark Core, Stacy Marsella, David Pynadath, and John Hart. 2009. BiLAT: A game-based environment for practicing negotiation in a cultural context. *Int. J. Artific. Intell. Edu.* 19, 3 (2009), 289–308.

[8] Colin F. Camerer. 2003. Behavioural studies of strategic thinking in games. *Trends Cogn. Sci.* 7, 5 (2003), 225–231.

[9] Max Velmans. 1991. Is human information processing conscious? *Behav. Brain Sci.* 14, 4 (1991), 651–726.

[10] John F. Kihlstrom. 1996. Perception without awareness of what is perceived, learning without awareness of what is learned. *The Science of Consciousness*. Taylor & Frances/Routledge, 23–46.

[11] Simon Van Gaal and Victor A. F. Lamme. 2012. Unconscious high-level information processing: Implication for neurobiological theories of consciousness. *Neuroscientist* 18, 3 (2012), 287–301.

[12] J. A. Reggia. 2013. The rise of machine consciousness: Studying consciousness with computational models. *Neural Netw.* 44,(2013), 112–131. DOI : 10.1016/j.neunet.2013.03.011

[13] Giulio Tononi, Melanie Boly, Marcello Massimini, and Christof Koch. 2016. Integrated information theory: From consciousness to its physical substrate. *Nature Rev. Neuroscience* 17, 7 (2016), 450–461.

[14] B. Baars. 1988. *A Cognitive Theory of Consciousness*. Cambridge University Press, Cambridge, UK.

[15] Stanislas Dehaene, Hakwan Lau, and Sid Kouider. 2017. What is consciousness, and could machines have it? *Science* 358, 6362 (2017), 486–492.

[16] Anil K. Seth, Keisuke Suzuki, and Hugo D. Critchley. 2012. An interoceptive predictive coding model of conscious presence. *Front. Psychol.* 2 (2012), 395.

[17] I. Aleksander. 2005. *The World in My Mind My Mind in the World: Key Mechanisms of Consciousness in Humans, Animals, and Machines*. Imprint Academic, Exeter, UK.

[18] Antonio R. Damasio. 1999. *The Feeling of What Happens: Body and Emotion in the Making of Consciousness*. Houghton Mifflin Harcourt.

[19] Olaf Blanke. 2012. Multisensory brain mechanisms of bodily self-consciousness. *Nature Rev. Neurosci.* 13, 8 (2012), 556–571.

[20] A. Chella and R. Manzotti. 2009. Machine consciousness: A manifesto for robotics. *Int. J. Mach. Conscious.* 1, (2009), 33–51.

[21] R. Manzotti and A. Chella. 2018. Good old-fashioned artificial consciousness and the intermediate level fallacy. *Front. Robot. AI* 5, 39 (2018).

[22] William James, Frederick Burkhardt, Fredson Bowers, and Ignas K. Skrupskelis. 1890. *The Principles of Psychology*. Vol. 1. Macmillan, London.

[23] Thomas Nagel. 1974. What is it like to be a bat? *Philos. Rev.* 83, 4 (1974), 435–450.

[24] Steven M. Lehar. 2003. *The World in your Head: A Gestalt View of the Mechanism of Conscious Experience*. Psychology Press.

[25] Bjorn Merker. 2012. From probabilities to percepts A subcortical "global best estimate buffer" as locus of phenomenal experience. *Being Time: Dynam. Models Phenom. Exp.* 88 (2012), 37.

[26] David Rudrauf, Daniel Bennequin, Isabela Granic, Gregory Landini, Karl Friston, and Kenneth Williford. 2017. A mathematical model of embodied consciousness. *J. Theor. Biol.* 428 (2017), 106–131.

[27] G. Riva. 2018. The neuroscience of body memory: From the self through the space to the others. *Cortex* 104 (2018), 241–260.

[28] Louise McHugh and Ian Stewart. 2012. *The Self and Perspective Taking: Contributions and Applications from Modern Behavioral Science*. New Harbinger Publications, California.

[29] Luc Ciompi. 1991. Affects as central organising and integrating factors a new psychosocial/biological model of the psyche. *Brit. J. Psych.* 159, 1 (1991), 97–105.

[30] Simon Baron-Cohen. 1989. Joint-attention deficits in autism: Towards a cognitive analysis. *Dev. Psychopathol.* 1, 3 (1989), 185–189.

[31] David Premack and Guy Woodruff. 1978. Does the chimpanzee have a theory of mind? *Behav. Brain Sci.* 1, 4 (1978), 515–526.

[32] A. K. Seth. 2018. Consciousness: The last 50 years (and the next). *Brain Neurosci. Adv.* 2 (2018).

[33] A. K. Seth and J. Hohwy. 2014. Elusive phenomenology, counterfactual awareness, and presence without mastery. *Cogn. Neurosci.* 5, 2 (2014), 127–128.

[34] A. K. Seth. 2015. Presence, objecthood, and the phenomenology of predictive perception. *Cogn. Neurosci.* 6, 2-3 (2015), 111–117.

[35] A. Revonsuo. 2005. *Consciousness as a Biological Phenomenon*. MIT Press, Cambridge, MA.

[36] Francis Crick and Christof Koch. 1990. Towards a neurobiological theory of consciousness. In *Seminars in the Neurosciences*, Vol. 2. 203.

[37] Bjorn Merker, Kenneth Williford, and David Rudrauf. 2022. The integrated information theory of consciousness: A case of mistaken identity. *Behav. Brain Sci.* 45 (2022).

[38] Kenneth Williford, Daniel Bennequin, Karl Friston, and David Rudrauf. 2018. The projective consciousness model and phenomenal selfhood. *Front. Psychol.* 9 (2018), 2571.

[39] David Rudrauf, Daniel Bennequin, and Kenneth Williford. 2020. The moon illusion explained by the projective consciousness model. *J. Theor. Biol.* 507 (2020), 110455.

[40] David Rudrauf and Martin Debbané. 2018. Building a cybernetic model of psychopathology: Beyond the metaphor. *Psychol. Inq.* 29, 3 (2018), 156–164.

[41] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, and G. Pezzulo. 2017. Active inference: A process theory. *Neural Comput.* 29, 1 (2017), 1–49.

[42] M. A. Apps and M. Tsakiris. 2014. The free-energy self: A predictive coding account of self-recognition. *Neurosci. Biobehav. Rev.* 41 (2014), 85–97.

[43] K. Friston. 2010. The free-energy principle: A unified brain theory? *Nature Rev. Neurosci.* 11, 2 (2010), 127–138.

[44] Jakub Limanowski and Felix Blankenburg. 2013. Minimal self-models and the free-energy principle. *Front. Hum. Neurosci.* 7 (2013), 547.

[45] A. K. Seth. 2014. The cybernetic Bayesian brain. *Open MIND*. MIND Group, Frankfurt am Main.

[46] K. Friston and C. Frith. 2015. A duet for one. *Conscious Cogn.* 39 (2015), 390–405.

[47] Axel Constant, Maxwell J. D. Ramstead, Samuel P. L. Veissiere, John O. Campbell, and Karl J. Friston. 2018. A variational approach to niche construction. *J. Roy. Soc. Interface* 15, 141 (2018), 20170685.

[48] Axel Constant, Maxwell J. D. Ramstead, Samuel P. L. Veissière, and Karl Friston. 2019. Regimes of expectations: An active inference model of social conformity and human decision making. *Front. Psychol.* 10 (2019), 679.

[49] Samuel P. L. Veissière, Axel Constant, Maxwell J. D. Ramstead, Karl J. Friston, and Laurence J. Kirmayer. 2020. Thinking through other minds: A variational approach to cognition and culture. *Behav. Brain Sci.* 43 (2020).

[50] Mateus Joffily and Giorgio Coricelli. 2013. Emotional valence and the free-energy principle. *PLoS Comput. Biol.* 9, 6 (2013), e1003094.

[51] E. Kalbe, F. Grabenhorst, Matthias Brand, J. Kessler, R. Hilker, and Hans J. Markowitsch. 2007. Elevated emotional reactivity in affective but not cognitive components of theory of mind: A psychophysiological study. *J. Neuropsychol.* 1, 1 (2007), 27–38.

[52] Simon Baron-Cohen. 1991. Precursors to a theory of mind: Understanding attention in others. In *Natural Theories of mind: Evolution, Development and Simulation of Everyday Mindreading*, vol. 1. Andrew Whiten and R. W. Byrne (Eds.). Basil Blackwell, Oxford, UK, 233–251.

[53] Heinz Wimmer and Josef Perner. 1983. Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* 13, 1 (1983), 103–128.

[54] C. Lamm, C. Lamm, C. D. Batson, and J. Decety. 2007. The neural substrate of human empathy: Effects of perspective-taking and cognitive appraisal. *J. Cogn. Neurosc.* 19, 1 (2007), 42–58.

[55] Alain Berthoz and Bérangère Thirioux. 2010. A spatial and perspective change theory of the difference between sympathy and empathy. *Paragrana* 19, 1 (2010), 32–61.

[56] J. J. Gross. 1988. The emerging field of emotion regulation: An integrative review. *Rev. Gen. Psychol.* 2, 3 (1988), 271–299.

[57] F. Clément and D. Dukes. 2013. The role of interest in the transmission of social values. *Front. Psychol.* 4 (2013), 349.

[58] G. Gergely and J. S. Watson. 1996. The social biofeedback theory of parental affect-mirroring: The development of emotional self-awareness and self-control. *Info. Int. J. Psycho-Analy.* 77 (1996), 1181–1212.

[59] L. Beckes and J. A. Coan. 2011. Social baseline theory: The role of social proximity in emotion and economy of action. *Soc. Personal. Psychol. Comp.* 5, 12 (2011), 976–988.

[60] P. Fonagy and E. Allison. 2014. The role of mentalizing and epistemic trust in the therapeutic relationship. *Psychotherapy* 51, 3 (2014), 372.

[61] R. Kalisch, M. B. Müller, and O. Tüscher. 2015. A conceptual framework for the neurobiological study of resilience. *Behav. Brain Sci.* 38 (2015).

[62] Geraldine Dawson, Andrew N. Meltzoff, Julie Osterling, Julie Rinaldi, and Emily Brown. 1998. Children with autism fail to orient to naturally occurring social stimuli. *J. Autism Develop. Disord.* 28, 6 (1998), 479–485.

[63] Etienne Koechlin and Alexandre Hyafil. 2007. Anterior prefrontal function and the limits of human decision-making. *Science* 318, 5850 (10 2007), 594–598. DOI:http://dx.doi.org/10.1126/science.1142995

[64] Julie A. Hadwin, Patricia Howlin, and Simon Baron-Cohen. 2015. *Teaching Children with Autism to Mind-read: The Workbook.* John Wiley & Sons, NY.

[65] Antonia F. de C. Hamilton, Rachel Brindley, and Uta Frith. 2009. Visual perspective taking impairment in children with autistic spectrum disorder. *Cognition* 113, 1 (2009), 37–44.

[66] Simon Baron-Cohen, Alan M. Leslie, Uta Frith, et al. 1985. Does the autistic child have a "theory of mind." *Cognition* 21, 1 (1985), 37–46.

[67] Alan M. Leslie and Uta Frith. 1988. Autistic children's understanding of seeing, knowing and believing. *Brit. J. Develop. Psychol.* 6, 4 (1988), 315–324.

[68] Kristine H. Onishi and Renée Baillargeon. 2005. Do 15-month-old infants understand false beliefs? *Science* 308, 5719 (2005), 255–258.

[69] Michael Georgeff, Barney Pell, Martha Pollack, Milind Tambe, and Michael Wooldridge. 1998. The belief-desire-intention model of agency. In *Proceedings of the International Workshop on Agent Theories, Architectures, and Languages.* Springer, 1–10.

[70] Joost Broekens, Doug Degroot, and Walter A. Kosters. 2008. Formal models of appraisal: Theory, specification, and computational model. *Cogn. Syst. Res.* 9, 3 (2008), 173–197.

[71] Colin F. Camerer. 1997. Progress in behavioral game theory. *J. Econ. Perspect.* 11, 4 (1997), 167–188.

[72] Tarek Alskaif, Manel Guerrero Zapata, and Boris Bellalta. 2015. Game theory for energy efficiency in wireless sensor networks: Latest trends. *J. Netw. Comput. Appl.* 54 (5 2015), 33–61. DOI:http://dx.doi.org/10.1016/j.jnca.2015.03.011

[73] Ismael T. Freire, Xerxes D. Arsiwalla, Jordi-Ysard Puigbó, and Paul Verschure. 2019. Modeling Theory of Mind in Multi-Agent Games Using Adaptive Feedback Control. Retrieved from https://arXiv:1905.13225.

[74] Rosemarie Nagel, Andrea Brovelli, Frank Heinemann, and Giorgio Coricelli. 2018. Neural mechanisms mediating degrees of strategic uncertainty. *Soc. Cogn. Affect. Neurosci.* 13, 1 (2018), 52–62.

[75] Piotr J. Gmytrasiewicz and Edmund H. Durfee. 2000. Rational coordination in multi-agent environments. *Auton. Agents Multi-Agent Syst.* 3, 4 (2000), 319–350.

[76] Piotr J. Gmytrasiewicz, Sanguk Noh, and Tad Kellogg. 1998. Bayesian update of recursive agent models. *User Model. User-Adapt. Interact.* 8, 1 (1998), 49–69.

[77] Chris L. Baker, Rebecca Saxe, and Joshua B. Tenenbaum. 2009. Action understanding as inverse planning. *Cognition* 113, 3 (2009), 329–349.

[78] David V. Pynadath and Stacy C. Marsella. 2005. PsychSim: Modeling theory of mind with decision-theoretic agents. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI'05)*, Vol. 5. 1181–1186.

[79] Wako Yoshida, Ray J. Dolan, and Karl J. Friston. 2008. Game theory of mind. *PLoS Comput. Biol.* 4, 12 (Dec. 2008). DOI:http://dx.doi.org/10.1371/journal.pcbi.1000254

[80] Wako Yoshida, Isabel Dziobek, Dorit Kliemann, Hauke R. Heekeren, Karl J. Friston, and Ray J. Dolan. 2010. Cooperation and heterogeneity of the autistic mind. *J. Neurosci.* 30, 26 (2010), 8815–8818.

[81] Dario Bombari, Marianne Schmid Mast, Elena Canadas, and Manuel Bachmann. 2015. Studying social interactions through immersive virtual environment technology: Virtues, pitfalls, and future challenges. *Front. Psychol.* 6 (2015), 869.

[82] Julien Bruneau, Anne-Helene Olivier, and Julien Pettre. 2015. Going through, going around: A study on individual avoidance of groups. *IEEE Trans. Visual. Comput. Graph.* 21, 4 (2015), 520–528.

[83] Tekla S. Perry. 2015. Virtual reality goes social. *IEEE Spectrum* 53, 1 (2015), 56–57.

[84] Sahil Narang, Andrew Best, and Dinesh Manocha. 2019. Inferring user intent using Bayesian theory of mind in shared avatar-agent virtual environments. *IEEE Trans. Visual. Comput. Graph.* 25, 5 (2019), 2113–2122.

[85] Thanh Thi Nguyen, Ngoc Duy Nguyen, and Saeid Nahavandi. 2020. Deep reinforcement learning for multiagent systems: A review of challenges, solutions, and applications. *IEEE Trans. Cybernet.* 50, 9 (2020), 3826–3839.

[86] Yvain Tisserand, Ruth Aylett, Marcello Mortillaro, and David Rudrauf. 2020. Real-time simulation of virtual humans' emotional facial expressions, harnessing autonomic physiological and musculoskeletal control. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents.* 1–8.

[87] Adi Botea, Martin Müller, and Jonathan Schaeffer. 2004. Near optimal hierarchical path-finding. *J. Game Dev.* 1, 1 (2004), 1–30.

[88] Rosenberg Ekman. 1997. *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression using the Facial Action Coding System.* Oxford University Press, NY.

[89] M. A. Wiering and M. van Otterlo. 2012. *Reinforcement Learning: State-of-the-Art.* Springer, Berlin.

[90] Lorenzo Canese, Gian Carlo Cardarilli, Luca Di Nunzio, Rocco Fazzolari, Daniele Giardino, Marco Re, and Sergio Spanó. 2021. Multi-agent reinforcement learning: A review of challenges and applications. *Appl. Sci.* 11, 11 (2021). DOI:http://dx.doi.org/10.3390/app11114948

[91] Julian Jara-Ettinger. 2019. Theory of mind as inverse reinforcement learning. *Curr. Opin. Behav. Sci.* 29 (2019), 105–110. DOI : http://dx.doi.org/10.1016/j.cobeha.2019.04.010

[92] Ying Wen, Yaodong Yang, Rui Luo, Jun Wang, and Wei Pan. 2019. Probabilistic recursive reasoning for multi-agent reinforcement learning. In *Proceedings of the International Conference on Learning Representations*. Retrieved from https://openreview.net/forum?id=rkl6As0cF7.