BIOLOGICAL REVIEWS

Cambridge Philosophical Society

Biol. Rev. (2023), pp. 000–000.
doi: 10.1111/brv.12994

Reading minds or reading scripts? De-intellectualising theory of mind

Derry Taylor[1], Gökhan Gönül[2,*], Cameron Alexander[3], Klaus Züberbühler[1], Fabrice Clément[2] and Hans-Johann Glock[3,4]

[1]*Faculty of Science, Institute of Biology, Department of Comparative Cognition, University of Neuchâtel, Rue-Emile-Argand 11, Neuchâtel 2000, Switzerland*

[2]*Faculty of Letters and Human Sciences, Institute of Language and Communication Sciences, Cognitive Science Centre, University of Neuchâtel, Pierre-à-Mazel 7, Neuchâtel 2000, Switzerland*

[3]*Department of Philosophy, University of Zürich, Zürichbergstrasse 43, Zurich CH-8044, Switzerland*

[4]*Institute for the Study of Language Evolution, University of Zürich, Affolternstrasse 56, Zürich CH-8050, Switzerland*

ABSTRACT

Understanding the origins of human social cognition is a central challenge in contemporary science. In recent decades, the idea of a 'Theory of Mind' (ToM) has emerged as the most popular way of explaining unique features of human social cognition. This default view has been progressively undermined by research on 'implicit' ToM, which suggests that relevant precursor abilities may already be present in preverbal human infants and great apes. However, this area of research suffers from conceptual difficulties and empirical limitations, including explanatory circularity, over-intellectualisation, and inconsistent empirical replication. Our article breaks new ground by adapting 'script theory' for application to both linguistic and non-linguistic agents. It thereby provides a new theoretical framework able to resolve the aforementioned issues, generate novel predictions, and provide a plausible account of how individuals make sense of the behaviour of others. Script theory is based on the premise that pre-verbal infants and great apes are capable of basic forms of agency-detection and non-mentalistic goal understanding, allowing individuals to form event-schemata that are then used to make sense of the behaviour of others. We show how script theory circumvents fundamental problems created by ToM-based frameworks, explains patterns of inconsistent replication, and offers important novel predictions regarding how humans and other animals understand and predict the behaviour of others.

Key words: implicit theory of mind, belief attribution, script theory, schema, development of social cognition, evolution of social cognition.

CONTENTS

* Author for correspondence (Tel.: + 41 32 718 1687; E-mail: gokhan.gonul@unine.ch).

I. INTRODUCTION

Humans stand out in nature by possessing unique features, like articulated language, sophisticated tools, and social institutions. Many uniquely human characteristics are underpinned by a capacity for complex social cognition (Laland & Seed, 2021). In philosophy, psychology, biology, and neuroscience, complex social cognition is thought to consist of a capacity to infer the existence and contents of another individual's inner psychological states. This is known as 'theory of mind' (ToM) (Premack & Woodruff, 1978; Schurz *et al.*, 2021; Slaughter, 2015; Wellman, 2017). According to standard views, ToM allows subjects to reason about the mental states of others and thereby predict their behaviour and communicate (e.g. Wellman, 2014, 2018). However, this label is both ambiguous and presumptuous. For some, 'theory of mind' is simply another term for social understanding; for others, it is a particular highly demanding account of social understanding. The term is therefore associated not only with a topic, but also with a particular approach to that topic, 'theory-theory'. Unfortunately, these terminological problems cannot be circumvented since the term 'theory of mind' is so ubiquitous as to be indispensable. In this review, we use 'theory of mind' to denote the *ability* to hold, and draw inferences from, metarepresentations, i.e. representations of the mental states of others. 'Mentalizing' and 'mind-reading' are considered as synonymous in this context. Similarly, ToM tasks are tasks designed to find evidence of this ability.

There are two classic accounts of how ToM works. According to *theory-theory*, mental states are internal, unobservable entities that we attribute to others based on a folk-psychological theory which allows us to explain and predict their behaviour. This theory can result from innate mind-reading modules (Baron-Cohen, 1995), or be akin to an empirical scientific theory that undergoes change and development during ontogeny (Gopnik & Wellman, 1992, p. 199). It can also be implicit, i.e. beyond the immediate awareness of the subject (Carruthers & Smith, 1996). According to *simulation theory*, we use our own subjective experience to simulate someone else's states of mind. The simulation is either conceived as an explicit process of imaginative inference (Goldman, 1992), or as an implicit process of automatic mirroring (Gallese & Sinigaglia, 2012). Hybrid theories borrow elements from both approaches (e.g. Nichols & Stich, 2003). Except for implicit simulation, both theory-theory and simulation theory share the assumption that subjects form – and carry out inferences on the basis of – representations of the mental representations of others.

Due to an already large and growing dissatisfaction with classic theories of ToM in humans (Section I.1) and non-human animals (Section I.2), the need to provide a de-intellectualised account of social cognition is pressing (Section I.3). We propose that scripts de-intellectualise social cognition, without falling prey to problems that affect *some* of the extant alternatives to ToM, whilst borrowing from the strengths of others (Section II).

Briefly, script theory posits that individuals understand the behaviour of others using event schemata that emerge bottom-up, largely through their own experience, i.e. through participating in events either as agents or as observers of others' interactions. Such ontogenetically acquired experience is thought to be organised around primary knowledge of how basic event structure is perceived in terms of agents, objects, and actions. As a result, individuals accumulate a catalogue of possible scripts through experience that are then applied top-down to recognise and predict the behaviour of others. Script theory avoids many of the problems of ToM accounts (Sections I.1 and I.2), improves on some extant theories that have attempted to de-intellectualise interpretations of human behaviour, and pinpoints a potential basis upon which genuinely mentalistic forms of human social cognition emerge in later ontogeny. Finally, script theory makes novel predictions for experimental paradigms, including the classic change-of-location tasks (Onishi & Baillargeon, 2005).

(1) How might humans make sense of others?

Central to the classic paradigms for explaining ToM is that humans understand and predict how others will behave by attributing mental states to them (Cole, Smith & Atkinson, 2015). Such mental states include beliefs, desires, visual perspectives, and intentions (Apperly, 2012). But the capacity of belief attribution is considered to be the most complex form and therefore merits special attention. Testing this capacity for attributing false beliefs is of paramount importance (Premack & Woodruff, 1978) because it requires simultaneously representing two conflicting views of the world: the test subject's true representation and the agent's false representation.

Belief attribution is typically tested by using some version of the Sally–Anne task (Baron-Cohen, Leslie & Frith, 1985; Wimmer & Perner, 1983). In this task, the subject observes an object changing its location while its owner is absent. Consequently, on her return the owner has a 'false belief' about the actual current location of her object. To evaluate whether subjects can attribute to the owner a false belief, they are asked where the owner will look for her object. Since the measures in this version of the Sally–Anne task are verbal responses, it is an 'explicit' ToM task. Pre-linguistic human infants and animals cannot respond to language-based requests. Therefore, researchers later developed non-verbal tasks which use 'implicit' measures of mental state attribution. In one landmark study (Onishi & Baillargeon, 2005), researchers used a Violation-of-Expectation (VoE) version of the Sally–Anne paradigm to assess prelinguistic infants' anticipation of the owner's search behaviour – whether she would look for her object in its current location (i.e. where it had been moved in her absence), or in its previous location (i.e. where she last saw it). Results indicated a greater degree of surprise (measured by looking time) in response to the former, i.e. the owner searching in the current location. This was interpreted as evidence that infants possess an 'implicit' understanding of others' (false) beliefs. Similar results have been obtained using anticipatory looking (AL) measurements (Southgate, Senju & Csibra, 2007), which further reinforced the distinction between implicit *versus* explicit forms of mental state attribution. At the same time, despite its broad appeal, the very idea of ToM – whether of the theory-theory or simulation theory variety, whether explicit or implicit – has become increasingly controversial, for theoretical, empirical, and methodological reasons.

First, there is growing concern that the classic approach to deciding whether or not an individual possesses ToM profoundly over-intellectualises the cognitive processes of people and animals when they make sense of each other's behaviour (Heyes, 2014; Meunier, 2017; Quesque & Rossetti, 2020; van der Vaart & Hemelrijk, 2014). Over-intellectualisation stems from the fact that ToM requires an individual being able to hold beliefs about the beliefs, desires, and intentions of others, i.e. to have other-regarding meta-representations. But even if it makes theoretical sense to ascribe such meta-representational capacities in the absence of language, empirically they are very difficult to diagnose. This raises the problem of explaining the ontogeny and phylogeny of mind-reading (Liszkowski, 2018). Moreover, one view regarding the emergence of sophisticated communication in both ontogeny and in the human lineage goes back to Grice (1957). It maintains that this emergence is due to the fact that subjects acquire(d) the ability to form complex representations of the communicative intentions of others (Origgi & Sperber, 2000; Scott-Phillips, 2015; Tomasello, 2008). However, the ability to form such meta-representations regarding the beliefs and intentions of others is plausibly an outcome of linguistic competence, at least in part (Lohmann & Tomasello, 2003). Admittedly, others adopt a more pluralistic view: while such cognition *can* play a role in language, it does not necessarily *always* play this role (Andrews, 2009). But even this view faces a serious problem: the meta-representational capacity required to represent communicative intentions can involve up to 4th order intentionality. For example, to inform a recipient of some state of affairs, a signaller must hold a complex intention, such as: 'I *intend* that you *believe* that I *intend* that you *believe* that it is sunny' (Moore, 2017; Sperber, 2000). However, 4th order intentionality regarding the mental states of others appears much later in ontogeny than language – at around 11–12 years (Liddle & Nettle, 2007; Moore, 2017). There is thus the threat of an explanatory circle: how can meta-representational forms of social cognition – i.e. ToM – be acquired in advance of capacities for the kind of interaction and communication that ToM is supposed to explain? Regarding phylogeny, if one accepts the view that communication involves recognising and responding to communicative intentions, and if one accepts the idea that non-linguistic animals lack language because they lack ToM (e.g. Tomasello, 2008), then the difficulties of explaining the emergence of language are simply transposed to the no less difficult problem of explaining the emergence of ToM in phylogeny (Bar-On, 2013). But a theoretically useful account of the nature of ToM and its relation to linguistic communication should not make it difficult, let alone impossible, to explain its phylogenetic and ontogenetic emergence; instead, it should facilitate such a theoretical explanation (Heyes, 2014).

Second, controversy around ToM also stems from difficulties with replicating paradigmatic ToM experiments that use implicit ToM tasks (Baillargeon, Buttelmann & Southgate, 2018; Kulke *et al.*, 2018) to detect ToM in non- and pre-linguistic subjects. Findings in developmental psychology have lowered the age thresholds for various forms of social understanding, with some experiments suggesting implicit recognition of false beliefs in very young infants (Onishi & Baillargeon, 2005; Southgate *et al.*, 2007). These findings have been achieved through both methodological and theoretical shifts. Methodologically, studies began to focus on behavioural indicators of surprise to gauge subjects' expectations about others' behaviour implicitly rather than relying on explicit linguistic responses. Theoretically, the concept of 'implicit' ToM has emerged; it is typically considered to refer to automatic and unconscious tracking of others' mental states (Low & Perner, 2012). This has culminated in a surprising inconsistency of results of ToM experiments across age cohorts (Newen & Wolf, 2020): how can young infants tested on implicit ToM tasks already understand implicitly that others have false beliefs, while even 3-year-olds fail explicit ToM tasks? Further, from an evolutionary perspective it is unclear how implicit ToM could be of use to an organism incapable of acting upon that implicit understanding (Martin & Santos, 2016). Interestingly, concerning both humans (Grosse Wiesmann *et al.*, 2017) and other primates (Kano, Call & Krupenye, 2020) it has been argued that socially dynamic stimuli appear to produce more reliable results, suggesting implicit ToM in the absence of explicit ToM may have some social utility (but see Kulke & Hinrichs, 2021).

As an added complexity, key results demonstrating implicit false belief understanding in young infants have not been consistently replicated, which has stirred further debate (Baillargeon *et al.*, 2018; Kampis, Frances & Kovács, 2020). Attempts to replicate the core findings have failed, either fully or partially, with VoE paradigms generally showing better replication than anticipatory-looking studies (Barone, Corradi & Gomila, 2019) – but see Powell *et al.* (2018).

Non-verbal (i.e. 'implicit') change-of-location tasks have also been administered to human adults (Schneider *et al.*, 2012) and, here again, replication problems emerged (Kulke *et al.*, 2018). A similar inconsistency has been reported across paradigms in animal cognition research, with primates failing action-based but passing implicit false-belief tasks – e.g. using looking time or gaze-direction measures (Flombaum & Santos, 2005; Kaminski, Call & Tomasello, 2008; Krupenye & Call, 2019).

Third, controversy surrounding ToM also arises from the emergence of theories that offer more parsimonious ways to interpret data from ToM experiments (Burnside, Severdija & Poulin-Dubois, 2020; Krupenye & Call, 2019; Quesque & Rossetti, 2020) by putting less emphasis on rich mental representations than, e.g. theory-theory or explicit simulation theories of mind. For each cognitive interpretation of a task performance in animals, a 'lower level' alternative can be proposed (Andrews, 2009; Povinelli & Vonk, 2006). It has been argued that performance in implicit ToM tasks might be explicable by non-ToM mechanisms such as behavioural rules (Perner & Ruffman, 2005) or domain-general cognitive mechanisms (Heyes, 2014); this calls into question the very idea of an implicit ToM. However, many advocates of infant ToM maintain that ToM is more parsimonious than explanations such as behavioural rules (He, Bolz & Baillargeon, 2011; Song & Baillargeon, 2008). They argue that a subject would have to learn excessively many rules to understand an agent's behaviour in different situations, whereas it is in fact more parsimonious for subjects to represent the beliefs of others by means of the subject's ToM. But this response has been objected to on multiple fronts (Perner, 2010; Povinelli & Vonk, 2006; Ruffman, 2014). One counterargument is that, even if it is granted that subjects represent the beliefs of others, it is still necessary for the subject to connect the agent's belief to the specific circumstances that cause the agent to form a belief, and, moreover, to take into account features of the specific context to predict the behaviour that ensues when an agent has a belief (Ruffman, 2014). Simply attributing beliefs to an agent (i.e. a one-step process) is not enough for a subject to understand what the agent will do next. Instead, the subject must first recognise the features in the social environment that lead to an agent forming a belief, then, secondly, attribute the belief to the agent, and finally, relate the agent's belief back to features in the social environment to predict behaviour accurately (i.e. a three-step process). Behavioural rules accounts simply cut out the middle step: subjects encode information about how features of a social environment lead to different behaviours (i.e. a two-step process). More generally, parsimony is always relative to a metric of comparison. Advocates of ToM deem it to be a parsimonious account of social cognition with respect to the number of types of cognitive process involved (just one, albeit a demanding form of meta-cognition), whereas behavioural reading advocates appeal to parsimony with respect to the number of cognitive steps involved in social cognition.

Fourth, evidence of adults not making mental state attributions in ToM tasks counts against the idea that ToM is the only – let alone indispensable – way that humans understand others. Even in experimental studies in which adult humans are explicitly made aware of others' false beliefs, they often fail to use this knowledge to make sense of the actions of others (Keysar, Lin & Barbar, 2003) – but for weak evidence of adult implicit ToM, based on a non-systematically replicated paradigm, see Schuwerk *et al.* (2018). Additionally, adult humans do not engage in mental state attribution in the videos used in the original Onishi & Baillargeon (2005) study, unless they are clearly instructed to focus on the presumed mental states of the agents (Low & Edwards, 2018). It is likely that adult humans often tend to explain the behaviour of others in terms of *behavioural routines* (familiar action sequences), while attributing mental states only exceptionally, mainly when agents behave in unusual ways or in breach of expected routines (Kaufmann & Clément, 2014; Korman & Malle, 2016). Although adult humans are certainly capable of meta-representational attributions and explicit false belief understanding, the extent to which they rely on such processes in their daily lives is subject to empirical investigation. Furthermore, neuroscientific theory and data indicate that neuro-computational resources are finite, implying that simpler processing mechanisms will be favoured whenever possible (Lieder & Griffiths, 2020).

To conclude, it should not simply be assumed that other-regarding meta-representations are the default way that humans assess everyday social situations (Corradi-Dell'Acqua *et al.*, 2015; Kulke & Hinrichs, 2021). Instead, we expect that even adult humans typically deploy cognitively simpler means to understand others (Fiebich, Gallagher & Hutto, 2016; Kaufmann & Clément, 2014), in particular social scripts.

(2) How might animals make sense of others?

The term 'theory of mind' was originally coined by seminal research with captive chimpanzees (*Pan troglodytes*) (Premack & Woodruff, 1978). Decades of subsequent work has shown that a range of non-human animals may possess at least some cognitive abilities that are assumed to be related to the broader spectrum of ToM in humans (Krupenye & Call, 2019): non-human apes can assess the visual perspectives of others (Okamoto-Barth, Call & Tomasello, 2007), understand their goals (Yamamoto, Humle & Tanaka, 2012), and potentially even their knowledge states (Kaminski *et al.*, 2008).

At the same time, research on false belief understanding in great apes and other animals has largely failed to find evidence for this capacity (Call & Tomasello, 1999; Kaminski *et al.*, 2008; Krachun *et al.*, 2009). One reason for this, some have argued, might be that many false belief tasks place demands on other cognitive capacities, such as working memory or inhibitory control, which may impair great apes in their performance during false-belief tasks (Krupenye & Call, 2019). However, in a recent eye-tracking study based on anticipatory looking it has been possible to remove some of these extrinsic cognitive demands. This has provided the first evidence of false belief understanding in great apes (Kano *et al*., 2019; Krupenye *et al*., 2016) which appears to be robust *vis-à-vis* some alternative interpretations (e.g. submentalising; see Krupenye *et al*., 2017). More recently, this ability has been claimed for Japanese macaques (*Macaca fuscata*); there it appears to be related to neural activity in the medial prefrontal cortex, just as in humans (Hayashi *et al*., 2020). Although results from these studies are compatible with the hypothesis that great apes, and possibly other non-human primates, attribute false beliefs to others, there are nonetheless alternative explanations (e.g. behavioural rules; Krupenye & Call, 2019). The poor replicability of similar human studies has also been cited as a cause for concern (Horschler, MacLean & Santos, 2020). Importantly, great apes possess well-documented long-term and episodic memory capabilities (for a comprehensive review, see Lewis, Berntsen & Call, 2019). These are likely to structure their expectations about others' behaviours and social interactions (Kano & Hirata, 2015), potentially offering an alternative non-ToM explanation of great apes' apparent success in some implicit ToM studies (e.g. Krupenye *et al*., 2016).

(3) Outstanding issues in de-intellectualising ToM

Prompted by the difficulties facing ToM, there have been many recent attempts to de-intellectualise how humans interpret the behaviour of others. These include, but are not limited to, behaviour reading/behavioural rules (Perner & Ruffman, 2005), statistical learning/developmental enrichment (Ruffman, 2014), submentalising (Heyes, 2014), interaction theory (Gallagher & Hutto, 2008; Newen, Welpinghus & Juckel, 2015), teleology (Gergely & Csibra, 2003; Perner, Priewasser & Roessler, 2018; Perner & Roessler, 2012), mental files and vicarious representations (Doherty & Perner, 2020; Nanay, 2020; Newen & Wolf, 2020), the two-systems account (Apperly & Butterfill, 2009; Butterfill & Apperly, 2013), the awareness relations account (Martin & Santos, 2016), knowledge-first approaches (Phillips *et al*., 2021), and naive sociology (Kaufmann & Clément, 2014). These resemble each other in their efforts to make understanding others cognitively less demanding than ToM requires. In this section we briefly discuss the most salient of these extant de-intellectualising approaches, highlighting some difficulties that we propose to overcome through script theory.

Some accounts arguably have limited applicability beyond a specific context, namely explaining performance in implicit ToM tasks. As a result, they shed little light on how human social cognition develops in ontogeny beyond the essentials of predicting what others will do next. For example, the *behavioural rules* account was proposed in part to explain the finding that infants can predict what the agent will do next following a toy's translocation from a yellow box to a green box in the false-belief (FB)-yellow condition of Onishi & Baillargeon's (2005) implicit ToM study. The observer uses a behavioural rule, 'agents look for an object where they last saw it', to predict that the agent will search in the green box. Importantly, the rule can be known and followed by the observing infant, *without* the infant having a conception of agents as having minds that mediate between seeing and acting (Perner & Ruffman, 2005). The strength of the behavioural rules approach lies in its 'lean' interpretation of the data (Ruffman, 2014). However, it is not clear how phenomena associated with ToM going beyond action-prediction, such as pretend play, deception, or communication, can be explained by behavioural rules alone. For example, play and deception depend on a means–end dissociation – using an object or communicative act for some purpose other than the one it is typically used for; but in a behavioural rule, the means and the ends are always closely coupled, despite some variation in the perceptual characteristics in means and ends that endow behavioural rule accounts with some degree of generality. And since these capacities appear in a variety of non-human animals, the benefits of explaining action prediction in terms of behavioural rules starts to wane. Another shortcoming of explaining social cognition exclusively through behavioural rules is that it fails to do justice to the fact that, even below the threshold of full-blown mind-reading in the form of belief attribution, subjects can make sense of goal-directed behaviour guided by perception. For example, captive chimpanzees select appropriate tools to transfer to conspecifics if they have visual access to what the conspecific is trying to do (Yamamoto *et al*., 2012). This capacity suggests a flexible sensitivity to the goals pursued by others and to what they have visual access to, rather than the rigid application of general rules in ways that are insensitive to context. Script theory bears similarities with behaviour-reading accounts, but there are significant differences, such as the central role of goal-understanding, the hierarchical structure in goals and routines give rise to each other, flexibility, and the representation of salient features of social environments (see Sections II.1 and II.2).

The *statistical learning* approach developed by Ruffman (2014) also builds on the idea that infants are behaviour readers. They can pass implicit ToM tasks because they are excellent learners of statistical regularities in behavioural patterns, and they have biased attention towards human faces (especially eyes) and human motion, allowing them to predict how events unfold. Statistical learning in infants has been demonstrated concerning word-segmentation (Saffran, Aslin & Newport, 1996), word-object pairings (Estes *et al*., 2007; Hay *et al*., 2011; Smith & Yu, 2008) and word categorisation (Erickson, Thiessen & Graf Estes, 2014; Saffran & Kirkham, 2018). Furthermore, it is not limited to language acquisition but is domain general, as evidenced by infants learning statistical patterns related to visually presented object sequences (Kirkham, Slemmer & Johnson, 2002). According to the statistical learning view, infants solve implicit ToM tasks by learning statistical regularities in the familiarisation phase of experiments, related to the agent's gaze direction and motor behaviours, which enables them to predict subsequent behaviour (Ruffman, 2014). Infants also bring general knowledge of statistical regularities of gaze direction and behaviour learned in everyday life to bear on various aspects of implicit ToM tasks, such as behavioural rules of the kind spelled out earlier: 'agents look for an object where they last saw it'. The statistical learning account draws attention to a crucial precondition of the ontogenetic emergence of social cognition. However, the possibility of one-shot learning in non-human animals (Deshpande, Van Boekholt & Zuberbuhler, 2022; León *et al.*, 2022) and humans (Lee, O'Doherty & Shimojo, 2015; Roediger & Arnold, 2012) alike, and the importance of episodic memory in solving experimental tasks (e.g. Clayton & Dickinson, 1998; Gershman & Daw, 2017; Kano & Hirata, 2015) speaks in favour of the idea that even in the absence of repeated exposure to stimuli both pre- and non-linguistic subjects can form the suggestively sophisticated event representations required by scripts.

Interaction theory proposes that social cognition has its roots in directly perceiving the emotions and goals of others by observing bodily movements, gestures, and facial expressions (Gallagher & Hutto, 2008). Direct perception is further developed *via* familiarisation with narrative practices (i.e. regularly repeating scenarios which embed basic forms of direct perception in broader social expectations about how agents should behave; Gallagher, 2020). Notably, infants and non-human primates can directly perceive the goals of conspecifics, whilst the sophisticated socio-cognitive abilities of adult humans are further enabled by exposure to narrative practices of increasing complexity. On the face of it, interaction theory and our account of scripts have much in common (see Section II.1). However, it is difficult to characterise direct perception with much precision (Nanay, 2020), and this leads to problems formulating testable hypotheses about the limits of what infants and non-human primates can perceive 'directly'. Furthermore, with respect to the apparent success of infants in some implicit ToM tasks, interaction theory faces some difficulties. Its proponents (Gallagher & Povinelli, 2012) argue that 15-month-old infants pass Onishi & Baillargeon's (2005) test thanks to their capacity to pick up affordances pertinent to other agents. This capacity, Gallagher & Povinelli (2012) argue, develops through repeated interactions with caregivers. However, it is left unexplained why an infant's capacity to interact successfully based on attending to affordances pertinent to another agent should transpose so readily from interactive to observational contexts. Our own account, scripts, suggests that it is because infants form rudimentary representations of routine behaviours that they encounter in *both* interactive *and* observational contexts (see Sections II.1 and II.2). Further, we contend that performance in active choice tasks is often enhanced relative to purely observational tasks (e.g. Kaminski *et al.*, 2008; Setoh, Scott & Baillargeon, 2016) because subjects have more active experience in interaction with the social and physical world than in purely observational experience.

According to *knowledge-first approaches*, non-linguistic subjects ascribe knowledge to others in advance of ascribing beliefs (Clément & Koenig, 2018; Phillips *et al.*, 2021). Whilst this approach may explain why 3-year-olds fail explicit ToM tasks, which rely on subjects attributing to agents' false beliefs rather than simply ignorance, that line of reasoning cannot explain infants' apparent *success in some implicit* ToM tasks. This suggests that, in the case of infants, something else must be at play. Phillips *et al.* (2021) rightly point out the replication problems with implicit ToM tasks, yet without explaining the conflicting results.

Submentalising, by contrast, explicitly tries to explain the findings of Onishi & Baillargeon's (2005) study. If mentalising is equivalent to ToM, then submentalising is keeping track of mental states without representing them. Submentalising takes the 'social' out of social cognition: domain-general cognitive mechanisms applied in a social context are all that is required for an observer to predict what others will do next, without requiring that the observer understands the mental states of others (Heyes, 2014, 2017). Whilst the study of Onishi & Baillargeon (2005) can be interpreted in terms of submentalising, findings from subsequent implicit ToM tasks administered to infants, adults, and apes, have not been amenable to this interpretation (Kano *et al.*, 2017; Krupenye *et al.*, 2017; Surian & Franchin, 2020). Submentalising has also been experimentally tested and ruled out as an explanation for great ape success in an implicit ToM task. Heyes (2017) suggested that the apparent success of great apes in a ToM task using eye-tracking (Krupenye *et al.*, 2016) might be explained in terms of the test subjects submentalising (i.e. the test subjects detected the demonstrator's behavioural cues). However, in a follow up experiment, Krupenye *et al.* (2017) controlled for the submentalising hypothesis by using an inanimate stimulus which lacked agential or social features in one of the test conditions. If submentalising was how great apes pass this task, then such inanimate controls would not alter their performance in the task, but this was not the case. At the very least great apes but also human infants can extract and process richly social cues, although we maintain that this is not necessarily to be explained in terms of mentalising. Finally, there is also converging evidence that humans and non-human animals have a dedicated cognitive mechanism for detecting gaze direction. This mechanism is *domain specific* and, in adult humans at least, modulated by higher-level *social* cognition (Teufel *et al.*, 2009). Moreover, inanimate stimuli such as arrows cause subjects to shift attention to any object on the congruent side of the arrow whilst eyes cause subjects to attend to the specific location indicated by gaze direction (Marotta *et al.*, 2012; Phillips, 2021). This difference, as well as signatures of different attentional mechanisms triggered by animate and inanimate cues (Marotta, Román-Caballero & Lupiáñez, 2018), does not support the submentalising hypothesis that domain-general mechanisms are responsible for detecting gaze direction.

Another influential account is the *two-systems account* (Apperly *et al.*, 2006; Low *et al.*, 2016). It is designed to solve two problems touched on already: why do infants but not 3-year-olds appear to pass a version of the ToM task and how can adult reasoning about beliefs be at times automatic and at other times non-automatic? To explain these discrepancies, the two-systems account proposes that a minimal mind-reading system, which is evolutionarily ancient and develops early in human ontogeny, is used in automatic belief-reasoning, whilst a flexible mind-reading system is used in non-automatic belief reasoning. The flexible system emerges following developments in executive function and the acquisition of language (Apperly, 2010). Importantly, the minimal system is used in the absence of language, i.e. in change-of-location tasks with looking time and gaze direction as measures. It does not encode information about an agent's beliefs in propositional form (as typically assumed by theory-theorists), but as 'belief-like' states whose contents are determined by relations pertaining between an agent, objects, and object locations. Minimal mind-readers, such as pre-linguistic infants and non-linguistic great apes, form representations of the belief-like states of others – so-called 'registrations'. However, it has been argued that if this minimal system is the same across human infants and non-human primates, then a problem arises. Infants and non-human primates seem to perform differently in implicit ToM tasks: infant looking time is different across 'true-belief' and 'false-belief' conditions, whilst rhesus macaques (*Macaca mulatta*) show no difference in looking duration (Marticorena *et al.*, 2011; Martin & Santos, 2014, 2016).

The *awareness relations* account picks up on this discrepancy between infant and non-human primate performance in implicit ToM tasks. It suggests that non-human primates form representations of awareness relations exclusively between agents and true (yet not false) information (Martin & Santos, 2016). The account is right in pointing out that subjects can know what conspecifics have perceptual access to, without knowing what they believe. The awareness relations account denies that non-human primates even form representations of the knowledge and ignorance states of others (Call & Santos, 2012; Okamoto-Barth *et al.*, 2007; Rosati, Santos & Hare, 2010; Whiten, 2013). So, in some respects it is the very opposite of the knowledge-first approach. But there is as yet no awareness relations account of *how* observers' undisputed knowledge of the *perceptual access* of agents interacts with an understanding of agents' *goals* (Horschler, Santos & MacLean, 2021). This is a weakness, since the way observers process information regarding perceptual access *and* goals is especially important when it comes to making sense of the complex and hierarchically structured action sequences highlighted by script theory.

Teleology contends that observers make sense of agents by representing them as pursuing an objective (non-mental) goal and as operating rationally in the pursuit of that goal given circumstantial constraints. We share this general outlook, while also emphasising that the observer represents goal-directed behaviour as *guided by perception*. We also specify explicitly the circumstantial constraints of agency, as causal interrelations between actions and events in the form of 'entry' and 'exit' conditions and as relevant features of the environment in the form of 'props' (see Section II.1). Finally, we share with the teleology account the view that while understanding action as guided by perception is a lower-level form of 'mind-reading', it does not require full-blown belief attribution. Instead, these contributions of perception to action are assumed to be obvious to an observer in what one might call 'naïve teleology' (Gergely & Csibra, 2003). Furthermore, according to naïve teleology, the observer simply assumes that the agent shares its evaluative perspective of a state achieved by its behaviour and its perspective on facts pertinent to achieving it. It therefore has a problem accounting for what happens when the agent has a different view of facts and/or values.

By contrast, 'teleology-in-perspective' tries to account for this by assuming that children acquire the ability 'to use their teleology within the other's [agent's] perspective' (Perner *et al.*, 2018, p. 106, see also pp. 103–104). But, as teleologists-in-perspective realise, they face the problem of explaining *how the observer knows* what the agent's perspective is. At this point they ultimately rely on *mental files*. Mental files are internal units of stored or acquired information about an entity (Nanay, 2020; Recanati, 2012). They have been invoked to explain social cognition, and in particular the findings of false-belief tasks. The idea is that observers can have two files of a single object in a change-of-location set-up: a normal file reflecting the observer's own perspective on the situation, and a 'vicarious file' indexed to the agent the observer is making sense of. Implicit understanding reflects the fact that the observer has a precarious file for the agent, explicit understanding is achieved once the observer is able to link that file to its normal file (Doherty & Perner, 2020; Nanay, 2020; Newen & Wolf, 2020). Mental file theory is right to emphasise the role of different kinds of awareness of perspectives in different types of social cognition. But it postulates a complex 'cognitive architecture' involving both information to which the observer has direct access and information on behalf of an agent. It assumes that the observer is capable of indexing information to agents, thereby assuming a capacity to track different agents and their perspectives, and to remember these pieces of information. Perhaps this additional complexity in the number of cognitive feats can be defended by appealing to the fact that file theory does not assume meta-cognition. In fact, however, it has simply replaced meta-cognitive representations with representations 'imported' into the observer's mind *on behalf of agents*, so to speak. And it has yet to explain what the linking of vicarious and normal files amounts to and how it comes about. In order for the agent to distinguish its normal file of a particular object from its vicarious file of that same object, and to identify the latter as indexed to another subject, the capacity to distinguish the two perspectives is presupposed rather than explained.

The *naïve sociology* hypothesis shares many features with our script account. Its main claim is that social cognition cannot be reduced to a calculus based on others' mental states. In daily life, many actions are predictable because people normally react to the social environment in specific ways (Fiske, 1992; Kaufmann & Fabrice, 2003). On the one hand, social behaviour depends on social relationships linking individuals. When a dominant individual crosses the path of a subordinate, for instance, observers expect that the latter will adopt a respectful attitude, follow the dominant's instructions, and allow her to benefit from more resources. Such expectations are already present in infants (Mascaro & Csibra, 2012; Pun, Birch & Baron, 2017; Thomsen *et al*., 2011), and they develop further during childhood (Charafeddine *et al*., 2015). Similar dispositions seem to exist within non-human primate communities, notably in relation to hierarchy and alliances (Goodall, 1986; Pun *et al*., 2017). Besides this relational aspect, human interactions are often organised by rules and conventions. Given the role endorsed by an agent and the social context, an observer familiar with that culture can predict what will happen next with a high level of confidence (Kaufmann & Clément, 2014). For instance, one can add a social rule to the Sally–Anne false-belief task specifying that the object that Sally is looking for *must be placed* in the first location (where Sally put it). Given such a rule, 3-year-old children's predictions no longer differ from those of older children: using the rule, they predict that she will look for it at that location after the change of location by Anne (Clément, Bernard & Kauffmann, 2011). This account is therefore in line with the script framework that we will develop in Section II. However, naïve sociology aims more specifically at contexts with explicit social norms, which, furthermore, are specific to a given culture. A more general framework is therefore needed to account for non-human primates and for explaining how observers can predict goal-directed behaviours in more diverse contexts.

In what follows, we propose a theoretical framework, *script theory*, that, we believe, adequately explains current findings concerning alleged mind-reading capacities, resolves some of the aforementioned problems in the literature, provides a more general account of how individuals make sense of others, and gives rise to novel experimental predictions within ToM paradigms and beyond. In addition to downgrading the cognitive complexity required to make sense of others' behaviour, the account offered here is a general framework for studying social cognition that goes beyond the topic of ToM. The novelty of our approach lies in adapting an account originally developed for artificial intelligence (AI) research, scripting (Abelson, 1981; Schank & Abelson, 1977), to explain how linguistic and non-linguistic agents comprehend event sequences. A related attempt was made earlier by Worden (1996), who sought to explain how monkeys understand their social worlds in terms of a catalogue of familiar event sequences. We expand this line of reasoning by substantially modifying the AI script account to unravel how *non-linguistic* agents, including great apes and human infants, make sense of others' behaviour.

II. SCRIPTS *versus* MINDREADING: EXPLANATIONS AND PREDICTIONS

(1) Explanations

Script theory is rooted in the idea of *schema* (Mandler, 2014), which are remembered frameworks of knowledge about objects or topics (Bartlett, 1932; McVee, Dunsmore & Gavelek, 2005). It is closely related to and supported by some earlier accounts of social understanding (Worden, 1996), and some recent accounts of event cognition (Farag *et al*., 2010; Franklin *et al*., 2020; Zacks, 2020). Scripts are a particular type of schema; they outline *action sequences* comprising scenes. Scenes are typically *hierarchically organised* (Farag *et al*., 2010; Mesoudi & Whiten, 2004) and *causally interrelated* as *chunks* (Abelson, 1981; Mandler, 2014; Nelson & Gruendel, 1986). Such schematas can have at least three different kinds of function for, respectively, *agents*, *observers*, and us as *theoreticians*. Scripts can *guide* the activities of an *agent*, who follows the script in a sequentially structured activity. They can enable an *observer* to *make sense* of the activities of one or more agents. And they can be invoked by a theorist to *explain the accomplishments* of agents and/or observers. Herein, we shall mainly be concerned with scripts in the second capacity, i.e. as predictive tools for an observer. We argue that scripts emerge through experience of one's own behaviour and that of others. Accordingly, they are acquired in a bottom-up manner, in that they are built up through experience and can be applied subsequently in a top-down manner such that the resultant script can be used to interpret and predict the behaviour of others (e.g. Zacks, 2020). Importantly, scripts do not prescribe specific motor actions, although some scripts may be more rigid or 'strong' than others (Abelson, 1981; Albarracin *et al*., 2021; Gioia & Poole, 1984). Instead, they are defined by *goals* and *sub-goals*. This means that agents can seek equifinal means for fulfilling a script, if a particular means typically used to achieve a given end is unavailable (Schank & Abelson, 1977). It is crucial to note that we use the term 'goal' in a more nuanced sense than earlier proponents of script theory. From our perspective, goals are objective states of the world which are brought about by actions targeted towards objects and other agents. The relation between agent and goal is learned through the observer's experience. If the goal is not prioritised, the observer may focus on other characteristics, such as how the action is performed or its trajectory. Later in development, agents can also ascribe having a goal to other agents (e.g. the goal of learning something by reading a book).

Scripts have two main structural features: *ontology* and *scenes* (Abelson, 1981; Schank & Abelson, 1977). Scenes are *sequences of acts*; each of these acts serves a *sub-goal* that contributes, in varying degrees, to fulfilling the overall goal of the script. Consider the paradigmatic example of a script: the restaurant script (Schank & Abelson, 1977) (see Fig. 1). From the perspective of a customer, going to a restaurant comprises various scenes (e.g. entering the restaurant, consulting the menu, ordering food, eating food, and paying the bill). Each of these scenes contributes to fulfilling the customer's overall goal of eating food prepared by the restaurant. Outside observers familiar with restaurants can use the script to make sense of a customer's behaviour in the restaurant (e.g. consulting the menu before ordering). The ontology of a script refers to what is pertinent within the scene. It includes *roles* which are executed by agents

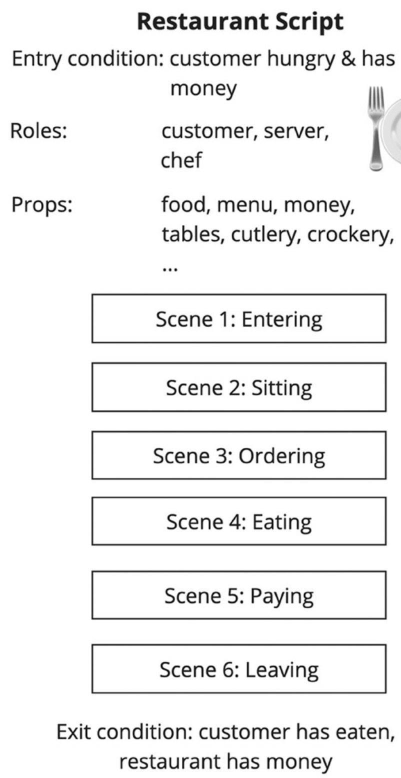


Fig. 1. The restaurant script.

who perform particular functions, *props* (objects that are used by agents) and a *goal* (or goals). For instance, customers and servers are roles performed by agents who order food and take orders for food respectively. The props in a script can vary: tools (e.g. cutlery, crockery, money as a means of payment, etc.), food, or conventional signs (e.g. menu, table numbers, etc.). Props and roles are interrelated; together they afford ways for the goal of a scene to be fulfilled. For example, a customer picking up a menu, consulting it and then looking up to find the server affords the server an opportunity to take the order from the customer. This fulfils the goal of one scene of the restaurant script. From the theorist's perspective, roles, props and overarching goals are types, in that they are not confined to specific occasions. At the same time, observers may recognise those types on the basis of specific cues in particular contexts. For example, for young children to recognise dominance roles, personal history and/or perceptual cues such as body size may suffice (Thomsen, 2020).

The conditions required for a scene to start are known as *entry conditions*. The conditions required to move on from a scene are the *exit conditions*. Exit conditions can also be considered as the goal of the scene or indeed of the script more generally. With respect to the food-ordering scene in the restaurant script, having ordered food is both the exit condition for this scene and the entry condition for the next scene of the script to commence. There may be multiple conditions that must be satisfied for one scene to end and another to begin. For example, for the food-ordering scene to commence, it might be necessary that all customers at a table have finished consulting the menu *and* that they agree to place their order. If the exit condition(s) of a scene are not yet satisfied or if the entry condition(s) for the next scene are not met, that scene is not expected to start. For example, if an observer notices that the customer in the restaurant continues to consult the menu, the entry condition for ordering food is not met and the observer's script does not proceed further until that condition is met. Entry and exit conditions illustrate the causal interrelations within the script. *Causal chaining* in script theory is a principle for generating expectations about behaviour that rely on treating actions and/or scenes as connected by causal links. For instance, an agent and the props of a scene need to be related in such a way that the agent can detect and/or manipulate the prop. The agent's actions are expected to unfold in accordance with the observer's causal understanding of actions and scenes. If the causal relations between an agent fulfilling a particular role and the props are disrupted, such that entry/exit conditions are not fulfilled, these expectations will be disappointed.

At this point it is crucial to distinguish two different cases. In the first, *Suspended Script*, if the causal interrelations between the props and roles are upset, yet the agent behaves as if they had not been disrupted and continues to pursue the path predicted in the observer's script, the observer's script is 'suspended', because the entry conditions required for moving on to the next scene are not yet met. For example, if the customer stops eating to go to the bathroom, the observer's restaurant script is suspended, and resumes upon the customer's return. Similarly, if a server responds to unusual noises in the kitchen, the observer's restaurant script is suspended, and resumes upon the server resuming the actions characteristic of that role. In the second, *De-activated Script*, if the agent ceases to pursue the goal, the script becomes 'deactivated', and the observer no longer expects the agent to act in accordance with the script. These include cases where a highly salient stimulus triggers an alternative script (e.g. a fire alarm triggering an escape script), or where agents start to pursue different goals (e.g. a customer suddenly leaves the restaurant to greet a friend she has seen outside). We contend that observers will update their expectations about how agents behave in such scenarios, according to how they experience the script unfolding over time.

Historically, script theory has often presupposed linguistic competence (but see Worden, 1996). Thus far we have introduced the various features of script theory with the example of a restaurant script, which presupposes established conventions (e.g. a division of labour and implicit contractual obligations between customers and the restaurant owners) and linguistic competence (e.g. the ability to read menus and request food verbally or with sign-language). However, focusing solely on such cases gives a misleading picture of the scope of scripts. We argue here that scripts also provide a plausible *non-linguistic cognitive architecture for social cognition* based on observable action. To exemplify the plausibility of non-linguistic scripts, we show how a script with the aforementioned framework is instantiated by the case of chimpanzee honey foraging (Fig. 2).

We suggest that the principle of causal chaining, and the different ways in which an observer's expectations can be violated (i.e. script suspension and script de-activation) apply in non-linguistic cases, i.e. cases in which neither agent nor observer possesses language. For example, an observing chimpanzee's script for a conspecific collecting honey can be suspended if the chimpanzee who is collecting honey appears to be distracted by food provisioning calls but shows signs of aiming to resume collecting honey. By contrast, an observing chimpanzee's honey-collecting script might be deactivated if the chimpanzee collecting honey responds to alarm calls or is herself threatened and shows no signs of aiming to resume collecting honey.

We contend that scripts are plausible cognitive mechanisms (Franklin *et al.*, 2020; Stawarczyk, Bezdek & Zacks, 2021), one that even pre-linguistic subjects may utilise to make sense of the world (Maffongelli, Antognini & Daum, 2018; Pace, Carver & Friend, 2013). This holds for three reasons.

First, script theory builds on a number of findings regarding agency-attribution and simple goal detection in primates. Many primates can understand self-propelled actions and distinguish animate from inanimate from very early on (Rosa Salva, Mayer & Vallortigara, 2015), thereby providing a basis for distinguishing between roles and props in a script. There is evidence for a cross-species ability to attribute basic forms of agency to others, in the absence of representations of mental states. It also appears that young children (Kosie & Baldwin, 2021) and perhaps adult great apes (Kano & Call, 2014) are able to recognise goal-directed behaviour, provided they have enough contextual information such as salience, familiarity, and other external cues (Elsner & Adam, 2021; Paulus *et al.*, 2017).

The second reason arises from an evolutionary perspective. Scripts capture organised behavioural patterns in response to a given situation. For this reason, we expect that scripts share a basis with simpler behavioural responses to regularly repeating states/stimuli, such as *fixed action patterns*, stereotypical and species-specific action sequences such as courtship play that have been well documented across species (Raby & Clayton, 2009; Tinbergen, 1951). Moreover, in line with other accounts of behaviour (Byrne & Russon, 1998; Orbell & Verplanken, 2018; Radvansky & Zacks, 2014), scripts involve goal-directed behaviours of both habitual and flexible kinds. In this respect, they mirror the action execution capacities of many higher mammals. And because of the presence of these *execution* capacities, there is also a premium for conspecifics to *understand* the interplay between flexible and habitual behaviours in their action anticipation.

The third reason arises from a neuroscientific perspective. Ever since the discovery of so-called mirror neurons and mirror brain areas, there has been growing evidence that executing and observing actions share a pattern of cortical activations in both monkeys and humans (Breveglieri *et al.*, 2019). The same line of empirical research further indicates that this common neural substrate enables an understanding not just of biomechanical contractions and movements, but also of the goals at which an action is directed (Uithol *et al.*, 2012). And it has finally been suggested that this provides a building block for social cognition that precedes knowledge of the propositional contents of mental states like believing, desiring and intending (Gallese & Sinigaglia, 2018). Without taking a stance on all the details of and conclusions drawn from these ongoing investigations, they make it at least *prima facie* plausible that a capacity for employing scripts in understanding and explaining the behaviour of others as an observer can recruit some of the same neurophysiological resources as the capacity to engage in scripted behaviour as an agent.

(2) Development of scripts and action understanding

How scripts develop in early ontogeny has not been well studied, unlike perception of simple events, or action understanding. The developmental literature is mostly based on later developments of script formation and understanding over preschool or school years, stemming from verbal narratives of children (Hudson, Fivush & Kuebli, 1992; Mandler, 2014; Schank & Abelson, 1977). Nevertheless, from the literature on object, agent and event perception we may infer that infants have some building blocks to form scripts already in their first year. As indicated above, scripts represent spatio-temporally organised wholes that are structured hierarchically around goals (and sub-goals) and consist of causally interrelated scenes as chunks. Scripts create expectations by utilising prediction and learning. The following outlines hypotheses on how human infants form and understand scripts.

In human infants, script acquisition and use are based on the early emergence of some mutually complementary perceptual and cognitive abilities for (*i*) recognising objects, agents, actions, (*ii*) recognising goals, (*iii*) segmentation and hierarchisation of event sequences, and (*iv*) generalisation and transfer (Burge, 2018; Wilson, Zuberbühler & Bickel, 2022). In subsequent development, objects turn into props, agents acquire roles, object–agent–action interrelations are constructed to realise agents' goals, and event sequences are segmented into meaningful units (scenes) based on the goal hierarchy of the script.

(a) *From objects and agents to props and roles*

Infants have an intuitive understanding of the physical world which facilitates their *inanimate object* detection and representation (Spelke & Kinzler, 2007). Infants can also create *perceptual categories* of objects based on colour, size, texture, and shape in their first months (Mandler, 2000; Quinn

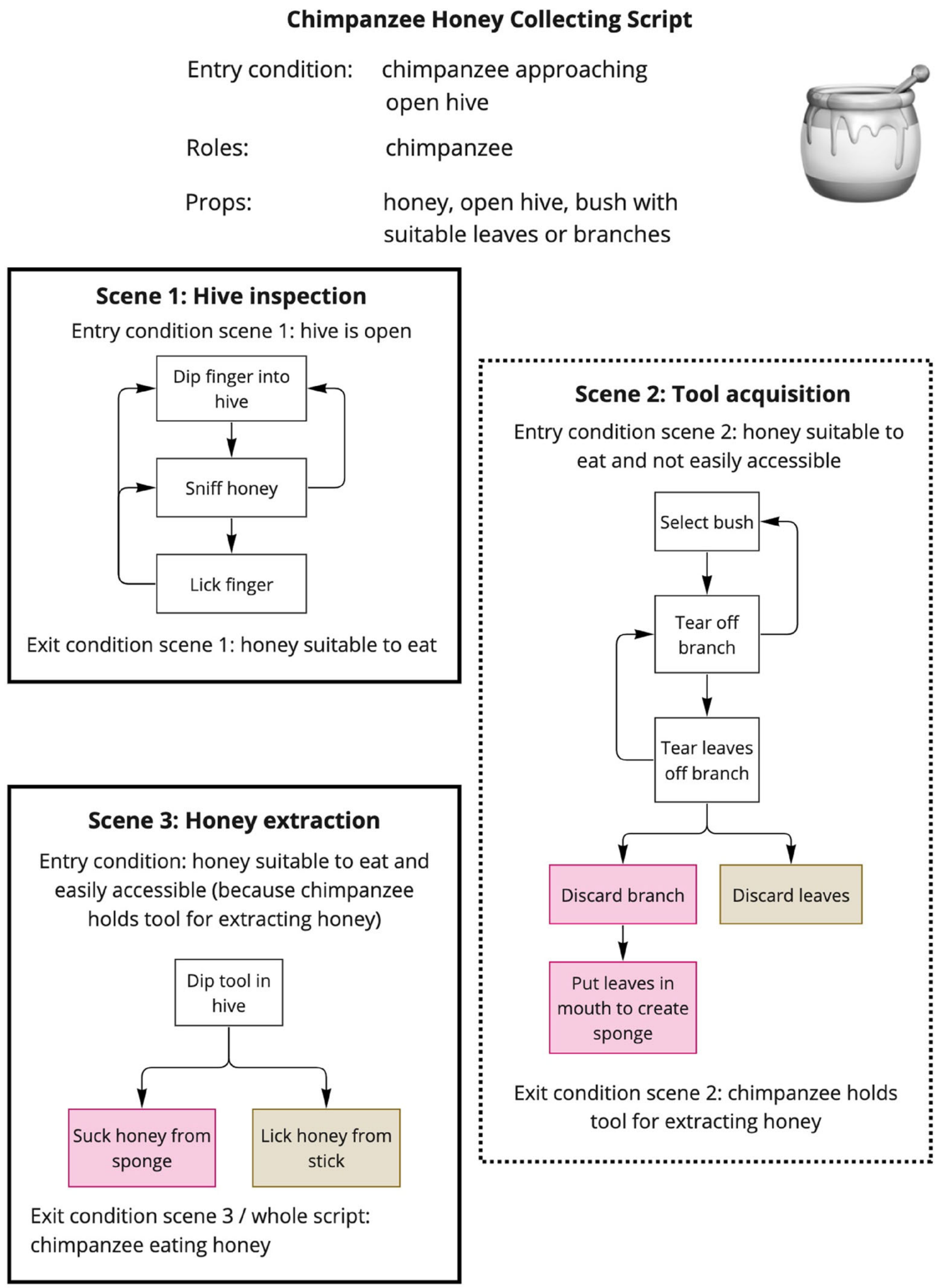


Fig. 2. Hypothesised honey-collecting script of a chimpanzee (*Pan troglodytes schweinfurthii*). The script is comprised of three scenes: hive inspection, tool acquisition (optional, indicated by dotted outline), and honey extraction. Some actions can be repeated as required to obtain the exit condition of the scene. These are indicated by loops to a previous action. For example, in scene 1, the chimpanzee is likely to inspect the honey, but it may do so repeatedly before going on to acquire a tool. Not all possible details are listed, for sake of simplicity. For example, in scene 1, a chimpanzee may initially use its left hand to inspect the honey but repeat the action with its right hand. We leave out such details, since they do not impinge on the chimpanzee acting to obtain the exit condition of the scene (namely, either to acquire a tool or to continue eating the honey). In scenes 2 and 3, actions indicated by pink nodes (on the left) reflect the sponge-dipping strategy of the Sonso chimpanzee community, whereas actions indicated by brown nodes (on the right) reflect the stick-dipping strategy of the Kanyawara chimpanzee community (Gruber *et al.*, 2009, 2011, 2012). Even though both Ugandan chimpanzee communities (of the Budongo Forest and the Kibale National Park, respectively) employ tool-based strategies for collecting honey, they may well employ the same basic honey-collecting script, yet take equifinal actions in order to fashion different tools from the available props. Even though both Sonso and Kanyawara communities fashion tools from the same kind of props, this does not mean that the two chimpanzee communities can easily switch between the two strategies. For instance, Sonso chimpanzees have not been observed deploying a stick-dipping strategy, even if a stick is available and visually salient (Gruber *et al.*, 2011).

et al., 2001). Through experience, they go beyond processing objects merely based on perceptual similarity, and form object categorisations based on *what objects do* – through exploration and observation (Mandler, 2000). That is, objects turn from perceptual units into functional *props*, based on how they function or are used in specific contexts (Mandler, 2000, 2014; Oakes & Madole, 2008). Moreover, infants can distinguish between inanimate objects and animate beings already in their first year (Morton & Johnson, 1991; Opfer & Gelman, 2011; Sanefuji *et al.*, 2014; Setoh *et al.*, 2013). They attribute agency and action, role, perception and affect.

Unlike the beliefs and desires attributed in ToM, these perceptual and conative states are directly manifest in agents' behavioural capacities (Burge, 2018). Nevertheless, they are representationally richer than recognising objects and agents. Very young infants use biological motion as an agency cue, and expect an object that displays agency cues to move itself without external force, through self-propulsion (Bardi, Regnolin & Simion, 2011; Di Giorgio *et al.*, 2017). In the second half of the first year, they also start to track agent–patient relations and to *ascribe roles* to agents. For example, in one study, infants were habituated to a scenario where a disk 'chases' another disk of a different colour, and the roles were reversed in the dishabituation test trials, such that the *chaser* became *chasee* and *vice versa* (Rochat, Striano & Morgan, 2004). Results indicated that 8-to-9-month olds are sensitive to role-reversal, suggesting that infants follow *roles* or '*who did what to whom*'. Infants also attribute sense and affect to other agents. Already in their first months, they are sensitive to gazes and head direction of others, and within their first 6 months, they start to connect the gaze direction and a cued object in that direction, i.e. they use gaze as a cue to look at an object in that direction (see Del Bianco *et al.*, 2019). Furthermore, infants are sensitive to other's perceptual access. Six-month-olds are surprised (look longer) if agents select an object that was not perceptually available to them during the familiarisation trials (Luo & Johnson, 2009). Infants also attribute affect to others, and use these affective cues to learn about the world (Ruba & Repacholi, 2020; Skerry & Spelke, 2014; Sorce *et al.*, 1985; Striano & Vaish, 2006; Walden & Ogan, 1988).

(*b*) *Understanding actions and goals*

Infants and young children not only detect and track objects and agents, but they also identify and track interrelations between agents, objects, and actions. Most of our actions are directed towards objects, and young children may encode the *how*, *where*, and *what* properties while following actions, such as when an agent makes a *grasping* action with a *specific grip* ('how') towards a specific *trajectory* ('where'), to hold *a cup* ('what') (Thompson, Bird & Catmur, 2019; Uithol & Paulus, 2014). When observing familiar and salient target-directed actions, infants display predictive eye movements towards the target, indicating that they recognise the type of action (e.g. grasping) and predict that an agent will follow a specific trajectory or path to take it (e.g. holding a cup). In short, infants can understand various characteristics of others' simple actions, both where/how (how the action is performed and its spatio-temporal properties) and what information guides it (the target of the action).

(*c*) *Segmentation and hierarchisation of event sequences*

Detecting agents, props, actions and goals, and attributing action, perception, and affective states to others are crucial elements to make sense the social world. In addition, script formation also requires segmenting discrete units (scenes), and organising these scenes and lower-level elements within scenes into a hierarchy based on goals and sub-goals. Infants are sensitive to the overall structure of goal-directed events and surprised when this structure is violated; they are able to segment goal-directed actions into meaningful units (Levine *et al.*, 2019). For example, 10–11-month-old infants were presented with some everyday actions executed by an adult (e.g. reaching towards a towel on a surface, picking it up and putting it onto a towel rack). Infants were more surprised if the video is paused in the middle of an action, compared to the pause in between two actions. This suggests that infants can segment dynamic actions into meaningful units based on goals and sub-goals, and recognise event boundaries (Baldwin *et al.*, 2001). Infants are also sensitive to overall goal-directed hierarchical structure in familiar actions (Maffongelli *et al.*, 2019). Based on event-related-potential electroencephalogram (EEG) results, preliminary evidence suggests that 6–7-month-old infants recognise the structural changes in ongoing goal-directed actions, an ability which can be called 'action syntax' (Maffongelli *et al.*, 2018).

(*d*) *Generalisation and transfer*

Generalisation and transfer are crucial components of script understanding and use, because on any given occasion, the specific subjects, props, or actions involved may vary (Hudson *et al.*, 1992; Schank & Abelson, 1977). The ability to transfer allows children to generalise script properties (specifically props and agents with roles) from one occasion to another. For example, in a feeding script for a toddler, the agents (mother, father, etc.), the size, colour, shape, and composition of the props (spoon, fork, baby bottle), the target food, and the context may change. Infants in their first year demonstrate some capacity for such perceptual transfer and generalisation (Bahrick, 2002; Mandler, 2000). For instance, 5.5-month-old infants are more attentive to the type of actions agents perform in dynamic events rather than to objects and faces, indicating that infants are more interested in actions performed by agents independently of changes of objects and faces (Bahrick, Gogate & Ruiz, 2002). This suggests that infants can transfer action information across different agents and props. At the end of their first year, and more so during their second year, infants go beyond generalisation based on perceptual properties, and transfer information based on *functions* (Brown, 1990; Mandler, 2000, 2014;

Quinn *et al.*, 2001). For example, even though the perceptual properties of a spoon change, the functional and affordance properties will perhaps remain the same in a 'feeding script'. Considering that script formation also requires hierarchical organisation, infants should be able to go beyond simple object, agent, or action generalisations; they should also transfer knowledge based on simple hierarchical rules that are not immediately perceptually available, which requires some inference capacities. Recent studies indicate that infants older than 7 months indeed learn and transfer simple hierarchical rules (Lewkowicz, Schmckler & Mangalindan, 2018; Werchan *et al.*, 2015).

(*e*) *Developmental changes*

As indicated above, infants develop abilities to understand objects, agents, actions, segment events, and form simple hierarchical representations based on events from early on. Nevertheless, over the years, children gain more knowledge about the world and develop perceptual and cognitive abilities for categorising props (e.g. food, toys, utensils, etc.), agents (e.g. teacher, family, etc.), and contexts involving props and agents (school, supermarket, etc.) (Gelman & Meyer, 2011; Mandler, 2000, 2014). Even though children can quickly form simple scripts after some familiarisation (e.g. an agent puts a specific object in a green box), children's experience will affect their script understanding. At any given time, a child's script understanding will be dependent on her prior experience, and the degree of the variability of the props, agents, roles, and actions she observes or interacts with during script acquisition. While some scripts include high variability, some others may be more rigid and habitual. For example, the restaurant script has a high degree of variability for props, agents, and context; thus, children may expect variation within this script (Hudson *et al.*, 1992).

(*f*) *From scripts to mind-reading*

It is central to our script account that it does *not rule out* mentalistic forms of social cognition emerging later in both phylogeny and ontogeny. In fact, the computational demands of learning a plethora of scripted action sequences may well require that an observer develops a way of arbitrating when a conspecific agent's actions activate multiple, and potentially competing scripts. For all intents and purposes that capacity can be called mind-reading, although its roots probably lie in the over-proliferation of scripts. Based on obvious differences between humans and great apes concerning the complexity of social life on the one hand (Glock, 2012; Schmidt & Tomasello, 2012) and apparent differences concerning long-term memory capacities on the other (Schwartz & Evans, 2001), we expect that great apes learn fewer scripts than humans. Accordingly, there is little need – evolutionarily speaking – for other great apes to have evolved the complex mind-reading capacities that could resolve a clash between scripts, since they do not have enough scripts stored in memory to experience significant clashes.

Turning to ontogeny, children's scripts get more flexible over the years. Two–three-year-old children can engage in pretend play, which we interpret as creating make-belief scenarios for scripted familiar actions (Harris, 2021). For example, they can use an object as a prop (e.g. a banana) of the 'calling script' and pretend the object is a mobile phone. They can also predict the outcome of pretend actions when they observe pretend play (e.g. if a toy bear pretends to pour milk from an empty carton onto the floor, 2-year-olds can imagine the floor as wet) (Harris *et al.*, 1993). That is, young children can flexibly change the props and agents in scripted scenarios with which they are familiar. This raises the question: why and how do children go beyond scripts and attribute mental states to others? We do not yet have a definitive answer to this question, but believe that various factors are at play.

Children learn many scripts over the years. As a result, they need to develop a way of arbitrating when agent's actions activate multiple, competing scripts. This might require realising or explicitly understanding that their own perspective may diverge from another's perspective, learning mental-state concepts, executive function skills to create meta-representations for competing scripts and going beyond what is perceptually available, and considering both the factual ongoing script and the counter-factual script (counterfactual thinking). It is known in the literature that children's acquiring of mental state concepts, their executive function skills, and their counterfactual thinking are related to their mind-reading performance (Adrian *et al.*, 2005; Guajardo, Parker & Turley-Ames, 2009; Guajardo & Turley-Ames, 2004; Ruffman, Slade & Crowe, 2002). However, understanding when and how children go beyond explaining the world through scripts and explicitly attribute mental states to others requires more in-depth examination. Since our aim herein is to examine what is measured by 'implicit theory of tasks' in human infants and other great apes, we limit our examination to implicit ToM tasks in the next section.

(3) Scripting an implicit ToM task

We now argue that script theory can explain present patterns in the implicit ToM literature (namely, that apes and infants have passed such tests), account for other patterns that implicit ToM cannot explain, and generate new predictions in the implicit ToM paradigm and beyond. We limit our focus to change-of-location paradigms, since only these paradigms have been used with both humans and apes. Implicit ToM tasks with human infants and apes begin with a 'familiarisation' trial, wherein the subject learns the end state of an action within a script – typically retrieving a concealed object (Kano *et al.*, 2019; Onishi & Baillargeon, 2005; Southgate *et al.*, 2007). Such familiarisation can be seen as a process of learning the script of the situation. From a script-theoretic perspective, and taking

Onishi & Baillargeon (2005) as an exemplar, it might be represented as shown in Fig. 3A.

In scene 1, the test subject observes a demonstrator picking up and placing a toy in a green box. In scene 2, the test subject observes the demonstrator reaching towards the green box. This is the script of the demonstrator's behaviour which the test subject learns during the three familiarisation trials. In this script of the implicit ToM familiarisation trials, it is important that a causal chain connecting the demonstrator to the props is established. That is, the demonstrator in scene 2 can retrieve the toy from the green box because the precondition for this in scene 1 (i.e. that the toy is placed in the green box, in view of the demonstrator) has been fulfilled.

The false-belief (FB) experimental conditions are essentially identical to the script learned in the familiarisation trial, but in the FB-yellow condition there is one crucial addition. Unlike the other conditions, in the belief-induction trial of the FB-yellow condition, the demonstrator's visual perspective is occluded, and the toy changes location while it is out of the demonstrator's sight (Fig. 3B). Importantly, the visual perception of the demonstrator would be an essential precondition for retrieving the toy, and this is implicit in the learned script. According to an implicit ToM account, if subjects understand that others can have false beliefs, they should expect the demonstrator to search for the toy in the yellow box, where it was last seen and where the demonstrator believes it to be located.

We submit that script theory offers an alternative interpretation of these findings, following the principle of causal chaining of events (summarised in Fig. 3C). Recall that to predict the behaviour of others using scripts, one need only remember the relationship between who (roles), what (props), as well as where and when (scenes and acts). If these relations change (i.e. the causal chain is disrupted), the script is not expected to proceed. Subsequent scenes in a script can become 'suspended'. By contrast, the observer's script is 'de-activated' if the agent appears no longer to pursue the goal of the script or perform any of its key actions (e.g. if the demonstrator no longer pursues the goal of retrieving the concealed object). In the belief-induction trial FB-yellow condition, the arrangement of props and the visual perspective of the demonstrator in the script has changed in such a way that the entry condition for the second scene is not met, leading to the test subject's script becoming suspended. It is an open empirical question to what extent each of the actions specified in Fig. 3B lead to the observing test subject's script being suspended. However, it is likely that the disruptions introduced by the occluder and the location transfer would be sufficient when taken together to suspend the test subject's script by the end of the belief-induction trial FB-yellow condition.

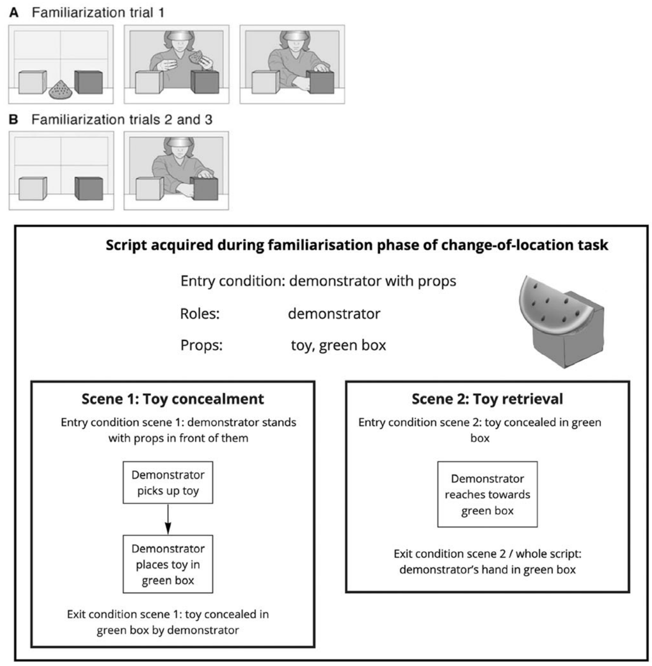
Within this framework, the interpretation of VoE paradigms is simple – subjects show surprise (as measured by looking time) if the agent successfully retrieves the toy in the experimental condition; for this violates the expectations set up by the causal interrelations within the script. The observer's script is 'suspended' because the entry conditions of scene 2 are not met; yet the demonstrator continues to pursue the goal of the script regardless and performs its key action (i.e. retrieving the toy). Therefore, since the demonstrator nonetheless proceeds to scene 2 (i.e. retrieving the toy), an observer using a script to predict the demonstrator's behaviour is expected to show surprise.

It is important to note that the memories formed during the familiarisation trials and retained in a script are first-order representations of objective states of the world, *not second or higher order* representations of other minds which the ToM approach posits to predict behaviour in this situation. In the belief-incongruent FB trial, the location of the toy has changed. Since the demonstrator was not present, the entry conditions are not satisfied, and the observer should not expect the script learned in the familiarisation trials to continue. For according to that script, the toy must be in the location where it was when the demonstrator was last present, for her to retrieve it. Although our focus here is on false belief conditions, our account also explains why greater surprise is observed in the belief-incongruent true belief condition (TB-yellow condition) as well. In this condition, the demonstrator does not continue to the second scene despite satisfying the necessary entry conditions which is unexpected according to the learned script.

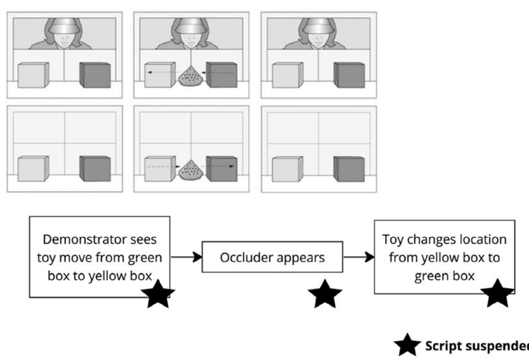
As mentioned in Section I.1, paradigms used to test for mind-reading are beset by replication problems. In many infant studies, difficulties in replicating findings can be due to factors such as caregiver disruptions, fussiness of the infant, and failure to meet familiarisation criterion (Bergmann *et al.*, 2018), while in other cases the reasons may be methodological/theoretical, and some may result from publication bias (Barone *et al.*, 2019; Margoni & Shepperd, 2020). Whilst we do not propose that scripts can account for every aspect of the failed replications, our script account addresses some challenges in replication. In particular, it offers an explanation of why results from anticipatory-looking paradigms have often failed to replicate. If the entry condition is not met because the agent was not present, the observer suspends the script and will not expect its next scene to occur. In consequence, she is neutral between the two options yellow box search and green box search. When the experimenters are tracking the observer's gaze, there is therefore no more chance that she looks in one direction than that she looks in another. The results are therefore distributed between the two boxes, precisely as replications of this type of task seem to indicate (Kulke *et al.*, 2018). This is true of replications within paradigms, such as the anticipatory-looking paradigm. It is also true of replications across paradigms, where, for example, anticipatory-looking measures are found not to correlate with VoE measures (Dörrenberg, Rakoczy & Liszkowski, 2018).

Replication attempts are still on-going (Schuwerk *et al.*, 2021). Nonetheless we conjecture that the interpretation offered here for replication attempts conducted thus far will prove its worth concerning results that emerge from these on-going replication attempts. We surmise that attempted replications using the same methods and sampling

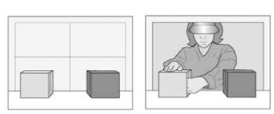
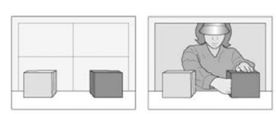
A



B



C

	Yellow box search	**Green box search**
		
	Demonstrator's hand in yellow box	Demonstrator's hand in green box
ToM (Onishi & Baillargeon, 2005)	No violation of expectation, because the demonstrator's behaviour is congruent with her false belief.	Violation of expectation, because the demonstrator's behaviour is incongruent with her false belief.
Behavioural rules (Perner & Ruffman, 2005)	No violation of expectation, because the demonstrator behaves according to the rule 'agents look for objects in the location they last saw them'.	Violation of expectation, because the demonstrator does not behave according to the rule 'agents look for objects in the location they last saw them'.
Script theory	No violation of expectation, because the script is still suspended due to the interruption to both spatio-temporal interrelations of the props and the visual perspective of the demonstrator during the belief-induction trial.	Violation of expectation, because the observer's suspended script is reactivated for two reasons: (*i*) the demonstrator executes the *key action* of scene 2 (reaching towards the green box), even though the script was previously suspended; (*ii*) the demonstrator satisfies the *exit condition* for scene 2/whole script (hand in the green box), even though the script was previously suspended.

(Figure 3 legend continues on next page.)

test subjects with similar backgrounds would be more likely to be replicated, since subjects would more likely acquire the same script during the familiarisation phase. That said, direct replications may still fail, since the previously acquired scripts which subjects bring to bear on the scenario presented in the familiarisation phase of the experiment may still differ slightly according to differences in what matters to different subjects (e.g. the importance of boxes in subjects' previously acquired scripts may differ).

Script theory also offers a more nuanced additional interpretation for the slight preference for the box where the agent last came into contact with the toy in anticipatory-looking studies. This is so because in cases in which there is a preference for that box, the observer still remembers features from a recently suspended script. Subjects may tend to look in the direction of the box where the agent last encountered the object to be retrieved (in Fig. 3, the yellow box), because the alternative (green) box *has not entered* into the causal chaining of events according to the script of the preceding scene. That script requires the agent to have visual access to the actual location of the toy, in order to retrieve it. In the terms set by the above script, from the subject's perspective, the agent cannot proceed to scene 2 (retrieving the toy) in the test trial because the causal chain has been disrupted. On this interpretation, anticipatory looking towards the yellow box tends to occur *not* because subjects anticipate the agent approaching the yellow box; instead, it occurs because the green box has *not featured* in the script according to the principle of causal chaining from the agent's perspective, meaning subjects seldom look towards such areas. That is, the eye-tracking data reflect what subjects are less interested in (i.e. low amounts of looking to props that are causally irrelevant to the script). According to this interpretation, findings from anticipatory-looking studies may have been inconsistently replicated due to a sampling error. In the experimental condition, the subject's looking behaviour is biased against areas that are irrelevant for the agent, such as the green box, meaning more gaze is directed towards more relevant areas such as the yellow box. To be clear, we are referring to a sampling error in the extent to which the data reflect the population from which they are drawn, rather than error in the sampling of the environment on the part of the subject within the experiment.

Importantly, we believe this argument still holds in versions of the change-of-location task (e.g. Southgate *et al.*, 2007) wherein the agent observes the object in both locations before it is ultimately removed in their absence. For our argument is that the expectations are biased towards where the agent last came into contact with the object, as is learned in the familiarisation trial and dictated by the causal-chaining of events learned in the corresponding script, as opposed to other places that the object also occupied.

We further contend that our argument holds in the face of measures taken to control for lower-level explanations of task performance. For example, in Onishi & Baillargeon (2005), the object changes location first in the agent's view, then changes location again when out of view, in order to rule out the possibility that the subject is simply more interested in one location than the other. Another measure that has been taken is that the agent touches both locations, to rule out the possibility that subjects' are simply more attracted to areas touched by the agent. These control measures, however, all take place in the test phase of the experiments. Our argument is that the script learned in the familiarisation phase guides expectations in the test phase, implying that these control measures have not played a role in generating subjects' expectations.

(4) Predictions

Script theory makes several novel experimental predictions in implicit mind-reading tasks. First, the script-theoretic interpretation of anticipatory-looking tasks predicts patterns of gaze behaviour, which may result in more biased looking towards particular locations. For example, in the experimental situation outlined above, if subjects are simply not looking at the green box, we might posit an attentional 'black hole' where the green box is located when the agent returns to retrieve the toy in the experimental trial, meaning the subject's gaze will be more likely to be directed towards the yellow box than the green box. Heat map analyses (Raschke, Blascheck & Burch, 2014) can be used to test this prediction. Importantly, we do not contend that gaze is never predictive of others' actions from the subjects' perspective. In cases where prediction is relevant to the causal chaining of events (perhaps, for example, in the case of socially co-ordinated actions such as dancing), gaze may indeed reflect prediction in order to facilitate co-ordinated action. However, this needs to be considered on a case-by-case basis depending on the goals, roles, and props of a given script and the relationship between them.

(Figure legend continued from previous page.)
Fig. 3. Three phases of scripting the change-of-location task. (A) *Observing infant's script.* This is a hypothesised script acquired by the infant during the *familiarisation* phase of the change-of-location task using the Violation of Expectation paradigm in Onishi & Baillargeon (2005). It is made up of two scenes: toy concealment and toy retrieval. (B) *Belief-induction trial: false-belief (FB)-yellow condition.* Here the demonstrator sees the toy move from the green box to the yellow box. An occluder is introduced, at which point the demonstrator's visual perspective is impaired. During this time the toy moves back to the green box. Black stars indicate factors which, taken together, lead to the observing infant's script being suspended. (C) *Violating expectations in the test trial.* There are two test conditions, one of which involves the same action as in the learned script of the familiarisation trial; namely, demonstrator reaches towards green box. Since the test subject's script is suspended, the test subject's initial expectation that the exit condition for the script as a whole is obtained is violated, as is the test subject's lack of expectation that the central action of scene 2 will be carried out (since the script has been suspended in the belief-induction trial FB-yellow condition). Change-of-location task visuals are adapted from Onishi & Baillargeon (2005) © 2022 Science. Reprinted with permission.

Secondly, script theory makes predictions about subjects' expectations in an extension of the implicit mind-reading scenario that we call a 'second try' condition. Scripts typically do not prescribe literal sequences of behaviour; instead, they outline a nested set of goals to be achieved in the service of an overarching goal (Mandler, 2014). As such, individuals may be required to solve problems using the ontology of the script, and they may try alternative means when necessary (Schank & Abelson, 1977). In a 'second try' trial wherein the agent has the opportunity to make a second choice after failing to locate the concealed object in the first experimental trial, script theory predicts that subjects would expect the agent to try searching in the green box after having searched in the yellow box yet failing to find the toy; for the green box is a part of the ontology of the scene and alternative/equifinal actions are expected to be used to fulfil the script. Still, it is important to note that since script theory emphases the role of experience in predicting behaviour, what happens in the test phase may influence what ultimately happens in the second try condition. By contrast, a ToM account makes no prediction about what would happen in a 'second try' trial. Following a ToM account, agents act according to their beliefs, but the ToM account does not account for how agents understand the causal interrelations between relevant features of the environment; but that would be required to make inferences about likely courses of action in such a circumstance. Thus, according to the subject, the agent no longer has a belief about where the toy is and would therefore not show a systematic preference in a 'second try' trial.

Beyond mind-reading tasks, script theory makes several other experimental predictions about how individuals learn about and remember the past behaviour of others, suggesting its ability to provide a more general account of social cognition. For example, scripts have a clear hierarchical structure with behaviours being chunked together in the service of goals (Mandler, 2014; Schank & Abelson, 1977). Consequently, script theory predicts that hierarchically organised sequences of goal-directed actions should be more easily remembered than non-hierarchically organised sequences (Byrne & Russon, 1998). As well as predicting what will be remembered, script theory predicts what will *not* be remembered. Changes in aspects of scenes that are not related to either the goals or ontology of the scene should be less likely to be remembered than changes in aspects of scenes that are relevant (Schank & Abelson, 1977). Relatedly, since scripts specify what is relevant within a scene, script theory is also consistent with robust psychological phenomena such as 'inattentional blindness', wherein subjects fail to recognise unusual features of scenes due to attentional biases which are prescribed by the goal of the situation (or the current script) (Mack, 2003).

Future empirical and modelling work on script acquisition and updating in non-human animals could draw on and modify work already conducted on the cognitive mechanisms underlying stereotypes in humans. There is evidence concerning different models of stereotype acquisition and updating, which are potentially informative for script theory: the bookkeeping model (i.e. current scripts are constantly updated to accommodate mildly script-discrepant information), the conversion model (i.e. current scripts are drastically updated to accommodate strongly script-discrepant information), and the subtyping model (i.e. new scripts are generated to accommodate schema-discrepant information) (Bicchieri & McNally, 2018; Crocker, Fiske & Taylor, 1984). The different models can be tested by comparing how subjects respond to familiar scenarios, after exposure to few *versus* many script-violating scenarios and after exposure to script-violating scenarios where few *versus* many aspects of the script have been violated. A study conducted by Weber & Crocker (1983) compared these different models in human adults, and found evidence for the bookkeeping and subtyping model, but not for the conversion model. Notably, they suggest that the conversion model may be applicable in cases where the subject is unfamiliar with a situation, and, in our terms, has not yet acquired a script for the event. In principle, similar studies could be conducted with non-human animals for cross-species comparison.

Future work on scripts can also compare which kinds of cues trigger different scripts across species and age groups, to gain a better understanding of how specific the entry condition needs to be for a specific script to be activated. Finally, the role of individual experience is important for script theory to address, and there is an opportunity to study the impact of individual experience in non-human animals in a ToM task. In a study with apes using a version of the goggles experiment (whereby the demonstrator wears a pair of goggles which the test subject has experience as being either opaque or translucent), Kano *et al.* (2019) showed that performance in this context depended on self-experience with the goggle apparatus. This is in line with our account, which emphasises the role of experience in understanding how the world works and setting up expectations for how scenarios will unfold, but follow-up experiments could test whether the impact of this kind of experience is manifested by great apes in non-ToM tasks as well. A critical step for future empirical research for contrasting script explanations of cognitive performance and mentalising explanations is to attend to the ways that familiarisation phases in social cognition tasks activate or themselves generate expectations which subjects use to structure their expectations about how events will unfold.

III. CONCLUSIONS

(1) We have argued that script theory avoids several shortcomings of currently available ToM accounts and provides a general de-intellectualised framework of social cognition, one that avoids the imputation of higher-order representations.

(2) Recognising that the organising principle of behavioural acts (e.g. to move, to put) and scenes (e.g. to hide a toy) is the agent's goal and that these goals are pursued in particular ways, allows humans and other cognitively advanced animals

to make predictions about the future (e.g. the agent will return to her hiding location), without the necessity to ascribe a false belief.

(3) Appreciating that others pursue goals that lead to predictable outcomes may be all that is cognitively required for what looks like an understanding of false belief.

(4) We argue that by categorising acts and scenes into scripts, humans and other animals may be able to interpret and predict actions using basic psychological skills.

(5) For these reasons, we argue that script theory is a promising framework capable of providing valuable new insights into the question of how individuals make sense of others.

IV. ACKNOWLEDGEMENTS

Open access funding provided by Universite de Neuchatel.

V. REFERENCES

Abelson, R. P. (1981). Psychological status of the script concept. *American Psychologist* **36**(7), 715–729.

Adrian, J. E., Clemente, R. A., Villanueva, L. & Rieffe, C. (2005). Parent–child picture-book reading, mothers' mental state language and children's theory of mind. *Journal of Child Language* **32**(3), 673–686.

Albarracin, M., Constant, A., Friston, K. J. & Maxwell, J. D. R. (2021). A variational approach to scripts. *Frontiers in Psychology* **12**, 585493.

Andrews, K. (2009). Understanding norms without a theory of mind. *Inquiry* **52**(5), 433–448.

Apperly, I. A. (2010). *Mindreaders: The Cognitive Basis of 'Theory of Mind'*. Psychology Press, Hove and New York.

Apperly, I. A. (2012). What is 'theory of mind'? Concepts, cognitive processes and individual differences. *Quarterly Journal of Experimental Psychology* **65**(5), 825–839.

Apperly, I. A. & Butterfill, S. A. (2009). Do humans have two systems to track beliefs and belief-like states? *Psychological Review* **116**(4), 953–970.

Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C. & Samson, D. (2006). Is belief reasoning automatic? *Psychological Science* **17**(10), 841–844.

Bahrick, L. E. (2002). Generalization of learning in three-and-a-half-month-old infants on the basis of amodal relations. *Child Development* **73**(3), 667–681.

Bahrick, L. E., Gogate, L. J. & Ruiz, I. (2002). Attention and memory for faces and actions in infancy: the salience of actions over faces in dynamic events. *Child Development* **73**(6), 1629–1643.

Baillargeon, R., Buttelmann, D. & Southgate, V. (2018). Invited commentary: interpreting failed replications of early false-belief findings: methodological and theoretical considerations. *Cognitive Development* **46**, 112–124.

Baldwin, D. A., Baird, J. A., Saylor, M. M. & Clark, M. A. (2001). Infants parse dynamic action. *Child Development* **72**(3), 708–717.

Bardi, L., Regolin, L. & Simion, F. (2011). Biological motion preference in humans at birth: role of dynamic and configural properties. *Developmental Science* **14**(2), 353–359.

Bar-On, D. (2013). Origins of meaning: must we 'go Gricean'? *Mind & Language* **28**(3), 342–375.

Baron-Cohen, S. (1995). *Mindblindness: An Essay on Autism and Theory of Mind*. MIT Press, Cambridge, MA.

Baron-Cohen, S., Leslie, A. M. & Frith, U. (1985). Does the autistic child have a 'theory of mind' ? *Cognition* **21**(1), 37–46.

Barone, P., Corradi, G. & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: a review and meta-analysis. *Infant Behavior and Development* **57**, 101350.

Bartlett, F. C. (1932). *Remembering: A Study in Experimental and Social Psychology*, p. 317. Cambridge University Press, Cambridge.

Bergmann, C., Tsuji, S., Piccinini, P. E., Lewis, M. L., Braginsky, M., Frank, M. C. & Cristia, A. (2018). Promoting replicability in developmental research through meta-analyses: insights from language acquisition research. *Child Development* **89**(6), 1996–2009.

Bicchieri, C. & McNally, P. (2018). Shrieking sirens: schemata, scripts, and social norms. How change occurs. *Social Philosophy and Policy* **35**(1), 23–53.

Breveglieri, R., Vaccari, F. E., Bosco, A., Gamberini, M., Fattori, P. & Galletti, C. (2019). Neurons modulated by action execution and observation in the macaque medial parietal cortex. *Current Biology* **29**(7), 1218–1225.

Brown, A. L. (1990). Domain-specific principles affect learning and transfer in children. *Cognitive Science* **14**(1), 107–133.

Burge, T. (2018). Do infants and nonhuman animals attribute mental states? *Psychological Review* **125**, 409–434.

Burnside, K., Severdija, V. & Poulin-Dubois, D. (2020). Infants attribute false beliefs to a toy crane. *Developmental Science* **23**(2), e12887.

Butterfill, S. A. & Apperly, I. A. (2013). How to construct a minimal theory of mind. *Mind & Language* **28**(5), 606–637.

Byrne, R. W. & Russon, A. E. (1998). Learning by imitation: a hierarchical approach. *Behavioral and Brain Sciences* **21**(5), 667–721.

Call, J. & Santos, L. R. (2012). Understanding other minds. In *The Evolution of Primate Societies* (eds J. C. Mitani and J. Call), pp. 664–681. University of Chicago Press, Chicago.

Call, J. & Tomasello, M. (1999). A nonverbal false belief task: the performance of children and great apes. *Child Development* **70**(2), 381–395.

Carruthers, P. & Smith, P. K. (eds) (1996). *Theories of Theories of Mind*. Cambridge University Press, Cambridge.

Charafeddine, R., Mercier, H., Clément, F., Kaufmann, L., Berchtold, A., Reboul, A. & der Henst, J.-B. V. (2015). How preschoolers use cues of dominance to make sense of their social environment. *Journal of Cognition and Development* **16**(4), 587–607.

Clayton, N. S. & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature* **395**(6699), 272–274.

Clément, F., Bernard, S. & Kaufmann, L. (2011). Social cognition is not reducible to theory of mind: when children use deontic rules to predict the behaviour of others. *British Journal of Developmental Psychology* **29**(4), 910–928.

Clément, F. & Koenig, M. (2018). Epistemology: knowledge in childhood. In *The Routledge Handbook of the Philosophy of Childhood and Children* (eds A. Gheaus, G. Calder and J. De Wispelaere), pp. 13–22. Routledge, New York.

Cole, G. G., Smith, D. T. & Atkinson, M. A. (2015). Mental state attribution and the gaze cueing effect. *Attention, Perception & Psychophysics* **77**(4), 1105–1115.

Corradi-Dell'Acqua, C., Turri, F., Kaufmann, L., Clément, F. & Schwartz, S. (2015). How the brain predicts people's behavior in relation to rules and desires. Evidence of a medio-prefrontal dissociation. *Cortex* **70**, 21–34.

Crocker, J., Fiske, S. T. & Taylor, S. E. (1984). Schematic bases of belief change. In *Attitudinal Judgment* (ed. J. R. Eiser), pp. 197–226. Springer, New York.

Del Bianco, T., Falck-Ytter, T., Thorup, E. & Gredebäck, G. (2019). The developmental origins of gaze-following in human infants. *Infancy* **24**(3), 433–454.

Deshpande, A., Van Boekholt, B. & Zuberbuhler, K. (2022). Preliminary evidence for one-trial social learning of vervet monkey alarm calling. *Royal Society Open Science* **9**(8), 210560.

Di Giorgio, E., Lunghi, M., Simion, F. & Vallortigara, G. (2017). Visual cues of motion that trigger animacy perception at birth: the case of self-propulsion. *Developmental Science* **20**(4), e12394.

Doherty, M. J. & Perner, J. (2020). Mental files: developmental integration of dual naming and theory of mind. *Developmental Review* **56**, 100909.

Dörrenberg, S., Rakoczy, H. & Liszkowski, U. (2018). How (not) to measure infant theory of mind: testing the replicability and validity of four non-verbal measures. *Cognitive Development* **46**, 12–30.

Elsner, B. & Adam, M. (2021). Infants' goal prediction for simple action events: the role of experience and agency cues. *Topics in Cognitive Science* **13**(1), 45–62.

Erickson, L. C., Thiessen, E. D. & Graf Estes, K. (2014). Statistically coherent labels facilitate categorization in 8-month-olds. *Journal of Memory and Language* **72**, 49–58.

Estes, K. G., Evans, J. L., Alibali, M. W. & Saffran, J. R. (2007). Can infants map meaning to newly segmented words?: Statistical segmentation and word learning. *Psychological Science* **18**(3), 254–260.

Farag, C., Troiani, V., Bonner, M., Powers, C., Avants, B., Gee, J. & Grossman, M. (2010). Hierarchical organization of scripts: converging evidence from fMRI and frontotemporal degeneration. *Cerebral Cortex* **20**(10), 2453–2463.

Fiebich, A., Gallagher, S. & Hutto, D. D. (2016). Pluralism, interaction, and the ontogeny of social cognition. In *The Routledge Handbook of Philosophy of the Social Mind*. Routledge, London.

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological Review* **99**, 689–723.

Flombaum, J. I. & Santos, L. R. (2005). Rhesus monkeys attribute perceptions to others. *Current Biology* **15**(5), 447–452.

Franklin, N. T., Norman, K. A., Ranganath, C., Zacks, J. M. & Gershman, S. J. (2020). Structured event memory: a neuro-symbolic model of event cognition. *Psychological Review* **127**(3), 327–361.

Gallagher, S. (2020). *Action and Interaction*. Oxford University Press, Oxford.

Gallagher, S. & Hutto, D. D. (2008). Understanding others through primary interaction and narrative practice. In *The Shared Mind: Perspectives on Intersubjectivity*, pp. 17–38. John Benjamins Publishing Company, Amsterdam.

Gallagher, S. & Povinelli, D. J. (2012). Enactive and behavioral abstraction accounts of social understanding in chimpanzees, infants, and adults. *Review of Philosophy and Psychology* **3**(1), 145–169.
Gallese, V. & Sinigaglia, C. (2012). Response to de Bruin and Gallagher: embodied simulation as reuse is a productive explanation of a basic form of mind-reading. *Trends in Cognitive Sciences* **16**(2), 99–100.
Gallese, V. & Sinigaglia, C. (2018). Embodied resonance. In *The Oxford Handbook of 4E Cognition* (eds A. Newen, L. De Bruin and S. Gallagher), pp. 417–432. Oxford University Press, Oxford, UK.
Gelman, S. A. & Meyer, M. (2011). Child categorization. *WIREs Cognitive Science* **2**(1), 95–105.
Gergely, G. & Csibra, G. (2003). Teleological reasoning in infancy: the naïve theory of rational action. *Trends in Cognitive Sciences* **7**(7), 287–292.
Gershman, S. J. & Daw, N. D. (2017). Reinforcement learning and episodic memory in humans and animals: an integrative framework. *Annual Review of Psychology* **68**, 101–128.
Gioia, D. A. & Poole, P. P. (1984). Scripts in organizational behavior. *Academy of Management Review* **9**(3), 449–459.
Glock, H.-J. (2012). The anthropological difference. In *Human Nature*, Royal Institute of Philosophy Supplement: 70 (eds C. Sandis and M. J. Cain). Cambridge University Press, Cambridge.
Goldman, A. I. (1992). In defense of the simulation theory. *Mind & Language* **7**, 104–119.
Goodall, J. (1986). Dominance. In *The Chimpanzees of Gombe: Patterns of Behavior*, pp. 409–422. The Belknap Press of Harvard University Press, Cambridge.
Gopnik, A. & Wellman, H. M. (1992). Why the child's theory of mind really is a theory. *Mind & Language* **7**(1–2), 145–171.
Grice, H. P. (1957). Meaning. *The Philosophical Review* **66**(3), 377–388.
Grosse Wiesmann, C., Friederici, A. D., Singer, T. & Steinbeis, N. (2017). Implicit and explicit false belief development in preschool children. *Developmental Science* **20**(5), e12445.
Gruber, T., Muller, M. N., Reynolds, V., Wrangham, R. & Zuberbühler, K. (2011). Community-specific evaluation of tool affordances in wild chimpanzees. *Scientific Reports* **1**(1), 128.
Gruber, T., Muller, M. N., Strimling, P., Wrangham, R. & Zuberbühler, K. (2009). Wild chimpanzees rely on cultural knowledge to solve an experimental honey acquisition task. *Current Biology* **19**(21), 1806–1810.
Gruber, T., Potts, K. B., Krupenye, C., Byrne, M.-R., Mackworth-Young, C., McGrew, W. C., Reynolds, V. & Zuberbühler, K. (2012). The influence of ecology on chimpanzee (*Pan troglodytes*) cultural behavior: a case study of five Ugandan chimpanzee communities. *Journal of Comparative Psychology* **126**(4), 446–457.
Guajardo, N. R., Parker, J. & Turley-Ames, K. (2009). Associations among false belief understanding, counterfactual reasoning, and executive function. *British Journal of Developmental Psychology* **27**(3), 681–702.
Guajardo, N. R. & Turley-Ames, K. J. (2004). Preschoolers' generation of different types of counterfactual statements and theory of mind understanding. *Cognitive Development* **19**(1), 53–80.
Harris, P. L. (2021). Early constraints on the imagination: the realism of young children. *Child Development* **92**(2), 466–483.
Harris, P. L., Kavanaugh, R. D., Wellman, H. M. & Hickling, A. K. (1993). Young children's understanding of pretense. *Monographs of the Society for Research in Child Development* **58**(1), 1–107.
Hay, J. F., Pelucchi, B., Estes, K. G. & Saffran, J. R. (2011). Linking sounds to meanings: infant statistical learning in a natural language. *Cognitive Psychology* **63**(2), 93–106.
Hayashi, T., Akikawa, R., Kawasaki, K., Egawa, J., Minamimoto, T., Kobayashi, K., Kato, S., Hori, Y., Nagai, Y., Iijima, A., Someya, T. & Hasegawa, I. (2020). Macaques exhibit implicit gaze bias anticipating others' false-belief-driven actions via medial prefrontal cortex. *Cell Reports* **30**(13), 4433–4444.e5.
He, Z., Bolz, M. & Baillargeon, R. (2011). False-belief understanding in 2.5-year-olds: evidence from violation-of-expectation change-of-location and unexpected-contents tasks. *Developmental Science* **14**(2), 292–305.
Heyes, C. (2014). Submentalizing: I am not really reading your mind. *Perspectives on Psychological Science* **9**(2), 131–143.
Heyes, C. (2017). Apes submentalise. *Trends in Cognitive Sciences* **21**(1), 1–2.
Horschler, D. J., MacLean, E. L. & Santos, L. R. (2020). Do non-human primates really represent others' beliefs? *Trends in Cognitive Sciences* **24**(8), 594–605.
Horschler, D. J., Santos, L. R. & MacLean, E. L. (2021). How do non-human primates represent others' awareness of where objects are hidden? *Cognition* **212**, 104658.
Hudson, J. A., Fivush, R. & Kuebli, J. (1992). Scripts and episodes: the development of event memory. *Applied Cognitive Psychology* **6**(6), 483–505.
Kaminski, J., Call, J. & Tomasello, M. (2008). Chimpanzees know what others know, but not what they believe. *Cognition* **109**(2), 224–234.
Kampis, D., Frances, B. & Kovács, Á. M. (2020). Developing a theory of mind: are infants sensitive to how other people represent the world? In *The Social Brain: A Developmental Perspective* (ed. J. Decety), pp. 143–160. MIT Press, Cambridge.
Kano, F. & Call, J. (2014). Great apes generate goal-based action predictions: an eye-tracking study. *Psychological Science* **25**(9), 1691–1698.
Kano, F., Call, J. & Krupenye, C. (2020). Primates pass dynamically social anticipatory-looking false-belief tests. *Trends in Cognitive Sciences* **24**(10), 777–778.
Kano, F. & Hirata, S. (2015). Great apes make anticipatory looks based on long-term memory of single events. *Current Biology* **25**(19), 2513–2517.
Kano, F., Krupenye, C., Hirata, S., Call, J. & Tomasello, M. (2017). Submentalizing cannot explain belief-based action anticipation in apes. *Trends in Cognitive Sciences* **21**, 633–634.
Kano, F., Krupenye, C., Hirata, S., Tomonaga, M. & Call, J. (2019). Great apes use self-experience to anticipate an agent's action in a false-belief test. *Proceedings of the National Academy of Sciences* **116**(42), 20904–20909.
Kaufmann, L. & Clément, F. (2014). Wired for society: cognizing pathways to society and culture. *Topoi* **33**(2), 459–475.
Kaufmann, L. & Fabrice, C. (2003). Is sociology core knowledge? The social nature of the cognitive architecture. *Intellectica* **36–37**, 421–457.
Keysar, B., Lin, S. & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition* **89**(1), 25–41.
Kirkham, N. Z., Slemmer, J. A. & Johnson, S. P. (2002). Visual statistical learning in infancy: evidence for a domain general learning mechanism. *Cognition* **83**(2), B35–B42.
Korman, J. & Malle, B. F. (2016). Grasping for traits or reasons? How people grapple with puzzling social behaviors. *Personality and Social Psychology Bulletin* **42**(11), 1451–1465.
Kosie, J. E. & Baldwin, D. A. (2021). Dwell times showcase how goal structure informs preschoolers' analysis of unfolding motion patterns. *Child Development* **92**(6), 2235–2243.
Krachun, C., Carpenter, M., Call, J. & Tomasello, M. (2009). A competitive nonverbal false belief task for children and apes. *Developmental Science* **12**(4), 521–535.
Krupenye, C. & Call, J. (2019). Theory of mind in animals: current and future directions. *WIREs Cognitive Science* **10**(6), e1503.
Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science* **354**(6308), 110–114.
Krupenye, C., Kano, F., Hirata, S., Call, J. & Tomasello, M. (2017). A test of the submentalizing hypothesis: apes' performance in a false belief task inanimate control. *Communicative & Integrative Biology* **10**(4), e1343771.
Kulke, L. & Hinrichs, M. A. B. (2021). Implicit theory of mind under realistic social circumstances measured with mobile eye-tracking. *Scientific Reports* **11**(1), 1215.
Kulke, L., von Duhn, B., Schneider, D. & Rakoczy, H. (2018). Is implicit theory of mind a real and robust phenomenon? Results from a systematic replication study. *Psychological Science* **29**(6), 888–900.
Laland, K. & Seed, A. (2021). Understanding human cognitive uniqueness. *Annual Review of Psychology* **72**(1), 689–716.
Lee, S. W., O'Doherty, J. P. & Shimojo, S. (2015). Neural computations mediating one-shot learning in the human brain. *PLoS Biology* **13**(4), e1002137.
León, J., Thiriau, C., Bodin, C., Crockford, C. & Zuberbühler, K. (2022). Acquisition of predator knowledge from alarm calls via one-trial social learning in monkeys. *IScience* **25**(9), 104853.
Levine, D., Buchsbaum, D., Hirsh-Pasek, K. & Golinkoff, R. M. (2019). Finding events in a continuous world: a developmental account. *Developmental Psychobiology* **61**(3), 376–389.
Lewis, A., Berntsen, D. & Call, J. (2019). Long-term memory of past events in great apes. *Current Directions in Psychological Science* **28**(2), 117–123.
Lewkowicz, D. J., Schmuckler, M. A. & Mangalindan, D. M. J. (2018). Learning of hierarchical serial patterns emerges in infancy. *Developmental Psychobiology* **60**(3), 243–255.
Liddle, B. & Nettle, D. (2007). Higher-order theory of mind and social competence in school-age children. *Journal of Cultural and Evolutionary Psychology* **4**(3–4), 231–244.
Lieder, F. & Griffiths, T. L. (2020). Resource-rational analysis: understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences* **43**, e1.
Liszkowski, U. (2018). Origins and complexities of infant communication and social cognition. In *The Oxford Handbook of 4E Cognition* (eds A. Newen, L. De Bruin and S. Gallagher), pp. 661–684. Oxford University Press, Oxford, UK.
Lohmann, H. & Tomasello, M. (2003). The role of language in the development of false belief understanding: a training study. *Child Development* **74**(4), 1130–1144.
Low, J., Apperly, I. A., Butterfill, S. A. & Rakoczy, H. (2016). Cognitive architecture of belief reasoning in children and adults: a primer on the two-systems account. *Child Development Perspectives* **10**(3), 184–189.
Low, J. & Edwards, K. (2018). The curious case of adults' interpretations of violation-of-expectation false belief scenarios. *Cognitive Development* **46**, 86–96.

Low, J. & Perner, J. (2012). Implicit and explicit theory of mind: state of the art. *British Journal of Developmental Psychology* **30**(1), 1–13.

Luo, Y. & Johnson, S. C. (2009). Recognizing the role of perception in action at 6 months. *Developmental Science* **12**(1), 142–149.

Mack, A. (2003). Inattentional blindness: looking without seeing. *Current Directions in Psychological Science* **12**(5), 180–184.

Maffongelli, L., Antognini, K. & Daum, M. M. (2018). Syntactical regularities of action sequences in the infant brain: when structure matters. *Developmental Science* **21**(6), e12682.

Maffongelli, L., D'Ausilio, A., Fadiga, L. & Daum, M. M. (2019). The ontogenesis of action syntax. *Collabra: Psychology* **5**(1), 21.

Mandler, J. M. (2000). Perceptual and conceptual processes in infancy. *Journal of Cognition and Development* **1**(1), 3–36.

Mandler, J. M. (2014). *Stories, Scripts, and Scenes: Aspects of Schema Theory*. Psychology Press, New York and Hove.

Margoni, F. & Shepperd, M. (2020). Changing the logic of replication: a case from infant studies. *Infant Behavior and Development* **61**, 101483.

Marotta, A., Lupiáñez, J., Martella, D. & Casagrande, M. (2012). Eye gaze versus arrows as spatial cues: two qualitatively different modes of attentional selection. *Journal of Experimental Psychology: Human Perception and Performance* **38**, 326–335.

Marotta, A., Román-Caballero, R. & Lupiáñez, J. (2018). Arrows don't look at you: qualitatively different attentional mechanisms triggered by gaze and arrows. *Psychonomic Bulletin & Review* **25**(6), 2254–2259.

Marticorena, D. C. W., Ruiz, A. M., Mukerji, C., Goddu, A. & Santos, L. R. (2011). Monkeys represent others' knowledge but not their beliefs. *Developmental Science* **14**(6), 1406–1416.

Martin, A. & Santos, L. R. (2014). The origins of belief representation: monkeys fail to automatically represent others' beliefs. *Cognition* **130**(3), 300–308.

Martin, A. & Santos, L. R. (2016). What cognitive representations support primate theory of mind? *Trends in Cognitive Sciences* **20**(5), 375–382.

Mascaro, O. & Csibra, G. (2012). Representation of stable social dominance relations by human infants. *Proceedings of the National Academy of Sciences* **109**(18), 6862–6867.

McVee, M. B., Dunsmore, K. & Gavelek, J. R. (2005). Schema theory revisited. *Review of Educational Research* **75**(4), 531–566.

Mesoudi, A. & Whiten, A. (2004). The hierarchical transformation of event knowledge in human cultural transmission. *Journal of Cognition and Culture* **4**(1), 1–24.

Meunier, H. (2017). Do monkeys have a theory of mind? How to answer the question? *Neuroscience & Biobehavioral Reviews* **82**, 110–123.

Moore, R. (2017). Gricean communication and cognitive development. *The Philosophical Quarterly* **67**(267), 303–326.

Morton, J. & Johnson, M. H. (1991). CONSPEC and CONLERN: a two-process theory of infant face recognition. *Psychological Review* **98**, 164–181.

Nanay, B. (2020). Vicarious representation: a new theory of social cognition. *Cognition* **205**, 104451.

Nelson, K. & Gruendel, J. M. (1986). Children's scripts. In *Event Knowledge: Structure and Function in Development* (ed. K. Nelson), pp. 21–46. Erlbaum, Hillsdale, NJ.

Newen, A., Welpinghus, A. & Juckel, G. (2015). Emotion recognition as pattern recognition: the relevance of perception. *Mind & Language* **30**(2), 187–208.

Newen, A. & Wolf, J. (2020). The situational mental file account of the false belief tasks: a new solution of the paradox of false belief understanding. *Review of Philosophy and Psychology* **11**(4), 717–744.

Nichols, S. & Stich, S. P. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*, p. 237. Clarendon Press/Oxford University Press, Oxford.

Oakes, L. M. & Madole, K. L. (2008). Function revisited: how infants construe functional features in their representation of objects. *Advances in Child Development and Behavior* **36**, 135–185.

Okamoto-Barth, S., Call, J. & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science* **18**(5), 462–468.

Onishi, K. H. & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science* **308**(5719), 255–258.

Opfer, J. E. & Gelman, S. A. (2011). Development of the animate-inanimate distinction. In *The Wiley-Blackwell Handbook of Childhood Cognitive Development*, Second Edition, pp. 213–238. Wiley Blackwell.

Orbell, S. & Verplanken, B. (2018). Progress and prospects in habit research. In *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts* (ed. B. Verplanken), pp. 397–409. Springer International Publishing, Cham.

Origgi, G. & Sperber, D. (2000). Evolution, communication and the proper function of language. In *Evolution and the Human Mind: Modularity, Language and Meta-Cognition* (eds A. Chamberlain and P. Carruthers), pp. 140–169. Cambridge University Press, Cambridge.

Pace, A., Carver, L. J. & Friend, M. (2013). Event-related potentials to intact and disrupted actions in children and adults. *Journal of Experimental Child Psychology* **116**(2), 453–470.

Paulus, M., Schuwerk, T., Sodian, B. & Ganglmayer, K. (2017). Children's and adults' use of verbal information to visually anticipate others' actions: a study on explicit and implicit social-cognitive processing. *Cognition* **160**, 145–152.

Perner, J. (2010). Who took the cog out of cognitive science? Mentalism in an era of anti-cognitivism. In *Cognition and Neuropsychology*, First Edition (Volume **1**, eds P. A. Frensch and R. Schwarzer). Psychology Press, Hove, UK.

Perner, J., Priewasser, B. & Roessler, J. (2018). The practical other: teleology and its development. *Interdisciplinary Science Reviews* **43**(2), 99–114.

Perner, J. & Roessler, J. (2012). From infants' to children's appreciation of belief. *Trends in Cognitive Sciences* **16**(10), 519–525.

Perner, J. & Ruffman, T. (2005). Infants' insight into the mind: how deep? *Science* **308**(5719), 214–216.

Phillips, B. (2021). Seeing seeing. *Philosophy and Phenomenological Research* **102**(1), 24–43.

Phillips, J., Buckwalter, W., Cushman, F., Friedman, O., Martin, A., Turri, J., Santos, L. & Knobe, J. (2021). Knowledge before belief. *Behavioral and Brain Sciences* **44**(e140), 1–75.

Povinelli, D. & Vonk, J. (2006). We don't need a microscope to explore the chimpanzee's mind. In *Rational Animals?* (eds S. Hurley and M. Nudds), pp. 1–28. Oxford University Press, Oxford.

Powell, L. J., Hobbs, K., Bardis, A., Carey, S. & Saxe, R. (2018). Replications of implicit theory of mind tasks with varying representational demands. *Cognitive Development* **46**, 40–50.

Premack, D. & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences* **1**(4), 515–526.

Pun, A., Birch, S. A. J. & Baron, A. S. (2017). Foundations of reasoning about social dominance. *Child Development Perspectives* **11**(3), 155–160.

Quesque, F. & Rossetti, Y. (2020). What do theory-of-mind tasks actually measure? Theory and practice. *Perspectives on Psychological Science* **15**(2), 384–396.

Quinn, P. C., Slater, A. M., Brown, E. & Hayes, R. A. (2001). Developmental change in form categorization in early infancy. *British Journal of Developmental Psychology* **19**(2), 207–218.

Raby, C. R. & Clayton, N. S. (2009). Prospective cognition in animals. *Behavioural Processes* **80**(3), 314–324.

Radvansky, G. A. & Zacks, J. M. (2014). *Event Cognition*, p. 272. Oxford University Press, New York.

Raschke, M., Blascheck, T. & Burch, M. (2014). Visual analysis of eye tracking data. In *Handbook of Human Centric Visualization* (ed. W. Huang), pp. 391–409. Springer, New York.

Recanati, F. (2012). *Mental Files*. Oxford University Press, Oxford.

Rochat, P., Striano, T. & Morgan, R. (2004). Who is doing what to whom? Young Infants' developing sense of social causality in animated displays. *Perception* **33**(3), 355–369.

Roediger, H. L. & Arnold, K. M. (2012). The one-trial learning controversy and its aftermath: remembering Rock (1957). *The American Journal of Psychology* **125**(2), 127–143.

Rosa Salva, O., Mayer, U. & Vallortigara, G. (2015). Roots of a social brain: developmental models of emerging animacy-detection mechanisms. *Neuroscience & Biobehavioral Reviews* **50**, 150–168.

Rosati, A. G., Santos, L. R. & Hare, B. (2010). Primate social cognition: thirty years after Premack and Woodruff. In *Primate Neuroethology*, pp. 117–143. Oxford University Press, New York.

Ruba, A. L. & Repacholi, B. M. (2020). Do preverbal infants understand discrete facial expressions of emotion? *Emotion Review* **12**(4), 235–250.

Ruffman, T. (2014). To belief or not belief: children's theory of mind. *Developmental Review* **34**(3), 265–293.

Ruffman, T., Slade, L. & Crowe, E. (2002). The relation between children's and mothers' mental state language and theory-of-mind understanding. *Child Development* **73**(3), 734–751.

Saffran, J. R., Aslin, R. N. & Newport, E. L. (1996). Statistical learning by 8-month-old infants. *Science* **274**(5294), 1926–1928.

Saffran, J. R. & Kirkham, N. Z. (2018). Infant statistical learning. *Annual Review of Psychology* **69**, 181–203.

Sanefuji, W., Wada, K., Yamamoto, T., Mohri, I. & Taniike, M. (2014). Development of preference for conspecific faces in human infants. *Developmental Psychology* **50**, 979–985.

Schank, R. C. & Abelson, R. P. (1977). *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*. Psychology Press, New Jersey.

Schmidt, M. F. H. & Tomasello, M. (2012). Young children enforce social norms. *Current Directions in Psychological Science* **21**(4), 232–236.

Schneider, D., Bayliss, A. P., Becker, S. I. & Dux, P. E. (2012). Eye movements reveal sustained implicit processing of others' mental states. *Journal of Experimental Psychology: General* **141**(3), 433–438.

Schurz, M., Radua, J., Tholen, M. G., Maliske, L., Margulies, D. S., Mars, R. B., Sallet, J. & Kanske, P. (2021). Toward a hierarchical model of social cognition: a neuroimaging meta-analysis and integrative review of empathy and theory of mind. *Psychological Bulletin* **147**(3), 293–327.

Schuwerk, T., Kampis, D., Baillargeon, R., Biro, S., Bohn, M., Byers-Heinlein, K., Dörrenberg, S., Fisher, C., Franchin, L., Fulcher, T., Garbisch, I., Geraci, A., Wiesmann, C. G., Hamlin, K., Haun, D. B. M., et al. (2021). Action anticipation based on an agent's epistemic state in toddlers and adults. Electronic file available at https://doi.org/10.31234/osf.io/x4jbm Accessed 14.02.2021.

Schuwerk, T., Priewasser, B., Sodian, B. & Perner, J. (2018). The robustness and generalizability of findings on spontaneous false belief sensitivity: a replication attempt. *Royal Society Open Science* **5**(5), 172273.

Schwartz, B. L. & Evans, S. (2001). Episodic memory in primates. *American Journal of Primatology* **55**(2), 71–85.

Scott-Phillips, T. (2015). *Speaking Our Minds: Why Human Communication Is Different, and How Language Evolved to Make it Special*. Bloomsbury Academic, London.

Setoh, P., Scott, R. M. & Baillargeon, R. (2016). Two-and-a-half-year-olds succeed at a traditional false-belief task with reduced processing demands. *Proceedings of the National Academy of Sciences* **113**(47), 13360–13365.

Setoh, P., Wu, D., Baillargeon, R. & Gelman, R. (2013). Young infants have biological expectations about animals. *Proceedings of the National Academy of Sciences* **110**(40), 15937–15942.

Skerry, A. E. & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition* **130**(2), 204–216.

Slaughter, V. (2015). Theory of mind in infants and young children: a review. *Australian Psychologist* **50**(3), 169–172.

Smith, L. & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition* **106**(3), 1558–1568.

Song, H. & Baillargeon, R. (2008). Infants' reasoning about others' false perceptions. *Developmental Psychology* **44**, 1789–1795.

Sorce, J. F., Emde, R. N., Campos, J. J. & Klinnert, M. D. (1985). Maternal emotional signaling: its effect on the visual cliff behavior of 1-year-olds. *Developmental Psychology* **21**, 195–200.

Southgate, V., Senju, A. & Csibra, G. (2007). Action anticipation through attribution of false belief by 2-year-olds. *Psychological Science* **18**(7), 587–592.

Spelke, E. S. & Kinzler, K. D. (2007). Core knowledge. *Developmental Science* **10**(1), 89–96.

Sperber, D. (2000). *Metarepresentations: A Multidisciplinary Perspective*. Oxford University Press USA, New York.

Stawarczyk, D., Bezdek, M. A. & Zacks, J. M. (2021). Event representations and predictive processing: the role of the midline default network core. *Topics in Cognitive Science* **13**(1), 164–186.

Striano, T. & Vaish, A. (2006). Seven- to 9-month-old infants use facial expressions to interpret others' actions. *British Journal of Developmental Psychology* **24**(4), 753–760.

Surian, L. & Franchin, L. (2020). On the domain specificity of the mechanisms underpinning spontaneous anticipatory looks in false-belief tasks. *Developmental Science* **23**(6), e12955.

Teufel, C., Alexis, D. M., Todd, H., Lawrance-Owen, A. J., Clayton, N. S. & Davis, G. (2009). Social cognition modulates the sensory coding of observed gaze direction. *Current Biology* **19**(15), 1274–1277.

Thompson, E. L., Bird, G. & Catmur, C. (2019). Conceptualizing and testing action understanding. *Neuroscience & Biobehavioral Reviews* **105**, 106–114.

Thomsen, L. (2020). The developmental origins of social hierarchy: how infants and young children mentally represent and respond to power and status. *Current Opinion in Psychology* **33**, 201–208.

Thomsen, L., Frankenhuis, W. E., Ingold-Smith, M. & Carey, S. (2011). Big and mighty: preverbal infants mentally represent social dominance. *Science* **331**(6016), 477–480.

Tinbergen, N. (1951). *The Study of Instinct*, p. 237. Clarendon Press/Oxford University Press, New York.

Tomasello, M. (2008). *Origins of Human Communication*, p. 393. MIT Press, United States.

Uithol, S. & Paulus, M. (2014). What do infants understand of others' action? A theoretical account of early social cognition. *Psychological Research* **78**(5), 609–622.

Uithol, S., van Rooij, I., Bekkering, H. & Haselager, P. (2012). Hierarchies in action and motor control. *Journal of Cognitive Neuroscience* **24**(5), 1077–1086.

van der Vaart, E. & Hemelrijk, C. K. (2014). 'Theory of mind' in animals: ways to make progress. *Synthese* **191**(3), 335–354.

Walden, T. A. & Ogan, T. A. (1988). The development of social referencing. *Child Development* **59**(5), 1230–1240.

Weber, R. & Crocker, J. (1983). Cognitive processes in the revision of stereotypic beliefs. *Journal of Personality and Social Psychology* **45**, 961–977.

Wellman, H. M. (2014). *Making Minds: How Theory of Mind Develops*. Oxford University Press, New York.

Wellman, H. M. (2017). The development of theory of mind: historical reflections. *Child Development Perspectives* **11**(3), 207–214.

Wellman, H. M. (2018). Theory of mind: the state of the art. *European Journal of Developmental Psychology* **15**(6), 728–755. https://doi.org/10.1080/17405629.2018.1435413.

Werchan, D. M., Collins, A. G. E., Frank, M. J. & Amso, D. (2015). 8-month-old infants spontaneously learn and generalize hierarchical rules. *Psychological Science* **26**(6), 805–815.

Whiten, A. (2013). Humans are not alone in computing how others see the world. *Animal Behaviour* **86**(2), 213–221.

Wilson, V. A. D., Zuberbühler, K. & Bickel, B. (2022). The evolutionary origins of syntax: event cognition in nonhuman primates. *Science Advances* **8**(25), eabn8464.

Wimmer, H. & Perner, J. (1983). Beliefs about beliefs: representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition* **13**(1), 103–128.

Worden, R. P. (1996). Primate social intelligence. *Cognitive Science* **20**(4), 579–616.

Yamamoto, S., Humle, T. & Tanaka, M. (2012). Chimpanzees' flexible targeted helping based on an understanding of conspecifics' goals. *Proceedings of the National Academy of Sciences* **109**(9), 3588–3592.

Zacks, J. M. (2020). Event perception and memory. *Annual Review of Psychology* **71**(1), 165–191.

(*Received 31 May 2022; revised 14 June 2023; accepted 19 June 2023*)