The psychological foundations of reputation-based cooperation

Héctor M. Manrique[1], Henriette Zeidler[2], Gilbert Roberts[3], Pat Barclay[4], Michael Walker[5], Flóra Samu[6], Andrea Fariña [7], Redouan Bshary[8], Nichola Raihani[9]

1 Universidad de Zaragoza, Spain.
2 Aston University, UK.
3 Newcastle, UK.
4 University of Guelph, Canada.
5 Universidad de Murcia, Spain.
6 Linköping University, Sweden.
7 Leiden University, NL.
8 University of Neuchâtel, Switzerland.
9 University College London, UK.

Abstract

Humans care about having a positive reputation, which may prompt them to help in scenarios where the return benefits are not obvious. Various game-theoretical models support the hypothesis that concern for reputation may stabilize cooperation beyond kin, pairs or small groups. However, such models are not explicit about the underlying psychological mechanisms that support reputation-based cooperation. These models therefore cannot account for the apparent rarity of reputation-based cooperation in other species. Here we identify the cognitive mechanisms that may support reputation-based cooperation in the absence of language. We argue that a large working memory enhances the ability to delay gratification, to understand others' mental states (which allows for perspective-taking and attribution of intentions), and to create and follow norms, which are key building blocks for increasingly complex reputation-based cooperation. We review the existing evidence for the appearance of these processes during human ontogeny as well as their presence in non-human apes and other vertebrates. Based on this review, we predict that most non-human species are cognitively constrained to show only simple forms of reputation-based cooperation.

Keywords: attributing intentions, cognition, delay of gratification, memory, normativity, perspective-taking, reputation-based cooperation, theory of mind.

1. Introduction

Concern for reputation is a key psychological mechanism for explaining the high levels of cooperation observed in humans. Obtaining a good reputation could lead to downstream benefits via one of two routes: individuals might be more likely to be chosen as a partner (reputation-based partner choice, Roberts, 1998) or they might be more likely to be rewarded by other individuals ('indirect reciprocity', Kandori, 1992; Ohtsuki & Iwasa, 2007; see Roberts et al. this issue for a detailed discussion and comparison). Despite the intensive focus on how cooperation can be theoretically promoted by concern for reputation, these theoretical models have tended to 'black-box' the psychology that underpins decision rules. In this review, we aim to highlight the psychological and cognitive mechanisms that might support reputation-based cooperation in humans. We begin by discussing the ontogeny of reputation-based cooperation in humans, and the cognitive mechanisms that likely underpin the ability to evaluate and manage reputation. We argue that the requirement for these mechanisms might largely preclude the emergence of reputation-based cooperation in other species. We end by presenting a few examples where reputation-based cooperation in non-human species appears to exist, illustrating how reputation-based cooperation might sometimes be achieved by simpler cognitive means.

2. Reputation-based cooperation in humans and other primates

Reputation-based cooperation relies on two distinct capacities: individuals must be able to evaluate the reputations of others as well as be able to strategically manage their own reputation. The cognition underpinning these two facets of reputation-based cooperation is likely to differ (Figure 1). Some evidence suggests that children begin to evaluate others on the basis of their prosociality from a very young age (reviewed in Van de Vondervoort & Hamlin, 2008 but see Salvadori et al., 2015 for failed replication efforts). Evidence also exists in non-human apes and other primates to suggest that individuals are able to evaluate and choose interaction partners on the basis of observed prosociality (Herrmann et al., 2013, Russell et al., 2008, Subiaul et al., 2008, Kawai et al., 2019, but see Bueno-Guerra et al., 2020).

In addition to evaluating others, humans also strategically manage their reputation by behaving more cooperatively when there is a possibility that other individuals will learn about their actions (see meta-analysis by Bradley et al., 2018). Observability increases cooperation in many domains, including tax compliance (Coricelli et al., 2010); voter turnout (Gerber et al., 2008); energy conservation (Yoeli et al., 2013); environmentalism (Barclay & Barker, 2020); blood donation (Lacetera & Macis, 2010); and more. Most researchers interpret this increased cooperation as being caused by people's concern for reputation.

However, unlike the ability to evaluate others' reputation, this tendency to strategically manage one's own reputation is not present at all stages of life and instead appears to emerge during development. Although young children (under two years old) are known to behave prosocially (Dunfield et al., 2011; Warneken & Tomasello, 2007; Vaish et al., 2009), such behaviour appears to stem from an intrinsic motivation to satisfy a partner's needs rather from attempts to strategically manage reputation. Children begin to show a concern for reputation from the age of around five, for example by refraining from stealing from others if they are observed, or making more generous or fairer donations to recipients when their generosity will be revealed to others (Grueneisen & Tomasello, 2017; Leimgruber et al., 2012, McAuliffe et al., 2020). Other work has shown that a concern with appearing to be prosocial or fair-minded increases over childhood (Shaw et al. 2014), and that children become especially concerned with self-presentation between the ages of 8 to 11 years old (Aloise-Young, 1993). At this age, children are increasingly able to inhibit behaviours that might result in social sanctions (Apfelbaum et al., 2008; Rutland et al., 2005) and attempt to present themselves in a positive light to others. At the same time, children become increasingly skeptical about the intentions of others, particularly when it comes to judging prosocial reputations (Heyman et al. 2014). Thus, it takes most of childhood for humans to hone their ability to understand how one's actions affect our reputations and to behave strategically so as to curate a positive reputation.

Unlike humans, there is scant evidence that non-humans primates attempt to strategically manage their reputation. One recent study found that capuchin monkeys were insensitive to the presence of an observer when deciding whether to share food (Schino et al., 2021), suggesting that capuchins do not attempt to strategically manage their reputation in this way. Studies in chimpanzees have also yielded null results. For instance, although chimpanzees increase effort in a resource acquisition task when watched by a potential competitor, they do not increase effort when watched by a potential cooperation partner (Engelmann et al., 2016). In the same task, four to-five-year-old children increased their efforts both in the presence of a competitive observer and in the presence of a potential future cooperation partner (Engelmann et al., 2016). Similarly, although five-year old children share more and steal less when observed by a peer, chimpanzees are not sensitive to the presence of an observer in the same paradigm (Engelmann, et al., 2012, see also Leimgruber et al., 2012, see also Nettle et al., 2013).

The findings above suggest that (1) cognitive strategies needed for reputation-based cooperation differ depending on whether we consider evaluation of partners versus managing one's own reputation and (2) that managing one's own reputation is likely to depend upon more sophisticated socio-cognitive mechanisms. In what follows, we present four socio-cognitive candidates that may frequently be involved in reputation-based cooperation. Most fundamentally, we propose that an extensive working memory is key to developing the sophisticated forms of reputation management seen in humans. Three additional socio-cognitive abilities derive from working memory that are likely to be involved in reputation-based cooperation. These abilities are: (i) delaying gratification (ii) understanding others' mental states; and (iii) following and enforcing social norms. We show how these building blocks recruit working memory and how they may impinge upon reputation-based cooperation – as well as distinguishing between the cognition needed for evaluating others' reputations and managing one's own reputation, respectively (Figure 1).

3. Cognitive mechanisms supporting reputation-based cooperation

3.1. Working memory

Following Fuster (2001), we define working memory as "a mechanism of temporal integration". Crucially, working memory is not synonymous with short-term memory (STM) but rather emphasises both the reactivation of long-term stored information and the integration of new inputs, both of which are likely to be involved in dynamically evaluating and managing reputation. Working memory can be metaphorically likened to a workstation, a place where information is temporarily held and manipulated. Working memory is engaged whenever sophisticated socio-cognitive calculations are needed, such as appreciating that our own perspectives, beliefs and intentions can differ from those of other individuals, and understanding that an individual's intentions might not be accurately represented by his actions.

The ability to successfully manage one's own reputation might often require individuals to monitor how they appear to others. Such monitoring requires the ability to entertain multiple perspectives simultaneously, which makes burdensome demands of working memory

(Manrique & Walker, 2017). Successfully managing one's own reputation might also involve mental time travel, which allows individuals to imagine how events might unfold in the future. This ability is also likely to involve working memory (Dere et al., 2019). Working memory is also likely to be involved in evaluating the reputations of others, for example by tracking cooperative behaviours (Milinski & Wedekind, 1998) and recalling what happened, with whom and when ('episodic memory'). The complexity of such tasks can be increased further when individuals compare observed behaviours against normative standards, or against behaviours adopted by other individuals. The all-round utility of working memory poses some intriguing questions for developmental and evolutionary psychology: at what age does children's working memory become capable of maintaining reputation-based cooperative systems? Do great apes have working memory complex enough to sustain reputation-based cooperative systems? By what processes might these abilities have evolved in humans?

Working memory increases linearly between ages ~7 months and 14 years (Diamond & Doar, 1989; Gathercole et al., 2004, Read, 2008). Meta-analytic evidence (Read, 2008) suggests that 6-year-olds have a working memory size of three (compared to seven in adults: Miller, 1956). Three is the minimum working memory size required to command relative clauses in sentences, which are complex recursive structures like those used to tracking other people's perspectives (e.g. John thinks that Mary knows he is supportive). Given that many reputational acts require such recursion (e.g., John knows that if he doesn't help Mary now, she will not trust him to reciprocate), it is reasonable to regard three as the minimum working memory size required for constructing complex reputation-based cooperative systems. The extent of working memory involvement in evaluation of others' reputations is likely to depend: evaluations that don't involve recursion (e.g., helping that signals physical ability) may need less working memory than those which do (e.g., helping that signals future intent to cooperate).

Studies directly measuring working memory in great apes are few and have yielded mixed results. Some studies suggest that the working memory capacity of non-human apes is likely to be limited. For instance, in a simplified version of the Wisconsin Card Sorting Test, that involves sorting cards along three dimensions (shape, colour, number), chimpanzees struggled to form a classificatory criterion or to change it flexibly to match the reinforcement contingencies (Moriguchi et al., 2011). Similarly, in a memory task where individuals had to turn over cards one at a time and find matching pairs, chimpanzees made four times more mistakes than humans when tasked with three pairs, which would involve holding three cards in working memory (Washburn et al., 2007). Nevertheless, other studies have reported remarkable performance in serial ordering tasks administered to chimpanzees, that involved memorizing up to 5 digits flashed on a screen in ascending order (Inoue & Matsuzawa, 2007), or presenting up to 6 closed boxes on a platform and having a subject chimpanzee encode and remember those boxes already emptied of food in previous trials to avoid re-opening them again (Völter et al., 2019).

An alternative approach to assessing working memory capacity involves measuring the extent to which individuals are able to hierarchically classify objects (Langer, 1980, 1986, 2000). The Langer protocol investigates spontaneous grouping of objects and allows performance to be rated as a function of complexity, ranging from first-order classifications, where only a single group of objects matching in shape and/or colour is formed (e.g. is set apart from the other objects), to classifications in which more than one group is formed contemporaneously (e.g. rings are grouped together and kept apart from the cubes). Second- (and higher) order classifications are assigned to groups of objects that are perceptually different, yet share the same classificatory criteria. Second- (and higher) order classification impose higher working memory demands on the classificatory rule as well as on the elements to be sorted, as their differing features need to be compared simultaneously and flexibly (Langer, 1980, 1986, 2000). Chimpanzees attain second-order combinativity around age 5 (Potì et al., 1999; Spinozzi et al., 1999) when still they rarely compose more than two sets at a time (Langer, 2000, p.225). In contrast, toddlers begin developing three-category classifications around age 3. Three-category classification allows children to hierarchize – such as two subordinate classes within one superordinate class – whereas two-category classification does not (Langer, 2000). This hierarchization indicates that children develop recursive structures that might help them tracking other people's perspectives and construct social reputation-based cooperative systems.

Other approaches have inferred working memory size based on the increasing complexity of manufactured stone tools in the fossil record. Making and using simple stone flakes is reported from Late Pliocene Africa 3,4 MYA, where bipedal Australopithecine existed from before 4 MYA. Australopithecines gave rise to the genus Homo, perhaps as early as 2,8 MYA, with which they coexisted until after 2 MYA. By 2,5 MYA there are several Palaeolithic assemblages of sharp conchoidal (i.e., shell-shaped) flakes struck by manual percussion with hard hammer-stones. Conchoidal fracturing requires simultaneously focusing on the core stone, the hammer stone, and the percussion angle, which implies a larger working memory than that required for simple flakes (Read & Van Der Leeuw, 2008). Homo predominated by 1,76 MYA years ago, and co-occur in the African archaeological record with flattish stone handaxes. These handaxes often resembled a large almond, were formed by manual percussion with a hard hammer-stone that removed small conchoidal flakes in a regular manner (e.g. bifacial stone-tool fashioning), from two surfaces of the handaxe to be. By 0.4-0.3 MYA, handaxes had 3D symmetry, which required their makers to simultaneously remember different perspectives of the core being worked on. To achieve ideal symmetry involves advanced foresight and the ability to represent mentally the intended final product to exert on-going corrections on the working substrate. Based on the increasing complexity of stone tools, and the working memory required to make them, a reasonable conjecture is that early Homo had a working memory greater than that of Australopithecines, which was in turn that of chimpanzees. Taken together, these various lines of evidence suggest that working memory capacity is likely to be higher in humans than in non-human apes (and specifically chimpanzees).

Although working memory capacity has been relatively understudied in other animals (Carruthers, 2013), there is some suggestive evidence for correlates of advanced working memory in some species. For example, scrub jays display evidence of episodic-like memory, being able to remember 'what', 'when' and 'where' during food caching events (Clayton & Dickinson, 1998) as well as flexibly altering their own caching strategies to avoid being parasitized by others (Correia et al., 2007). This example might provide the most compelling evidence for sophisticated working memory in non- primates. As such, if they would benefit from being able to choose partners for cooperative interactions, then they are a good species to test for reputation-based cooperative systems.

3.2 Delay of Gratification

Any form of costly cooperation based on investments requires the ability to resist the temptation to obtain immediate benefits (e.g. by cheating) in order to pursue a larger benefit in the future. In some cases, this problem may be solved by psychological mechanisms which render cooperative behaviour immediately subjectively rewarding (a phenomenon known as warm glow, Andreoni, 1990). In other cases, individuals may have to effortfully resist an immediately higher-paying option: they must be able to delay gratification.

Although people are systematically present-biased, the human ability to think long-term is extraordinary in nature (Roberts, 2002; Suddendorf, 2013). Human consciousness can produce mental simulations of possible futures, allowing decisions to be based on anticipated outcomes (Baumeister et al., 2018). Indeed, a large part of humans' mental processes seems to be prospective (Seligman et al., 2013), focusing on what ought to be done in the here and now in order to produce positive results in the future (Schacter, Addis & Buckner, 2007).

Investing in a prosocial reputation might sometimes require the ability to delay gratification, because the rewards for cooperation come from future (potentially unknown) partners instead of one's current partner and are therefore inherently more likely to be delayed and less certain to materialise. Several lines of evidence link the ability to delay gratification with cooperative tendency in humans. Focusing on the future makes participants more generous (Sjåstad, 2019), and spurs their willingness to incur personal costs to prevent damaging reputational information from spreading (Vonasch et al., 2018). Children's ability to delay gratification is positively related to their tendency to share, indicating that the ability to delay gratification might be a prerequisite for children's sharing and cooperation (Sebastián-Enesco & Warneken, 2015). Similar patterns have been observed in adults (Curry et al., 2008; Harris & Madden, 2002; though see Barclay & Barker, 2020; Wu et al., 2017), as well as in blue jays who are prevented from consuming rewards immediately (Stephens et al., 2002). Children are also better at delaying gratification in cooperative tasks than solo tasks (Koomen et al., 2020). A direct link between delay of gratification and reputational management has been suggested in 3- and 4-year-old children (Ma et al., 2020), although other work has shown that people are unable to anticipate the delayed indirect benefits from their own cooperative investments (Wu et al., 2016). To the extent that delay of gratification is involved in reputation-based cooperation, we expect it to be more important in reputation management than in evaluating the reputations of others (see Fig. 1).

In humans, the ability to delay gratification is measured using paradigms such as the 'marshmallow test' (Mischel & Ebbesen, 1970), which measures the willingness to forego a smaller, immediate reward when a larger, delayed reward is promised. Performance on such tasks is variable - and the strategies children use to resist temptation suggest the importance of two different cognitive systems ("automatic" vs "top-down") that affect self-control (Luerssen et al., 2015; Hare et al., 2009). By the age of six, children become aware that putting the rewards out of sight during the delay interval helps them to withhold and wait longer (Mischel

& Mischel, 1983). By the age of 12, children realise that not only seeing the food influences their performance, but also the way they talk about it – demonstrating the role of metacognition on performance in such settings. Qualitatively similar results have been observed in chimpanzees. In experimental settings, chimpanzees can delay gratification for up to 10 minutes (Beran & Evans, 2006), and seem to use similar strategies to human children to increase performance on these tasks. For example, chimpanzees engage in more play when higher self-restraint is needed in order to gain bigger rewards - suggesting that they are intentionally deploying strategies to increase their performance (Evans & Beran, 2007).

The delay-of-gratification test has by now been used on a variety of vertebrate species (Miller et al., 2019; Susini et al., 2020; Aellen et al., 2021) with varying results. Dogs (with their owners) as well as some fish and large-brained monkeys (macaques and capuchins) are all able to wait for extended periods to obtain larger rewards; cuttlefish have also been reported to wait up to two minutes (Schnell et al., 2021). By contrast, small monkeys, rats and various birds (pigeons, corvids, parrots) perform poorly in such tasks. Nevertheless, apart from dogs and chimpanzees, individuals of high performing species typically only wait 30-60 seconds for a larger amount or a preferred food, which offers a stark contrast with the circa 30 minutes reported in human children (Luerssen et al., 2015) in similar tasks – and the potential to delay gratification for much longer periods in adulthood. This reduced delay of gratification in other species may limit their ability to perform reputation-based cooperation.

3.3 Theory of Mind

Theory of mind is a multifaceted concept that refers to the ability to attribute mental states to oneself and to third-parties and encompasses different abilities, which vary in computational complexity. For example, taking another individual's visual perspective is simpler than attributing intentions, which is in turn simpler than attributing knowledge, which is again simpler than understanding complex perspectives (level 2 perspective-taking) or attributing beliefs. These latter two examples of theory of mind are extremely taxing in terms of computational demands, because they involve entertaining simultaneously alternative, often contradictory, representations of reality (for a more detailed explanation, see Manrique & Walker, 2017).

Here we introduce two theory of mind abilities that are likely to be involved in reputation management and evaluating the reputation of others: perspective-taking and attribution of intentions. Reputation-based cooperation may be more stable against erosion if bystanders or other third parties can correctly attribute intentions and beliefs to actors, and if actors can represent how they and their actions are perceived in the eyes of others. For example, an individual may fail to cooperate either because (s)he does not realise that a recipient needs help, or because (s)he currently lacks the resources to help. In other words, individuals with a willingness to help may sometimes behave uncooperatively. If bystanders can correctly identify uncooperative behaviour as a mistake or temporary inability, they can continue a cooperative relationship with those who didn't intend to defect. Therefore, the reputation system becomes less prone to errors undermining cooperation.

Errors are particularly problematic in indirect reciprocity models of cooperation. Indirect reciprocity is only stable if agents distinguish between justified defections and unjustified defections (i.e., defecting on defectors vs. defecting on cooperators; "Kandori" or "standing" strategies, Kandori, 1992; Ohtsuki & Iwasa, 2007). However, such systems are undermined by errors because they can cause two individuals to perceive the same situation differently. Under the Kandori strategy, an actor's reputation improves if (s)he either helps a partner in good standing or refuses to help a partner in bad standing. Conversely, an actor's reputation decreases if (s)he fails to help someone in good standing or helps someone in bad standing. Thus, if actors and bystanders evaluate a potential recipient's reputation differently, bystanders will alter the actor's reputation score in the opposite direction as the actor (or others) would have expected. Under the Kandori strategy, low frequencies of any type of error may therefore erode cooperation (Milinski et al., 2001). Perspective taking (and more broadly theory of mind) are crucial to overcome the limitations of Kandori, as players may acknowledge the possibility of missing information leading to the 'wrong' behaviour or the 'wrong' interpretation.

By contrast, reputation-based partner choice can function with or without theory of mind. In reputation-based partner choice, actors help others to signal their ability and/or willingness to help (Barclay, 2013). Theory of mind is not necessary to signal one's abilities or to interpret such signals: when people see a good hunter share his kill, they can infer that (s)he is physically skilled enough to catch it (e.g., Smith & Bliege Bird, 2000) without knowing anything of his or her mental state. Hunters needn't know anything about the audience's mental state either – they can learn that certain behaviours are rewarded (e.g., being chosen as a partner) via reinforcement learning. However, theory of mind can greatly aid reputation-based partner choice because it allows for more complex or targeted signals. For example, theory of mind allows audiences to infer a helper's intentions in order to predict future cooperation and thus allows individuals to signal not just their ability but their willingness to help. Therefore, although simple forms of reputation-based partner choice might be achieved without the advanced socio-cognitive mechanisms we discuss in this paper, we note that reputation-based partner choice can later evolve to become cognitively quite complex, particularly when helpful individuals have an incentive to misrepresent their type to others and when receivers take hidden intentions of partners into consideration when evaluating prosocial acts (see Raihani & Power, 2021 for a detailed discussion).

3.3.1 Perspective-taking

Perspective-taking can be broadly described as the ability to adopt the perspective of others (e.g. visual, informational, emotional). At around two years of age, children are able to differentiate what people can or cannot see (Moll & Tomasello, 2006). However, it is usually not until three to four years of age that children understand that the same item can look different from different perspectives (Moll & Meltzoff, 2011). This ability (level 2 perspective-taking) requires effortful control to suppress the child's own visual perception, and is often viewed as the precursor to full-blown theory of mind, in which the individual gains the ability to understand others' knowledge and beliefs.

Perspective-taking is likely to be involved in both reputation management and the evaluation of others' reputations. Reputation management involves not only behaving in a certain way, but also the ability to shift perspectives to represent how complying or failing to act in this manner will be perceived by others (Fig. 1). Thus, taking others' perspectives can make an organism much more effective at reputation management. Similarly, perspective-taking makes an organism better at detecting cheaters: organisms may dishonestly present themselves as cooperative, and it requires cognitive effort for observers to distinguish between genuine versus deceptive cooperators. For example, one individual might normally be a "cheater", but might temporarily act cooperatively when (s)he sees someone (s)he wants to deceive or impress (e.g., a potential mate). Detecting dishonesty involves being able to entertain simultaneously differing views of reality, an ability that can be equated in terms of computational complexity to attributing complex (level 2) visual perspective. Hence, even if perspective-taking is not strictly required to evaluate other's reputation, managing level 2 visual perspective-taking indicates that organisms have the cognitive potential to entertain simultaneously differing/contrasting views of reality (mine vs yours), and hence the ability to representing simultaneously overt and hidden intentions in other's actions.

Perspective taking covers a wide spectrum of abilities, from knowing what others can or cannot see ('level 1') to understanding that others see something differently as a function of their relative position ('level 2') (Flavell, 1977; Flavell et al., 1981) and is therefore a good proxy of other mentalising skills. Level 1 perspective-taking has been extensively investigated in chimpanzees with initially diverging results (Povinelli & Eddy, 1996; Hare et al., 2000). Karg et al. (2015) used a variation of the experience projection paradigm (Heyes, 1998) where chimpanzees were trained with different pairs of goggles that affected what they could see. When wearing one colour, the apes could see through the goggles but when wearing the other colour they could not see anything. It could be inferred that chimpanzees are able to shift perspectives if their own experience with the goggles (i.e., seeing vs. not seeing) affected their response to human experimenters wearing the goggles. However, in this study, chimpanzees' gaze-following was not influenced by their own previous experience with the googles (Karg et al. 2015). Subsequent results indicated that chimpanzees may be able to shift perspectives in a competitive context yet correct visual perspective attribution only approached a modest 60% (Karg et al. 2015; but see Okamoto-Barth et al., 2007 for more positive findings). Demonstrating level 2 visual perspective-taking in chimpanzees still proves elusive (Karg et al., 2016).

Outside apes, the basic forms of perspective taking have currently only been found in large-brained species. For example, rhesus monkeys steal more often from a human competitor whose face is hidden by an opaque barrier than a competitor whose body alone is hidden (Flombaum & Santos, 2005). Capuchin monkeys can also strategically conceal visual information (Flombaum & Santos, 2005), while macaques have been reported to know what others can or cannot hear (Santos et al., 2006). Ravens provide the best evidence for perspective taking in birds, being able to follow human gaze direction around obstacles (Bugnyar et al., 2004) and attributing visual perspectives even to unseen competitors (Bugnyar et al., 2016). Most recently, however, there is evidence that cleaner fish *Labroides dimidiatus* females are able to choose foraging sites where their male partners cannot observe them (McAuliffe et al., in review). Altogether, it appears that some other species may have some perspective-taking abilities which can aid reputation-based cooperation, but perhaps not to the same level as humans.

3.3.2 Attributing intentions

Having a good or bad reputation is not simply the consequence of performing good or bad deeds; the intention behind observed actions matters (although the tendency to take intentions into consideration when forming moral judgements varies across cultures, Barrett et al., 2016). Notwithstanding this cross-cultural variability, attributing intentionality is another skill that is key to evaluating third-party reputations (Fig.1).

As early as 14 months, infants selectively copy actions performed intentionally, as opposed to those that seem fortuitous (Meltzoff, 1995). Similarly, Gergely et al. (2002) showed that 14-month-old children imitate unusual actions (e.g., turning on a light with one's forehead) more often if those actions were voluntary than if the actions were necessary (e.g., the model's hands were full, thus necessitating use of their forehead). Nine to eighteen months-old toddlers show more patience towards adults who try but fail to hand them a toy than towards teasing adults (i.e., seem unwilling) (Behne et al., 2005). Similarly, 21-month-old children are more willing to help other children who had attempted but failed to hand them a toy in previous interactions, than to those who previously refused to offer the toy (Dunfield & Kuhlmeier, 2010). Therefore, it appears children at a very early age can differentiate outcomes from intentions when judging others' behaviour.

Other animals also appear capable of attributing intentions. In one study (Call & Tomasello, 1998), chimpanzees and orangutans preferentially selected boxes that were deliberately marked as containing rewards, more so than boxes that were accidentally marked by the experimenter. Similar attempts at gauging intention attribution in other nonhuman primates have met with mixed results: positive in cotton-top tamarins and rhesus macaques (Wood et al., 2007); negative in chimpanzees (Povinelli et al., 1998), Tonkean macaques and tufted capuchin monkeys (Costes-Thiré et al., 2015). Call et al. (2004) showed that chimpanzees leave a testing area sooner when confronted with an experimenter who was unwilling to give them food (e.g. a teasing human who took away the food) as opposed to one who was unable to do so. This paradigm has yielded similar results in capuchins and Tonkean macaques (Canteloup et al., 2016; Phillips et al., 2009). Some non-primates also seem able to consider both the intentions and the outcomes of performed actions: grey parrots (Péron et al., 2010) and even horses (Trösch et al., 2020) behave differently when confronted with an unwilling versus an unable experimenter offering food rewards.

Some intentions are simple and clear, or are even broadcasted, whereas other intentions are hidden – organisms may deliberately hide their intentions in order to trick others. Whereas non-humans may be capable of attributing simple intentions, we think that the ability to represent hidden intentions might be restricted to humans because it might require a full-blown theory of mind, a powerful working memory for simultaneously representing multiple realities or perspectives (Manrique & Walker, 2017), and possibly even the existence of language for representing knowledge propositionally.

3.4 The use of normative rules

The use of norms is a potential key complement to the socio-cognitive abilities discussed in the previous section. Normative/moral understanding is likely to be involved in managing own reputation and in evaluating others' reputations (Fig. 1). To have a good reputation, individuals must comply with some norms or moral standards and check that their behaviour aligns with those norms. The same goes for judging others' reputations, as individuals must contrast a potential partner's behaviour with the very same normative/moral standards. If humans did not possess an awareness of what the "right" behaviour is, it would become harder to choose partners based on whether they do the "right" thing. In indirect reciprocity models, the strong standing strategy makes a clear distinction between what is right and what is wrong, based on the standing of the recipient (Kandori, 1992; Ohtsuki & Iwasa, 2007). This can only work if all players converge on a specific norm that defines who is worthy of help, and who is unworthy of help. Thus, indirect reciprocity systems require a species to be able to use norms. In contrast, reputation-based partner choice can function without norms (e.g., if third parties only assess the actor's ability to help). That being said, reputation-based partner choice might also be affected by norms: the same helpful act may be seen as generous if the norm is to help less, or stingy if the norm is to help more (Barclay, 2013). It might be advantageous to compare potential partners to the norm to know whom to choose (McNamara et al., 2008), or to compare oneself to the norm and adjust one's own cooperation up or down accordingly (Barclay, 2013, 2016).

Human infants are born into a world filled with social norms. Throughout infancy, children learn how things are done and not done. By the age of around two, children can follow adults' requests and conform to others' social behaviours (Rakoczy & Schmidt, 2013). At around the age of three, children can infer norms by observing others acting in a certain way without needing adult directives. At the same time, they also start enforcing norms on others (Vaish et al., 2011). By around five years of age, children reach another milestone of normative development: the spontaneous creation of their own rules (Grueneisen & Tomasello, 2017). Although cultural norms vary widely in their content and implementation, children all over the world show similar abilities for understanding, following, and enforcing socially prescribed behaviours (Miller, 2007). The ways in which children create and deal with norms suggests a growing understanding that norms are mutual agreements which result in rights and obligations for each individual involved. Interestingly, children's concern about their own reputation (and attempts at actively managing it) seems to trail their normative development (Kelsey et al., 2018; Engelmann et al., 2012), i.e., children's reputation management develops after their ability to view norms as a mutually-agreed upon standards for collaborative interactions.

If normative development encompasses the ability to view norms as a set of standards for interactions, then it can only originate in species where collaborative interactions are initiated by joint agency. Given the lack of evidence for shared agency and intentionality in chimpanzees, the existence of a social system based on collective norms and influenced by reputation seems highly unlikely (Schmidt, & Rakoczy, 2019; Tomasello, 2019). Also, given the sparse evidence for social norms in chimpanzees, it is unsurprising that there is little evidence for norms in other species either. In both vervet monkeys and great tits, there is evidence that migrating individuals may give up previously learned preferences and conform to local arbitrary preferences (van de Waal et al., 2013; Aplin et al., 2015). If such conformity did represent norm-following, then these species might theoretically be capable of cooperative systems based on social norms. Without such norm-following, the evolution of reputation-based cooperation is less likely or less efficient.

4. Reputation-based cooperation in non-human species

Although cognitive constraints may prevent many non-human species from displaying complex forms of reputation-based cooperation (Izuma, 2012), they may have simpler forms that are less cognitively demanding. In social species, individuals often interact in communication networks, where bystanders may eavesdrop on interactions to extract valuable information (McGregor, 2005). Therefore, acting in a communication network has three potential payoff consequences: the payoff obtained from the current interaction, the effect of one's own action on the partner's future behaviour towards self, and the effect of one's own action on the future behaviour of any bystander that learns about the action. Interactions in a communication network therefore allow individuals to identify potentially cooperative or aggressive individuals in their social environment and to adjust their behaviour appropriately. Moreover, the possibility for bystander responsiveness might incentivise individuals to adjust their current behaviour when they are observed, a phenomenon known as 'audience effects' (Matos & Schlupp, 2005). This concept shares features with reputation management in humans.

While eavesdropping and audience effects are widespread among vertebrates and have even been documented in invertebrates, convincing evidence exists primarily in competitive contexts (McGregor, 2005). By contrast, in species other than our own, there is a paucity of evidence demonstrating that individuals show a concern for gaining a prosocial reputation. Various arguments can be made why signals are likely to be honest in a competitive context (Arnott & Elwood, 2009; Johnstone & Bshary, 2004) but less reliable in a cooperative context (Johnstone & Bshary, 2007; Barclay, 2013; André, 2010; Bebbington et al., 2017). In a competitive context, individual aggressiveness is likely to be correlated with strength, which is based on metastable features like size, muscle mass, agility and experience. Therefore, signals of formability are difficult to fake and more likely to be honest. The honesty of such signals can change the benefits associated with paying attention to them: eavesdropping in order to gain information on a potential partner's formidability is potentially self-serving. In return, strong individuals may benefit from signaling their strength to eavesdropping bystanders, for example by displaying after a victorious fight, or attacking those lower in the hierarchy after a defeat (Kazim & Aureli, 2005) in order to reduce the likelihood of being the target of future challenges. Strong individuals may even pick a fight that yields a short-term negative payoff to reduce the likelihood of being challenged by bystanders in the future (Johnstone & Bshary, 2004).

Nevertheless, there are a handful of examples from non-human species that are suggestive of reputation-based cooperation. In various species, individuals may temporarily act as a watchman by looking out for predators while the rest of the group forages. While such behaviour has been interpreted as immediately self-serving as it is mostly done by satiated individuals (Clutton-Brock et al., 1999), experiments involving dwarf mongooses have shown that playbacks of an individual's watchman calls increases the amount of grooming this individual receives later in the day (Kern & Radford, 2018). In vervet monkeys, males and females that contribute during territorial disputes receive more grooming by other group members (Arseneau-Robar et al., 2016). In Arabian babblers and Siberian jays, males act more aggressively towards predators in the presence of females, which is suggestive of males displaying in the context of female mate choice (Zahavi, 1995; da Cunha et al., 2017). In all these cases, there is no specific recipient of the initial helpful act, meaning that the source of eventual return benefits is uncertain.

Perhaps the best studied case is the marine cleaning mutualism involving the cleaner wrasse *Labroides dimidiatus* and its 'client' fish. Cleaners remove ectoparasites from clients, which benefits both partners (Côté, 2000). However, cleaners prefer to eat client mucus (Grutter & Bshary, 2003), which is detrimental to client health and hence constitutes cheating. As cleaners have about 2000 interactions per day (Grutter, 1995), ongoing interactions often take place in the presence of other clients. These bystanders observe the ongoing interaction and invite for inspection if the cleaner behaves cooperatively - but leave if they witness a conflict between cleaner and current client (Bshary, 2002), and may swim to another cleaner instead. As a consequence of this client decision rule, cleaners are more cooperative in the presence of bystanders (Bshary & Grutter, 2006; Pinto et al., 2011). Moreover, cleaners stop adjusting service quality if bystanders stop exerting such partner choice (Triki et al., 2018, 2020).

Some features of the cleaner-client interaction structure might facilitate reputation-based cooperation. First, memory requirements are minimal: bystanders need only consider the currently observed interaction to make an immediate decision whether to invite or to avoid inspection. Second, the bystander's decision is self-serving as there is short-term autocorrelation of cleaner service quality; and the clients get immediate feedback on their decisions, which facilitates learning (Skinner, 1953). Cleaners who feed against preference must delay immediate gratification, but the positive or negative feedback of this decision (clients inviting for inspection or swimming away) is almost immediate, which also facilitates learning. Thus, basic reinforcement learning might suffice to achieve reputation-based cooperation in this system.

One obvious distinction between reputation-based cooperation in humans and other animals is that humans use language (see other contributions to this theme issue). Language allows people to flexibly exchange information about other individuals (Wu et al., 2016) – and can potentially also increase the amount of information that can be exchanged. Language can also help humans to represent (and hence encode) and recall social norms and might also be a pre-requisite for expressing more complex aspects of social cognition that are likely to be involved in managing and evaluating reputations. Despite its likely importance, we do not discuss language in this review, because it acts more as a multiplier on other cognitive mechanisms, and we instead focus on other proximate cognitive mechanisms that form the basic building blocks of reputation-based cooperation in humans.

5. Discussion

We have presented four basic psychological building blocks that we consider important facilitators for complex reputation-based cooperation: working memory, delay of gratification, theory of mind, and social norms. Working memory allows for parallel processing of diverse information, to properly assess others' actions and update their reputation scores. Delay of gratification is useful for many types of cooperation, but may be particularly relevant for reputation-based cooperation where the returns come from a future interaction with an observer rather than an immediate reciprocation by one's current partner. Theory of mind makes it easier to properly assess others' actions, and reduces the risk that spreading errors will undermine cooperation. Finally, norms support theory of mind by giving individuals a benchmark of what is right or wrong. The more developed that each of these building blocks is, the more complex the interaction structure can become. We are aware that by picking these four socio-cognitive mechanisms we leave out other processes that might be involved, e.g. long-term memory, yet we think the ones we picked are more critical and better allow for comparison across species.

Reputation-based cooperation based on partner choice might often be less cognitively demanding than that based on indirect reciprocity. On the one hand, reputation-based partner choice might require a better ability to delay gratification (as it might take several acts of investment to outcompete competitors and be chosen by third parties), while IR games are typically set up in such a way that individuals alternate roles as helper and recipient. On the other hand, reputation-based partner choice can exist in cognitively simple forms like "walk away or reject partner if they seem uncooperative" (Aktipis, 2004; McNamara et al., 2008); this does not require high working memory, theory of mind, or normative behaviour, though these abilities can make reputation-based partner choice more efficient. In contrast, analyses of indirect reciprocity games have shown that Kandori is the simplest strategy yielding stable cooperation (Santos et al., 2018), and Kandori requires norms, theory of mind to identify errors, and as a consequence more computational power (e.g. working memory). Therefore, the vast majority of animal species may be cognitively constrained from implementing indirect reciprocity, and hence be limited to simple forms of reputation-based partner choice. In line with this hypothesis, the few non-human examples of reputation-based cooperation largely fit the concept of reputation-based partner choice, not indirect reciprocity. Most of the examples seem to be about one party gaining information about another, to know whom to cooperate or mate with, or whom to avoid in fights – a type of reputation-based partner choice based on eavesdropping (McGregor, 2005). As such, there is a clear evolutionary path for reputation-based partner choice: start with cognitively simple eavesdropping, which then evolves into an active signalling system (see Biernaskie et al., 2018 for cues evolving into signals), with more complex abilities arising later in both signallers and receivers in order to perform better within that signalling system.

Future work should further clarify the role of these cognitive mechanisms in reputation-based cooperation in both humans and non-humans. Studies could investigate reputation-based cooperation in humans when these cognitive mechanisms cannot function properly, such as experimental paradigms that increase cognitive load (e.g., Milinski & Wedekind, 1998), special populations that lack some of these cognitive mechanisms (e.g., Cage et al., 2013; Izuma et al., 2011), or online networks where one cannot use these mechanisms. Non-human studies could artificially grant these abilities to non-humans, for example by dissociating cooperative investments from ability to delay gratification (c.f. Stephens et al., 2002). Other studies could use other creative ways of outsourcing cognition to see how they affect reputation-based cooperation. We look forward to seeing further tests of the cognitive building blocks of reputation-based cooperation.

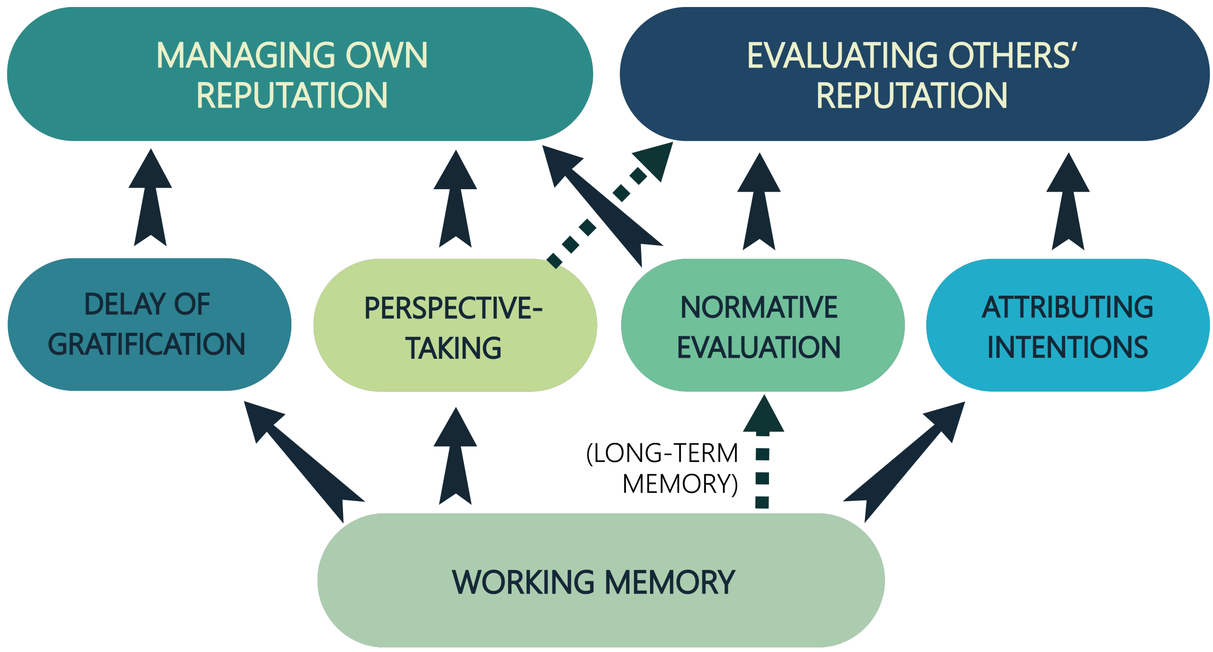

Figure 1. Depiction of how our four socio-cognitive mechanisms are recruited for the managing of one's own reputation as opposed to evaluating third-party reputations. No connecting lines indicate there is no need for the socio-cognitive mechanism in question to be recruited. Arrow continuity expresses the activation of the mechanism is heavily involved in reputation management and/or evaluation. Dotted lines indicate minor involvement. For instance, perspective taking is key to managing one's own reputation, as we need to see how our acts will appear to a putative observer, yet perspective taking matters less for evaluating third-party reputations. The opposite is true for attributing intentions. Delay of gratification might be involved in managing one's own reputation as it allows one to resist current temptations to exploit an interaction partner in order to obtain higher future payoffs associated with curating a good reputation. We expect delay of gratification to be less important for evaluating third-party reputations. Normative understanding is involved in both managing of one's own reputation and evaluating third-party reputations. Working memory is placed in a different level because it enhances the other psychological processes and greatly boosts their efficiency. While working memory is highly involved in delaying gratification, adopting the other's perspective, and attributing intentions, its involvement in moral evaluation is lower as norms are stored in long-term memory.

References

Aellen, M., Dufour, V., & Bshary, R. (2021). Cleaner fish and other wrasse match primates in their ability to delay gratification. *Animal Behaviour, 176*, 125-143. https://doi.org/10.1016/j.anbehav.2021.04.002

Aktipis, C. A. (2004). Know when to walk away: contingent movement and the evolution of cooperation. *Journal of theoretical biology*, *231*(2), 249-260. http://dx.doi.org/10.1016/j.jtbi.2004.06.020

André, J.-B. (2010). The evolution of reciprocity: social types or social incentives? *The American Naturalist*, *175*(2), 197-210. http://dx.doi.org/10.1086/649597

Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow giving. *The economic journal, 100*(401), 464-477. https://doi.org/10.2307/2234133

Aloise-Young, P. A. (1993). The development of self-presentation: Self-promotion in 6-to 10-year-old children. *Social Cognition*, *11*(2), 201-222. http://dx.doi.org/10.1521/soco.1993.11.2.201

Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist? Evaluating strategic colorblindness in social interaction. *Journal of personality and social psychology, 95*(4), 918-932. https://doi.org/10.1037/a0011990

Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C. (2015). Experimentally induced innovations lead to persistent culture via conformity in wild birds. *Nature*, *518*(7540), 538-541. http://dx.doi.org/10.1038/nature13998

Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests. *Animal Behaviour*, *77*(5), 991-1004. http://dx.doi.org/10.1016/j.anbehav.2009.02.010

Arseneau-Robar, T. J. M., Taucher, A. L., Müller, E., van Schaik, C., Bshary, R., & Willems, E. P. (2016). Female monkeys use both the carrot and the stick to promote male participation in intergroup fights. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1843), 20161817. http://dx.doi.org/10.1098/rspb.2016.1817

Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans. *Evolution & Human Behavior*, *34*(3), 164-175. http://dx.doi.org/10.1016/j.evolhumbehav.2013.02.002

Barclay, P. (2015). Reputation. In D. Buss (Ed.) *Handbook of Evolutionary Psychology (2nd Ed.)*, pp. 810-828. Hoboken, NJ: J. Wiley & Sons. http://dx.doi.org/10.1002/9781119125563.evpsych233

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology*, *7*, 33-38. http://dx.doi.org/10.1016/j.copsyc.2015.07.012

Barclay, P., & Barker, J.L. (2020). Greener than thou: people who protect the environment are more cooperative, compete to be environmental, and benefit from reputation. *Journal of Environmental Psychology*, *72*, 101441. http://dx.doi.org/10.1016/j.jenvp.2020.101441

Barrett, H.C., Bolyanatz, A., Crittenden, A.N., Fessler, D.M.T., Fitzpatrick, S., Gurven, M., Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B.A., Stich, S., von Rueden, C., Zhao, W., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the role of intentions in moral judgement. *Proceedings of the New York Academy of Sciences*, *113*(17), 4688-4693. http://dx.doi.org/10.1073/pnas.1522070113

Baumeister, R. F., Maranges, H. M., & Sjåstad, H. (2018). Consciousness of the future as a matrix of maybe: Pragmatic prospection and the simulation of alternative possibilities. *Psychology of Consciousness: Theory, Research, and Practice*, *5*(3), 223–238.

Bebbington, K., MacLeod, C., Ellison, T. M., & Fay, N. (2017). The sky is falling: evidence of a negativity bias in the social transmission of information. *Evolution and Human Behavior*, *38*(1), 92-101. http://dx.doi.org/10.1016/j.evolhumbehav.2016.07.004

Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants' understanding of intentional action. *Developmental psychology*, *41*(2), 328-337. http://dx.doi.org/10.1037/0012-1649.41.2.328

Beran, M. J., & Evans, T. A. (2006). Maintenance of delay of gratification by four chimpanzees (Pan troglodytes): The effects of delayed reward visibility, experimenter presence, and extended delay intervals. *Behavioural Processes*, *73*(3), 315-324. http://dx.doi.org/10.1016/j.beproc.2006.07.005

Biernaskie, J. M., Perry, J. C., & Grafen, A. (2018). A general model of biological signals, from cues to handicaps. *Evolution Letters, 2*(3), 201-209. https://doi.org/10.1002/evl3.57

Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality? *Proceedings of the Royal Society of London B*, *285*, 20180116. http://dx.doi.org/10.1098/rspb.2018.0116

Bshary, R. (2002). Biting cleaner fish use altruism to deceive image–scoring client reef fish. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1505), 2087-2093. http://dx.doi.org/10.1098/rspb.2002.2084

Bshary, R., & Grutter, A. S. (2006). Image scoring and cooperation in a cleaner fish mutualism. *Nature*, *441*(7096), 975-978. http://dx.doi.org/10.1038/nature04755

Bueno-Guerra, N., Colell, M., & Call, J. (2020). Effects of indirect reputation and type of rearing on food choices in chimpanzees (Pan troglodytes). *Behavioral Ecology and Sociobiology, 74*(6), 1-10. https://doi.org/10.1007/s00265-020-02861-w

Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen competitors. *Nature Communications*, *7*(1), 1-6. http://dx.doi.org/10.1038/ncomms10506

Bugnyar, T., Stöwe, M., & Heinrich, B. (2004). Ravens, Corvus corax, follow gaze direction of humans around obstacles. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1546), 1331-1336. http://dx.doi.org/10.1098/rspb.2004.2738

Cage, E., Pellicano, E., Shah, P., & Bird, G. (2013). Reputation management: evidence for ability but reduced propensity in autism. *Autism Research*, *6*, 433-442. http://dx.doi.org/10.1002/aur.1313

Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). `Unwilling' versus `unable': chimpanzees' understanding of human intentional action. *Developmental Science, 7*(4), 488-498. http://dx.doi.org/10.1111/j.1467-7687.2004.00368.x

Call, J., & Tomasello, M. (1998). Distinguishing intentional from accidental actions in orangutans (Pongo pygmaeus), chimpanzees (Pan troglodytes) and human children (Homo sapiens). *Journal of Comparative Psychology*, *112*(2), 192-206. http://dx.doi.org/10.1037/0735-7036.112.2.192

Canteloup, C., Piraux, E., Poulin, N., & Meunier, H. (2016). Do Tonkean macaques (Macaca tonkeana) perceive what conspecifics do and do not see? *PeerJ,* 4:e1693 http://dx.doi.org/10.7717/peerj.1693

Carruthers, P., (2013). Evolution of working memory. *Proceedings of the National Academy of Sciences of the USA*, *110*(Supplement 2), 10371-10378. http://dx.doi.org/10.1073/pnas.1301195110

Clutton-Brock, T. H., O'riain, M. J., Brotherton, P. N., Gaynor, D., Kansky, R., Griffin, A. S., & Manser, M. (1999). Selfish sentinels in cooperative mammals. *Science*, *284*(5420), 1640-1644. http://dx.doi.org/10.1126/science.284.5420.1640

Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by scrub jays. *Nature*, *395*(6699), 272-274. http://dx.doi.org/10.1038/26216

Coricelli, G., Joffily, M., Montmarquette, C., & Villeval, M.C. (2010). Cheating, emotions, and rationality: an experiment on tax evasion. *Experimental Economics*, *13*(2), 226-247.

Correia, S. P., Dickinson, A., & Clayton, N. S. (2007). Western scrub-jays anticipate future needs independently of their current motivational state. *Current Biology*, *17*(10), 856-861. http://dx.doi.org/10.1016/j.cub.2007.03.063

Costes-Thiré, M., Levé, M., Uhlrich, P., Pasquaretta, C., De Marco, A., & Thierry, B. (2015). Evidence that monkeys (Macaca tonkeana and Sapajus apella) read moves, but no evidence that they read goals. *Journal of Comparative Psychology*, *129*(3), 304-310. http://dx.doi.org/10.1037/a0039294

Côté, I. M. (2000). Evolution and ecology of cleaning symbioses in the sea. *Oceanography and Marine Biology*, *38*, 311–355. Taylor & Francis, CRS Press, United Kingdom.

da Cunha, F. C. R., Fontenelle, J. C. R., & Griesser, M. (2017). The presence of conspecific females influences male-mobbing behavior. *Behavioral Ecology and Sociobiology*, *71*(3), 52. http://dx.doi.org/10.1007/s00265-017-2267-7

Curry, O.S., Price, M.E., & Price, J.G. (2008). Patience is a virtue: cooperative people have lower discount rates. *Personality and Individual Differences*, *44*(3), 778-783. http://dx.doi.org/10.1016/j.paid.2007.09.023

Dere, D., Zlomuzica, A., Dere, E. (2019). Fellow travellers in cognitive evolution: Co-evolution of working memory and mental time travel? *Neuroscience & Biobehavioral Reviews*, *105*, 94-105. https://doi.org/10.1016/j.neubiorev.2019.07.016.

Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal cortex function, the delayed-response task. *Developmental Psychobiology*, *22*(3), 271-294. http://dx.doi.org/10.1002/dev.420220307

Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy. *Psychological science*, *21*(4), 523-527. http://dx.doi.org/10.1177/0956797610364119

Dunfield, K., Kuhlmeier, V. A., O'Connell, L., & Kelley, E. (2011). Examining the diversity of prosocial behavior: Helping, sharing, and comforting in infancy. *Infancy*, *16*(3), 227-247.

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not chimpanzees, attempt to manage their reputations. *PLoS On*e 7:e48433. http://dx.doi.org/10.1371/journal.pone.0048433

Engelmann, J. M., Herrmann, E., & Tomasello, M. (2016). The effects of being watched on resource acquisition in chimpanzees and human children. *Animal Cognition, 19*(1), 147-151. http://dx.doi.org/10.1007/s10071-015-0920-y

Evans, T. A., & Beran, M. J. (2007). Chimpanzees use self-distraction to cope with impulsivity. *Biology Letters*, 3, 599-602. http://dx.doi.org/10.1098/rsbl.2007.0399

Flavell, J. H. (1977). The development of knowledge about visual perception. Nebraska Symposium on Motivation, 25, 43–76.

Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental Psychology*, *17*(1), 99-103. http://dx.doi.org/10.1037/0012-1649.17.1.99

Flombaum, J. & Santos, L. (2005). Rhesus monkeys can assess the visual perspective of others when competing for food. *Current Biology*, *15*(5), 447-452. http://dx.doi.org/10.1016/j.cub.2004.12.076

Fuster, J.M. (2001). The prefrontal cortex – an update: Time is of the essence. *Neuron*, *30*(2), 319-333.

Gathercole, S.E., S.J. Pickering, B. Ambridge, B. & Wearing, H. (2004). The structure of working memory from 4 to 15 years of age. *Development Psychology*, *40*(2), 177-190. http://dx. doi.org/10.1037/0012-1649.40.2.177

Gerber, A.A., Green, D.P., & Larimer, C.W. (2008). Social pressure and voter turnout: evidence from a large-scale field experiment. *American Political Science Review*, *102*(1), 33-48. http://dx.doi.org/10.1017/S000305540808009X

Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants. *Nature*, *415*(6873), 755. http://dx.doi.org/10.1038/415755a

Grueneisen, S. & Tomasello, M. (2017). Children coordinate in a recurrent social dilemma by taking turns and along dominance asymmetries. *Developmental Psychology*, *53*(2), 265-273. http://dx.doi.org/10.1037/dev0000236

Grutter, A. S. (1995). The relationship between cleaning rates and ectoparasites loads in coral reef fishes. *Marine Ecology Progress Series, 118*, 51–58. http://dx.doi.org/10.3354/meps118051

Grutter, A.S., & Bshary, R. (2003). Cleaner wrasse prefer client mucus: support for partner control mechanisms in cleaning interactions. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270* (Supplement 2), S242–S244. http://dx.doi.org/10.1098/rsbl.2003.0077

Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771-785. http://dx.doi.org/10.1006/anbe.1999.1377

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves modulation of the vmPFC valuation system. *Science*, *324*(5927), 646–648. http://dx.doi.org/10.1126/science.1168450

Harris, A.C., & Madden, G.J. (2002). Delay discounting and performance on the Prisoner's Dilemma Game. *Psychological Record*, *52*(4), 429-440. http://dx.doi.org/10.1007/BF03395196

Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2013). Direct and indirect reputation formation in nonhuman great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla, Pongo pygmaeus) and human children (Homo sapiens). *Journal of Comparative Psychology, 127*(1), 63–75. https://doi.org/10.1037/a0028929

Heyes, C.M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences, 21*(1), 101-148. http://dx.doi.org/10.1017/S0140525X98000703

Heyman, G., Barner, D., Heumann, J., & Schenck, L. (2014). Children's sensitivity to ulterior motives when evaluating prosocial behavior. *Cognitive science*, *38*(4), 683-700. http://dx.doi.org/10.1111/cogs.12089

Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current Biology*, *17*(23), R1004-R1005. http://dx.doi.org/10.1016/j.cub.2007.10.027

Izuma, K. (2012). The social neuroscience of reputation. *Neuroscience research, 72*(4), 283-288.

Izuma, K., Matsumoto, K., Camerer, C.F., & Adolphs, R. (2011). Insensitivity to social reputation in autism. *Proceedings of the National Academy of Sciences*, *108(42)*, 17302-17307. http://dx.doi.org/10.1073/pnas.1107038108

Johnstone, R. A., & Bshary, R. (2004). Evolution of spite through indirect reciprocity. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1551), 1917-1922. http://dx.doi.org/10.1098/rspb.2003.2581

Johnstone, R. A., & Bshary, R. (2007). Indirect reciprocity in asymmetric interactions: when apparent altruism facilitates profitable exploitation. *Proceedings of the Royal Society B: Biological Sciences*, *274*(1629), 3175-3181. http://dx.doi.org/10.1098/rspb.2003.2581

Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic Studies*, *59*(1), 63-80. http://dx.doi.org/10.2307/2297925

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015). The goggles experiment: can chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour*, *105*, 211-221. http://dx.doi.org/10.1016/j.anbehav.2015.04.028

Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2016). Differing views: Can chimpanzees do Level 2 perspective-taking? *Animal Cognition*, *19*, 555-564. http://dx.doi.org/10.1007/s10071-016-0956-7

Kawai, N., Nakagami, A., Yasue, M., Koda, H., & Ichinohe, N. (2019). Common marmosets (Callithrix jacchus) evaluate third-party social interactions of human actors but Japanese monkeys (Macaca fuscata) do not. *Journal of Comparative Psychology, 133*(4), 488–495. https://doi.org/10.1037/com0000182

Kazim, A. J. N., & Aureli, F. (2005). Redirection of aggression: multiparty signalling within a network? In P. McGregor (Ed) Animal Communication Networks, pp. 191-218. Cambridge, UK: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511610363.013

Kelsey, C., Grossmann, T., & Vaish, A. (2018). Early reputation management: Three-year-old children are more generous following exposure to eyes. *Frontiers in psychology*, *9*, 698. http://dx.doi.org/10.3389/fpsyg.2018.00698

Kern, J. M., & Radford, A. N. (2018). Experimental evidence for delayed contingent cooperation among wild dwarf mongooses. *Proceedings of the National Academy of Sciences*, *115*(24), 6255-6260. http://dx.doi.org/10.1073/pnas.1801000115

Koomen, R., Grueneisen, S., & Herrmann, E. (2020). Children delay gratification for cooperative ends. *Psychological science*, *31*(2), 139-148. http://dx.doi.org/10.1177/0956797619894205

Lacetera, N., & Macis, M. (2010). Social image concerns and prosocial behavior: field evidence from a nonlinear incentive scheme. *Journal of Economic Behavior & Organization*, *76*(2), 225-237.

Langer, J. (1980). The Origins of Logic, Six to Twelve Months. New York, Academic Press.

Langer, J. (1986). The Origins of Logic: One to Two Years. New York: Academic Press.

Langer, J. (2000). The heterochronic evolution of Primate cognitive development. In: Parker, S.T., Langer, J., McKinney, M.L.(Eds.), Biology, Brains and Behavior. The Evolution of Human Development. Santa Fe, American School of Research Press, "Advanced Seminar Series", and Oxford, James Currey, pp. 215-235. http://dx.doi.org/10.1162/biot.2006.1.1.41

Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more generous when others are aware of their actions. *PloS one*, *7*(10), e48292. http://dx. doi.org/10.1371/journal.pone.0048292

Luerssen, A., Gyurak, A., Ayduk, O., Wendelken, C., & Bunge, S. A. (2015). Delay of gratification in childhood linked to cortical interactions with the nucleus accumbens. *Social cognitive and affective neuroscience*, *10*(12), 1769-1776. http://dx.doi.org/10.1093/scan/nsv068

Ma, F., Zeng, D., Xu, F., Compton, B. J., & Heyman, G. D. (2020). Delay of Gratification as Reputation Management. *Psychological Science*, *31*(9), 1174-1182. http://dx.doi.org/10.1177/0956797620939940

Manrique, H. M., Walker, M. J. (2017). *Early Evolution of Human Memory*. London: Palgrave Macmillan. http://dx.doi.org/10.1007/978-3-319-64447-9

Matos, R., & Schlupp, I. (2005). Performing in front of an audience: Signallers and the social environment. In P. McGregor (Ed.), *Animal Communication Networks* (pp. 63-83). Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511610363

McAuliffe, K., Blake, P. R., & Warneken, F. (2020). Costly fairness in children is influenced by who is watching. *Developmental Psychology, 56*(4), 773–782. https://doi.org/10.1037/dev0000888

McAuliffe, K., Drayton, L.A., Royka, A., Aellen, M., Santos, L.R. & Bshary, R. (submitted) Do cleaner fish know what others can and cannot see?

McGregor, P. (Ed.). (2005). *Animal Communication Networks*. Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511610363

McNamara, J. M., Barta, Z., Frohmage, L., & Houston, A. I. (2008). The coevolution of choosiness and cooperation. *Nature*, *451*(7175), 189-192. http://dx.doi.org/10.1038/nature06455

Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts by 18-month-old children. *Developmental psychology*, *31*(5), 838-850. http://dx.doi.org/10.1037/0012-1649.31.5.838

Milinski, M., & Wedekind, C. (1998). Working memory constrains human cooperation in the Prisoner's Dilemma. *Proceedings of the National Academy of Sciences*, *95*(23), 13755-13758. http://dx.doi.org/10.1073/pnas.95.23.13755

Milinski, M., Semmann, D., Bakker, T. C., & Krambeck, H. J. (2001). Cooperation through indirect reciprocity: image scoring or standing strategy? *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *268*(1484), 2495-2501. http://dx.doi.org/10.1098/rspb.2001.1809

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, *63*(2), 81-97. http://dx.doi.org/10.1037/h0043158

Miller, J. G. (2007). Cultural psychology of moral development. In S. Kitayama & D. Cohen (Eds.), Handbook of cultural psychology (p. 477–499). The Guilford Press.

Miller, R., Boeckle, M., Jelbert, S. A., Frohnwieser, A., Wascher, C. A., & Clayton, N. S. (2019). Self-control in crows, parrots and nonhuman primates. *Wiley Interdisciplinary Reviews: Cognitive Science*, *10*(6), e1504. http://dx.doi.org/10.1002/wcs.1504

Mischel, W., & Ebbesen, E.B. (1970). Attention in delay of gratification. *Journal of Personality and Social Psychology*, 16, 329–337. http://dx.doi.org/10.1037/h0029815

Mischel, H. N., & Mischel, W. (1983). The development of children's knowledge of self-control strategies. *Child Development*, *54*(3), 603-619. http://dx.doi.org/10.2307/1130047

Moll, H., & Meltzoff, A.N. (2011). How does it look? Level 2 perspective-taking at 36 months of age. *Child development*, 82(2), 661-673. http://dx.doi.org/10.1111/j.1467-8624.2010.01571.x

Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British Journal of Developmental Psychology*, *24*(3), 603-613. http://dx.doi.org/10.1348/026151005X55370

Moriguchi, Y., Tanaka, M., & Itakura, S. (2011). Executive function in young children and chimpanzees (Pan troglodytes): evidence from a non-verbal dimensional change card sort task. *Journal of Genetic Psychology, 172*(3), 252-265. http://dx.doi.org/10.1080/00221325.2010.534828

Nettle, D., Cronin, K. A., & Bateson, M. (2013). Responses of chimpanzees to cues of conspecific observation. *Animal Behaviour, 86*(3), 595-602. https://doi.org/10.1016/j.anbehav.2013.06.015

Ohtsuki, H. & Iwasa, Y. (2007). Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *Journal of Theoretical Biology*, *24*(3), 518–531. http://dx.doi.org/10.1016/j.jtbi.2006.08.01

Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other individuals' line of sight. *Psychological Science*, *18*(5), 462-468. http://dx.doi.org/10.1111/j.1467-9280.2007.01922.x

Péron, F., Rat-Fischer, L., Nagle, L., & Bovet, D. (2010). 'Unwilling'versus 'unable': do grey parrots understand human intentional actions?. *Interaction studies*, *11*(3), 428-441. http://dx.doi.org/10.1075/is.11.3.06per

Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009). 'Unwilling'versus 'unable': capuchin monkeys'(Cebus apella) understanding of human intentional action. *Developmental science*, *12*(6), 938-945. http://dx.doi.org/10.1111/j.1467-7687.2009.00840.x

Pinto, A.I., Oates, J., Grutter, A.S., & Bshary, R. (2011). Cleaner wrasse Labroides dimidiatus are more cooperative in the presence of an audience. *Current Biology*, *21(13)*, 1140–1144. http://dx.doi.org/10.1016/j.cub.2011.05.021

Potì, P., Langer, J., Savage-Rumbaugh, S., & Brakke, K. E. (1999). Spontaneous logicomathematical constructions by chimpanzees (Pan troglodytes, P. paniscus). *Animal Cognition*, *2*(3), 147-156. http://dx.doi.org/10.1007/s100710050035

Povinelli, D.J., & Eddy, T. (1996). What young chimpanzees know about seeing. *Monographs of the Society for Research in Child Development, 61*(3), 1-189. http://dx.doi.org/10.2307/1166159

Povinelli, D. J., Perilloux, H. K., Reaux, J. E., & Bierschwale, D. T. (1998). Young and juvenile chimpanzees'(Pan troglodytes) reactions to intentional versus accidental and inadvertent actions. *Behavioural Processes*, *42*(2-3), 205-218. http://dx.doi.org/10.1016/S0376-6357(97)00077-6

Raihani, N., & Power, E. A. (2021). No Good Deed Goes Unpunished: the social costs of prosocial behaviour. PsyArXiv. http://dx.doi.org/10.31234/osf.io/ebfrg

Rakoczy, H., & Schmidt, M. F. (2013). The early ontogeny of social norms. *Child Development Perspectives*, *7*(1), 17-21. http://dx.doi.org/10.1111/cdep.12010

Read, D. W. (2008). Working memory: A cognitive limit to non-human primate recursive thinking prior to hominid evolution. *Evolutionary Psychology*, *6*(4), 147470490800600413. http://dx.doi.org/10.1177/147470490800600413

Read, D., & Van Der Leeuw, S. (2008). Biology is only part of the story…. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 1959-1968.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 427-431. http://dx.doi.org/10.1098/rspb.1998.0312

Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin, 128*(3), 473–489.

Roberts, G., Raihani, N., Bshary, R., Manrique, H.M., Fariña, A., Samu, F., & Barclay, P. (submitted). The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. Manuscript submitted to this special issue of *Philosophical Transactions of the Royal Society*.

Russell, Y. I., Call, J., & Dunbar, R. I. (2008). Image scoring in great apes. *Behavioural processes, 78*(1), 108-111. https://doi.org/10.1016/j.beproc.2007.10.009

Rutland, A., Cameron, L., Bennett, L., & Ferrell, J. (2005). Interracial contact and racial constancy: A multi-site study of racial intergroup bias in 3–5 year old Anglo-British children. *Journal of Applied Developmental Psychology*, *26*(6),699-713. http://dx.doi.org/10.1016/j.appdev.2005.08.005

Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G. (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication attempts of Hamlin and Wynn (2011). *PloS one, 10*(11), e0140570. https://doi.org/10.1371/journal.pone.0140570

Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, Macaca mulatta, know what others can and cannot hear. *Animal Behaviour*, *71*(5), 1175–1181. http://dx.doi.org/10.1016/j.anbehav.2005.10.007

Santos, F. P., Santos, F. C., & Pacheco, J. M. (2018). Social norm complexity and past reputations in the evolution of cooperation. *Nature*, *555*(7695), 242-245. http://dx.doi.org/10.1038/nature25763

Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the future: The prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661.

Schino, G., Boggiani, L., Mortelliti, A., Pinzaglia, M., & Addessi, E. (2021). Testing the two sides of indirect reciprocity in tufted capuchin monkeys. *Behavioural Processes, 182*, 104290. https://doi.org/10.1016/j.beproc.2020.104290

Schmidt, M.F.H., & Rakoczy, H. (2019) On the uniqueness of human normative attitudes. In: Bayertz, K., Roughley, N. (eds) *The normative animal? On the anthropological significance of social, moral and linguistic norms*. Oxford University Press, Oxford

Schnell, A.K., Boeckle, M., Rivera, M., Clayton, N.S., & Hanlon, R.T. (2021). Cuttlefish exert self-control in a delay of gratification task. *Proceedings of the Royal Society of London. Series B: Biological Sciences*. *288*(1946), 20203161. https://doi.org/10.1098/rspb.2020.3161

Sebastián-Enesco, C., & Warneken, F. (2015). The shadow of the future: 5-year-olds, but not 3-year-olds, adjust their sharing in anticipation of reciprocation. *Journal of Experimental Child Psychology*, *129*, 40–54. http://dx.doi.org/10.1016/j.jecp.2014.08.007

Seligman, M. E. P., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the future or driven by the past. *Perspectives on Psychological Science*, *8*(2), 119–141. https://doi.org/10.1177/1745691612474317

Shaw, A., Montinari, N., Piovesan, M., Olson, K.R., Gino, F., & Norton, M.I. (2014). Children Develop a Veil of Fairness. *Journal of Experimental Psychology: General*, *143*(1), 363–375. https://doi.org/10.1037/a0031247

Sjåstad, H. (2019). Short-sighted greed? Focusing on the future promotes reputation-based generosity. *Judgment and Decision Making*, *14*(2), 199–213.

Smith, E. A., & Bliege Bird, R. (2000). Turtle hunting and tombstone opening public generosity as costly signaling. *Evolution and Human Behavior*, *21*(4), 245–261. https://doi.org/10.1016/S1090-5138(00)00031-3

Skinner, B. F. (1953). *Science and Human Behavior*. Free Press.

Spinozzi, G., Natale, F., Langer, J. & Brakke, K.E. (1999). Spontaneous class grouping behavior by bonobos (*Pan paniscus*) and common chimpanzees (*P. troglodytes*). *Animal Cognition*, *2*(3), 157-170. http://dx.doi.org/10.1007/s100710050036

Stephens, D.W., McLinn, C.M., & Stevens, J.R. (2002). Discounting and reciprocity in an iterated Prisoner's Dilemma. *Science*, *298*, 2216-2218. http://dx.doi.org/10.1126/science.1078498

Subiaul, F., Vonk, J., Okamoto-Barth, S. et al. (2008). Do chimpanzees learn reputation by observation? Evidence from direct and indirect experience with generous and selfish strangers. *Animal Cognition, 11*(4), 611–623. https://doi.org/10.1007/s10071-008-0151-6

Suddendorf, T. (2013). *The gap: The science of what separates us from other animals*. New York, NY, US: Basic Books.

Susini, I., Safryghin, A., Hillemann, F., & Wascher, C. A. F. (2020). Delay of gratification in non-human animals: A review of inter- and intra-specific variation in performance. BioRxiv. https://doi.org/10.1101/2020.05.05.078659

Tomasello, M. (2019). The moral psychology of obligation. *Behavioral and Brain Scien*ces, *43*, 1–33. http://dx.doi.org/10.1017/S0140525X19001742

Triki, Z., Wismer, S., Levorato, E., & Bshary, R. (2018). A decrease in the abundance and strategic sophistication of cleaner fish after environmental perturbations. *Global Change Biology, 24*, 481–489. http://dx.doi.org/10.1111/gcb.13943

Triki, Z., Emery, Y., Teles, M.C., Oliveira, R.F., & Bshary, R. (2020). Brain morphology predicts social intelligence in wild cleaner fish. *Nature Communications, 11*, 6423. http://dx.doi.org/10.1038/s41467-020-20130-2

Trösch, M., Bertin, E., Calandreau, L., Nowak R. & Lansade, L. (2020). Unwilling or willing but unable: can horses interpret human actions as goal directed? *Animal Cognition, 23*, 1035–1040. http://dx.doi.org/10.1007/s10071-020-01396-x

Van de Vondervoort, J. W., & Hamlin, J. K. (2018). The early emergence of sociomoral evaluation: infants prefer prosocial others. *Current Opinion in Psychology, 20*, 77-81. https://doi.org/10.1016/j.copsyc.2017.08.014

Van de Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent social learning and conformity shape a wild primate's foraging decisions. *Science*, *340*(6131), 483-485. http://dx.doi.org/10.1126/science.1232769

Völter, C. J., Mundry, R., Call, J., & Seed, A. M. (2019). Chimpanzees flexibly update working memory contents and show susceptibility to distraction in the self-ordered search task. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1907): 20190715. https://doi.org/10.1098/rspb.2019.0715

Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective taking and its relation to prosocial behavior in toddlers. *Developmental psychology*, *45*(2), 534-543. https://doi.org/10.1037/a0014322

Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-party moral transgressions. *British Journal of Developmental Psychology*, *29*(1), 124-130. http://dx.doi.org/10.1348/026151010X532888

Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2018). Death before dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality Science*, *9*(5), 604–613. https://doi.org/10.1177/1948550617720271

Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age. *Infancy, 11*(3), 271-294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x

Washburn, D. A., Gulledge, J. P., James, F., & Rumbaugh, D. M. (2007). A species difference in visuospatial working memory: Does language link "what" with "where"?. *International Journal of Comparative Psychology*, *20*(1), 55-64.

Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007). The perception of rational, goal-directed action in nonhuman primates. *Science*, *317*(5843), 1402-1405. http://dx.doi.org/10.1126/science.1144663

Wu, J., Balliet, D., & Van Lange, P. A. (2016). Reputation management: Why and how gossip enhances generosity. Evolution and Human Behavior, 37(3), 193-201. http://dx.doi.org/10.1016/j.evolhumbehav.2015.11.001

Wu, J., Balliet, D., Tybur, J. M., Arai, S., Van Lange, P. A., & Yamagishi, T. (2017). Life history strategy and human cooperation in economic games. *Evolution and Human Behavior*, *38*(4), 496-505. http://dx.doi.org/10.1016/j.evolhumbehav.2017.03.002

Yoeli, E., Hoffman, M., Rand, D.G., & Nowak, M.A. (2013). Powering up indirect reciprocity with a large-scale field experiment. *Proceedings of the National Academy of Sciences of the USA*, *110*(Supplement 2), 10424-10429. https://doi.org/10.1073/pnas.1301210110

Zahavi, A. (1995). Altruism as a handicap: the limitations of kin selection and reciprocity. *Journal of Avian Biology*, *26*(1), 1-3. http://dx.doi.org/10.2307/3677205

Author Note

Héctor M. Manrique wants to thank Professor Álvaro Arrizabalaga and the MINECO project with reference HAR2017-82483-C3-1-P for financial support.

Nichola Raihani was supported by a Royal Society University Research Fellowship and by the Leverhulme Trust.

Pat Barclay was supported by the Social Science & Humanities Research Council of Canada (SSHRC grant 430287).

Redouan Bshary was supported by the Swiss Science Foundation (grant 310030_192673).

Andrea Fariña was supported by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (AdG agreement n° 785635; PI Carsten K.W. De Dreu).

Flóra Samu was supported by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (grant agreement No. 648693, PI: Károly Takács).

Competing interests: NR is the author of the forthcoming book, The Social Instinct: How Cooperation Shaped the World

Correspondence concerning this article should be addressed to Héctor Marín Manrique, Department of Psychology and Sociology, Universidad de Zaragoza, Campus Universitario de Teruel, Ciudad Escolar, s/n. 44003 Teruel (Spain). E-mail: manrique@unizar.es; Tel: +34 – 978 – 645343; Fax: +34 – 978 – 618103

Héctor M. Manrique https://orcid.org/0000-0002-1943-340X