1

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.

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

69   many domains, including tax compliance (Coricelli et al., 2010); voter turnout (Gerber et al.,

70   2008); energy conservation (Yoeli et al., 2013); environmentalism (Barclay & Barker, 2020);

71   blood donation (Lacetera & Macis, 2010); and more. Most researchers interpret this increased

72   cooperation as being caused by people's concern for reputation.

73

74   However, unlike the ability to evaluate others' reputation, this tendency to strategically manage

75   one's own reputation is not present at all stages of life and instead appears to emerge during

76   development. Although young children (under two years old) are known to behave prosocially

77   (Dunfield et al., 2011; Warneken & Tomasello, 2007; Vaish et al., 2009), such behaviour

78   appears to stem from an intrinsic motivation to satisfy a partner's needs rather from attempts

79   to strategically manage reputation. Children begin to show a concern for reputation from the

80   age of around five, for example by refraining from stealing from others if they are observed,

81   or making more generous or fairer donations to recipients when their generosity will be

82   revealed to others (Grueneisen & Tomasello, 2017; Leimgruber et al., 2012, McAuliffe et al.,

83   2020). Other work has shown that a concern with appearing to be prosocial or fair-minded

84   increases over childhood (Shaw et al. 2014), and that children become especially concerned

85   with self-presentation between the ages of 8 to 11 years old (Aloise-Young, 1993). At this age,

86   children are increasingly able to inhibit behaviours that might result in social sanctions

87   (Apfelbaum et al., 2008; Rutland et al., 2005) and attempt to present themselves in a positive

88   light to others. At the same time, children become increasingly skeptical about the intentions

89   of others, particularly when it comes to judging prosocial reputations (Heyman et al. 2014).

90   Thus, it takes most of childhood for humans to hone their ability to understand how one's

91   actions affect our reputations and to behave strategically so as to curate a positive reputation.

92

93   Unlike humans, there is scant evidence that non-humans primates attempt to strategically

94   manage their reputation. One recent study found that capuchin monkeys were insensitive to the

95   presence of an observer when deciding whether to share food (Schino et al., 2021), suggesting

96   that capuchins do not attempt to strategically manage their reputation in this way. Studies in

97   chimpanzees have also yielded null results. For instance, although chimpanzees increase effort

98   in a resource acquisition task when watched by a potential competitor, they do not increase

99   effort when watched by a potential cooperation partner (Engelmann et al., 2016). In the same

100   task, four to-five-year-old children increased their efforts both in the presence of a competitive

101   observer and in the presence of a potential future cooperation partner (Engelmann et al., 2016).

102   Similarly, although five-year old children share more and steal less when observed by a peer,

103 chimpanzees are not sensitive to the presence of an observer in the same paradigm (Engelmann,
104 et al., 2012, see also Leimgruber et al., 2012, see also Nettle et al., 2013).

105

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

119

120 **3. Cognitive mechanisms supporting reputation-based cooperation**
121 **3.1. Working memory**

122

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

133

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

137   (Manrique & Walker, 2017). Successfully managing one's own reputation might also involve

138   mental time travel, which allows individuals to imagine how events might unfold in the future.

139   This ability is also likely to involve working memory (Dere et al., 2019). Working memory is

140   also likely to be involved in evaluating the reputations of others, for example by tracking

141   cooperative behaviours (Milinski & Wedekind, 1998) and recalling what happened, with whom

142   and when ('episodic memory'). The complexity of such tasks can be increased further when

143   individuals compare observed behaviours against normative standards, or against behaviours

144   adopted by other individuals. The all-round utility of working memory poses some intriguing

145   questions for developmental and evolutionary psychology: at what age does children's working

146   memory become capable of maintaining reputation-based cooperative systems? Do great apes

147   have working memory complex enough to sustain reputation-based cooperative systems? By

148   what processes might these abilities have evolved in humans?

149

150   Working memory increases linearly between ages ~7 months and 14 years (Diamond & Doar,

151   1989; Gathercole et al., 2004, Read, 2008). Meta-analytic evidence (Read, 2008) suggests that

152   6-year-olds have a working memory size of three (compared to seven in adults: Miller, 1956).

153   Three is the minimum working memory size required to command relative clauses in sentences,

154   which are complex recursive structures like those used to tracking other people's perspectives

155   (e.g. John thinks that Mary knows he is supportive). Given that many reputational acts require

156   such recursion (e.g., John knows that if he doesn't help Mary now, she will not trust him to

157   reciprocate), it is reasonable to regard three as the minimum working memory size required for

158   constructing complex reputation-based cooperative systems. The extent of working memory

159   involvement in evaluation of others' reputations is likely to depend: evaluations that don't

160   involve recursion (e.g., helping that signals physical ability) may need less working memory

161   than those which do (e.g., helping that signals future intent to cooperate).

162

163   Studies directly measuring working memory in great apes are few and have yielded mixed

164   results. Some studies suggest that the working memory capacity of non-human apes is likely

165   to be limited. For instance, in a simplified version of the Wisconsin Card Sorting Test, that

166   involves sorting cards along three dimensions (shape, colour, number), chimpanzees struggled

167   to form a classificatory criterion or to change it flexibly to match the reinforcement

168   contingencies (Moriguchi et al., 2011). Similarly, in a memory task where individuals had to

169   turn over cards one at a time and find matching pairs, chimpanzees made four times more

170   mistakes than humans when tasked with three pairs, which would involve holding three cards

171 in working memory (Washburn et al., 2007). Nevertheless, other studies have reported
172 remarkable performance in serial ordering tasks administered to chimpanzees, that involved
173 memorizing up to 5 digits flashed on a screen in ascending order (Inoue & Matsuzawa, 2007),
174 or presenting up to 6 closed boxes on a platform and having a subject chimpanzee encode and
175 remember those boxes already emptied of food in previous trials to avoid re-opening them
176 again (Völter et al., 2019).

177

178 An alternative approach to assessing working memory capacity involves measuring the extent
179 to which individuals are able to hierarchically classify objects (Langer, 1980, 1986, 2000). The
180 Langer protocol investigates spontaneous grouping of objects and allows performance to be
181 rated as a function of complexity, ranging from first-order classifications, where only a single
182 group of objects matching in shape and/or colour is formed (e.g. is set apart from the other
183 objects), to classifications in which more than one group is formed contemporaneously (e.g.
184 rings are grouped together and kept apart from the cubes). Second- (and higher) order
185 classifications are assigned to groups of objects that are perceptually different, yet share the
186 same classificatory criteria. Second- (and higher) order classification impose higher working
187 memory demands on the classificatory rule as well as on the elements to be sorted, as their
188 differing features need to be compared simultaneously and flexibly (Langer, 1980, 1986, 2000).
189 Chimpanzees attain second-order combinativity around age 5 (Potì et al., 1999; Spinozzi et al.,
190 1999) when still they rarely compose more than two sets at a time (Langer, 2000, p.225). In
191 contrast, toddlers begin developing three-category classifications around age 3. Three-category
192 classification allows children to hierarchize – such as two subordinate classes within one
193 superordinate class – whereas two-category classification does not (Langer, 2000). This
194 hierarchization indicates that children develop recursive structures that might help them
195 tracking other people's perspectives and construct social reputation-based cooperative systems.

196

197 Other approaches have inferred working memory size based on the increasing complexity of
198 manufactured stone tools in the fossil record. Making and using simple stone flakes is reported
199 from Late Pliocene Africa 3,4 MYA, where bipedal Australopithecine existed from before 4
200 MYA. Australopithecines gave rise to the genus Homo, perhaps as early as 2,8 MYA, with
201 which they coexisted until after 2 MYA. By 2,5 MYA there are several Palaeolithic
202 assemblages of sharp conchoidal (i.e., shell-shaped) flakes struck by manual percussion with
203 hard hammer-stones. Conchoidal fracturing requires simultaneously focusing on the core
204 stone, the hammer stone, and the percussion angle, which implies a larger working memory

205 than that required for simple flakes (Read & Van Der Leeuw, 2008). Homo predominated by

206 1,76 MYA years ago, and co-occur in the African archaeological record with flattish stone

207 handaxes. These handaxes often resembled a large almond, were formed by manual percussion

208 with a hard hammer-stone that removed small conchoidal flakes in a regular manner (e.g.

209 bifacial stone-tool fashioning), from two surfaces of the handaxe to be. By 0.4-0.3 MYA,

210 handaxes had 3D symmetry, which required their makers to simultaneously remember different

211 perspectives of the core being worked on. To achieve ideal symmetry involves advanced

212 foresight and the ability to represent mentally the intended final product to exert on-going

213 corrections on the working substrate. Based on the increasing complexity of stone tools, and

214 the working memory required to make them, a reasonable conjecture is that early Homo had a

215 working memory greater than that of Australopithecines, which was in turn that of

216 chimpanzees. Taken together, these various lines of evidence suggest that working memory

217 capacity is likely to be higher in humans than in non-human apes (and specifically

218 chimpanzees).

219

220 Although working memory capacity has been relatively understudied in other animals

221 (Carruthers, 2013), there is some suggestive evidence for correlates of advanced working

222 memory in some species. For example, scrub jays display evidence of episodic-like memory,

223 being able to remember 'what', 'when' and 'where' during food caching events (Clayton &

224 Dickinson, 1998) as well as flexibly altering their own caching strategies to avoid being

225 parasitized by others (Correia et al., 2007). This example might provide the most compelling

226 evidence for sophisticated working memory in non- primates. As such, if they would benefit

227 from being able to choose partners for cooperative interactions, then they are a good species to

228 test for reputation-based cooperative systems.

229

230 **3.2 Delay of Gratification**

231

232 Any form of costly cooperation based on investments requires the ability to resist the

233 temptation to obtain immediate benefits (e.g. by cheating) in order to pursue a larger benefit in

234 the future. In some cases, this problem may be solved by psychological mechanisms which

235 render cooperative behaviour immediately subjectively rewarding (a phenomenon known as

236 warm glow, Andreoni, 1990). In other cases, individuals may have to effortfully resist an

237 immediately higher-paying option: they must be able to delay gratification.

238

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

245

246 Investing in a prosocial reputation might sometimes require the ability to delay gratification,
247 because the rewards for cooperation come from future (potentially unknown) partners instead
248 of one's current partner and are therefore inherently more likely to be delayed and less certain
249 to materialise. Several lines of evidence link the ability to delay gratification with cooperative
250 tendency in humans. Focusing on the future makes participants more generous (Sjåstad, 2019),
251 and spurs their willingness to incur personal costs to prevent damaging reputational
252 information from spreading (Vonasch et al., 2018). Children's ability to delay gratification is
253 positively related to their tendency to share, indicating that the ability to delay gratification
254 might be a prerequisite for children's sharing and cooperation (Sebastián-Enesco & Warneken,
255 2015). Similar patterns have been observed in adults (Curry et al., 2008; Harris & Madden,
256 2002; though see Barclay & Barker, 2020; Wu et al., 2017), as well as in blue jays who are
257 prevented from consuming rewards immediately (Stephens et al., 2002). Children are also
258 better at delaying gratification in cooperative tasks than solo tasks (Koomen et al., 2020). A
259 direct link between delay of gratification and reputational management has been suggested in
260 3- and 4-year-old children (Ma et al., 2020), although other work has shown that people are
261 unable to anticipate the delayed indirect benefits from their own cooperative investments (Wu
262 et al., 2016). To the extent that delay of gratification is involved in reputation-based
263 cooperation, we expect it to be more important in reputation management than in evaluating
264 the reputations of others (see Fig. 1).

265

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

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

281

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

293

294 **3.3 Theory of Mind**
295 Theory of mind is a multifaceted concept that refers to the ability to attribute mental states to
296 oneself and to third-parties and encompasses different abilities, which vary in computational
297 complexity. For example, taking another individual's visual perspective is simpler than
298 attributing intentions, which is in turn simpler than attributing knowledge, which is again
299 simpler than understanding complex perspectives (level 2 perspective-taking) or attributing
300 beliefs. These latter two examples of theory of mind are extremely taxing in terms of
301 computational demands, because they involve entertaining simultaneously alternative, often
302 contradictory, representations of reality (for a more detailed explanation, see Manrique &
303 Walker, 2017).

304

305 Here we introduce two theory of mind abilities that are likely to be involved in reputation
306 management and evaluating the reputation of others: perspective-taking and attribution of

307   intentions. Reputation-based cooperation may be more stable against erosion if bystanders or
308   other third parties can correctly attribute intentions and beliefs to actors, and if actors can
309   represent how they and their actions are perceived in the eyes of others. For example, an
310   individual may fail to cooperate either because (s)he does not realise that a recipient needs
311   help, or because (s)he currently lacks the resources to help. In other words, individuals with a
312   willingness to help may sometimes behave uncooperatively. If bystanders can correctly
313   identify uncooperative behaviour as a mistake or temporary inability, they can continue a
314   cooperative relationship with those who didn't intend to defect. Therefore, the reputation
315   system becomes less prone to errors undermining cooperation.

316

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

331

332   By contrast, reputation-based partner choice can function with or without theory of mind. In
333   reputation-based partner choice, actors help others to signal their ability and/or willingness to
334   help (Barclay, 2013). Theory of mind is not necessary to signal one's abilities or to interpret
335   such signals: when people see a good hunter share his kill, they can infer that (s)he is physically
336   skilled enough to catch it (e.g., Smith & Bliege Bird, 2000) without knowing anything of his
337   or her mental state. Hunters needn't know anything about the audience's mental state either –
338   they can learn that certain behaviours are rewarded (e.g., being chosen as a partner) via
339   reinforcement learning. However, theory of mind can greatly aid reputation-based partner
340   choice because it allows for more complex or targeted signals. For example, theory of mind

341 allows audiences to infer a helper's intentions in order to predict future cooperation and thus
342 allows individuals to signal not just their ability but their willingness to help. Therefore,
343 although simple forms of reputation-based partner choice might be achieved without the
344 advanced socio-cognitive mechanisms we discuss in this paper, we note that reputation-based
345 partner choice can later evolve to become cognitively quite complex, particularly when helpful
346 individuals have an incentive to misrepresent their type to others and when receivers take
347 hidden intentions of partners into consideration when evaluating prosocial acts (see Raihani &
348 Power, 2021 for a detailed discussion).
349
350 **3.3.1 Perspective-taking**
351
352 Perspective-taking can be broadly described as the ability to adopt the perspective of others
353 (e.g. visual, informational, emotional). At around two years of age, children are able to
354 differentiate what people can or cannot see (Moll & Tomasello, 2006). However, it is usually
355 not until three to four years of age that children understand that the same item can look different
356 from different perspectives (Moll & Meltzoff, 2011). This ability (level 2 perspective-taking)
357 requires effortful control to suppress the child's own visual perception, and is often viewed as
358 the precursor to full-blown theory of mind, in which the individual gains the ability to
359 understand others' knowledge and beliefs.
360
361 Perspective-taking is likely to be involved in both reputation management and the evaluation
362 of others' reputations. Reputation management involves not only behaving in a certain way,
363 but also the ability to shift perspectives to represent how complying or failing to act in this
364 manner will be perceived by others (Fig. 1). Thus, taking others' perspectives can make an
365 organism much more effective at reputation management. Similarly, perspective-taking makes
366 an organism better at detecting cheaters: organisms may dishonestly present themselves as
367 cooperative, and it requires cognitive effort for observers to distinguish between genuine versus
368 deceptive cooperators. For example, one individual might normally be a "cheater", but might
369 temporarily act cooperatively when (s)he sees someone (s)he wants to deceive or impress (e.g.,
370 a potential mate). Detecting dishonesty involves being able to entertain simultaneously
371 differing views of reality, an ability that can be equated in terms of computational complexity
372 to attributing complex (level 2) visual perspective. Hence, even if perspective-taking is not
373 strictly required to evaluate other's reputation, managing level 2 visual perspective-taking
374 indicates that organisms have the cognitive potential to entertain simultaneously

375   differing/contrasting views of reality (mine vs yours), and hence the ability to representing

376   simultaneously overt and hidden intentions in other's actions.

377

378   Perspective taking covers a wide spectrum of abilities, from knowing what others can or cannot

379   see ('level 1') to understanding that others see something differently as a function of their

380   relative position ('level 2') (Flavell, 1977; Flavell et al., 1981) and is therefore a good proxy

381   of other mentalising skills. Level 1 perspective-taking has been extensively investigated in

382   chimpanzees with initially diverging results (Povinelli & Eddy, 1996; Hare et al., 2000). Karg

383   et al. (2015) used a variation of the experience projection paradigm (Heyes, 1998) where

384   chimpanzees were trained with different pairs of goggles that affected what they could see.

385   When wearing one colour, the apes could see through the goggles but when wearing the other

386   colour they could not see anything. It could be inferred that chimpanzees are able to shift

387   perspectives if their own experience with the goggles (i.e., seeing vs. not seeing) affected their

388   response to human experimenters wearing the goggles. However, in this study, chimpanzees'

389   gaze-following was not influenced by their own previous experience with the googles (Karg et

390   al. 2015). Subsequent results indicated that chimpanzees may be able to shift perspectives in a

391   competitive context yet correct visual perspective attribution only approached a modest 60%

392   (Karg et al. 2015; but see Okamoto-Barth et al., 2007 for more positive findings).

393   Demonstrating level 2 visual perspective-taking in chimpanzees still proves elusive (Karg et

394   al., 2016).

395

396   Outside apes, the basic forms of perspective taking have currently only been found in large-

397   brained species. For example, rhesus monkeys steal more often from a human competitor

398   whose face is hidden by an opaque barrier than a competitor whose body alone is hidden

399   (Flombaum & Santos, 2005). Capuchin monkeys can also strategically conceal visual

400   information (Flombaum & Santos, 2005), while macaques have been reported to know what

401   others can or cannot hear (Santos et al., 2006). Ravens provide the best evidence for perspective

402   taking in birds, being able to follow human gaze direction around obstacles (Bugnyar et al.,

403   2004) and attributing visual perspectives even to unseen competitors (Bugnyar et al., 2016).

404   Most recently, however, there is evidence that cleaner fish *Labroides dimidiatus* females are

405   able to choose foraging sites where their male partners cannot observe them (McAuliffe et al.,

406   in review). Altogether, it appears that some other species may have some perspective-taking

407   abilities which can aid reputation-based cooperation, but perhaps not to the same level as

408   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.

443

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

477 age of three, children can infer norms by observing others acting in a certain way without
478 needing adult directives. At the same time, they also start enforcing norms on others (Vaish et
479 al., 2011). By around five years of age, children reach another milestone of normative
480 development: the spontaneous creation of their own rules (Grueneisen & Tomasello, 2017).
481 Although cultural norms vary widely in their content and implementation, children all over the
482 world show similar abilities for understanding, following, and enforcing socially prescribed
483 behaviours (Miller, 2007). The ways in which children create and deal with norms suggests a
484 growing understanding that norms are mutual agreements which result in rights and obligations
485 for each individual involved. Interestingly, children's concern about their own reputation (and
486 attempts at actively managing it) seems to trail their normative development (Kelsey et al.,
487 2018; Engelmann et al., 2012), i.e., children's reputation management develops after their
488 ability to view norms as a mutually-agreed upon standards for collaborative interactions.
489
490 If normative development encompasses the ability to view norms as a set of standards for
491 interactions, then it can only originate in species where collaborative interactions are initiated
492 by joint agency. Given the lack of evidence for shared agency and intentionality in
493 chimpanzees, the existence of a social system based on collective norms and influenced by
494 reputation seems highly unlikely (Schmidt, & Rakoczy, 2019; Tomasello, 2019). Also, given
495 the sparse evidence for social norms in chimpanzees, it is unsurprising that there is little
496 evidence for norms in other species either. In both vervet monkeys and great tits, there is
497 evidence that migrating individuals may give up previously learned preferences and conform
498 to local arbitrary preferences (van de Waal et al., 2013; Aplin et al., 2015). If such conformity
499 did represent norm-following, then these species might theoretically be capable of cooperative
500 systems based on social norms. Without such norm-following, the evolution of reputation-
501 based cooperation is less likely or less efficient.
502

503 **4. Reputation-based cooperation in non-human species**
504

505 Although cognitive constraints may prevent many non-human species from displaying
506 complex forms of reputation-based cooperation (Izuma, 2012), they may have simpler forms
507 that are less cognitively demanding. In social species, individuals often interact in
508 communication networks, where bystanders may eavesdrop on interactions to extract valuable
509 information (McGregor, 2005). Therefore, acting in a communication network has three
510 potential payoff consequences: the payoff obtained from the current interaction, the effect of

511 one's own action on the partner's future behaviour towards self, and the effect of one's own
512 action on the future behaviour of any bystander that learns about the action. Interactions in a
513 communication network therefore allow individuals to identify potentially cooperative or
514 aggressive individuals in their social environment and to adjust their behaviour appropriately.
515 Moreover, the possibility for bystander responsiveness might incentivise individuals to adjust
516 their current behaviour when they are observed, a phenomenon known as 'audience effects'
517 (Matos & Schlupp, 2005). This concept shares features with reputation management in
518 humans.
519
520 While eavesdropping and audience effects are widespread among vertebrates and have even
521 been documented in invertebrates, convincing evidence exists primarily in competitive
522 contexts (McGregor, 2005). By contrast, in species other than our own, there is a paucity of
523 evidence demonstrating that individuals show a concern for gaining a prosocial reputation.
524 Various arguments can be made why signals are likely to be honest in a competitive context
525 (Arnott & Elwood, 2009; Johnstone & Bshary, 2004) but less reliable in a cooperative context
526 (Johnstone & Bshary, 2007; Barclay, 2013; André, 2010; Bebbington et al., 2017). In a
527 competitive context, individual aggressiveness is likely to be correlated with strength, which
528 is based on metastable features like size, muscle mass, agility and experience. Therefore,
529 signals of formability are difficult to fake and more likely to be honest. The honesty of such
530 signals can change the benefits associated with paying attention to them: eavesdropping in
531 order to gain information on a potential partner's formidability is potentially self-serving. In
532 return, strong individuals may benefit from signaling their strength to eavesdropping
533 bystanders, for example by displaying after a victorious fight, or attacking those lower in the
534 hierarchy after a defeat (Kazim & Aureli, 2005) in order to reduce the likelihood of being the
535 target of future challenges. Strong individuals may even pick a fight that yields a short-term
536 negative payoff to reduce the likelihood of being challenged by bystanders in the future
537 (Johnstone & Bshary, 2004).
538
539 Nevertheless, there are a handful of examples from non-human species that are suggestive of
540 reputation-based cooperation. In various species, individuals may temporarily act as a
541 watchman by looking out for predators while the rest of the group forages. While such
542 behaviour has been interpreted as immediately self-serving as it is mostly done by satiated
543 individuals (Clutton-Brock et al., 1999), experiments involving dwarf mongooses have shown
544 that playbacks of an individual's watchman calls increases the amount of grooming this

545 individual receives later in the day (Kern & Radford, 2018). In vervet monkeys, males and
546 females that contribute during territorial disputes receive more grooming by other group
547 members (Arseneau-Robar et al., 2016). In Arabian babblers and Siberian jays, males act more
548 aggressively towards predators in the presence of females, which is suggestive of males
549 displaying in the context of female mate choice (Zahavi, 1995; da Cunha et al., 2017). In all
550 these cases, there is no specific recipient of the initial helpful act, meaning that the source of
551 eventual return benefits is uncertain.

552

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

564

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

575

576 One obvious distinction between reputation-based cooperation in humans and other animals is
577 that humans use language (see other contributions to this theme issue). Language allows people
578 to flexibly exchange information about other individuals (Wu et al., 2016) – and can potentially

579 also increase the amount of information that can be exchanged. Language can also help humans

580 to represent (and hence encode) and recall social norms and might also be a pre-requisite for

581 expressing more complex aspects of social cognition that are likely to be involved in managing

582 and evaluating reputations. Despite its likely importance, we do not discuss language in this

583 review, because it acts more as a multiplier on other cognitive mechanisms, and we instead

584 focus on other proximate cognitive mechanisms that form the basic building blocks of

585 reputation-based cooperation in humans.

586

587 **5. Discussion**

588 We have presented four basic psychological building blocks that we consider important

589 facilitators for complex reputation-based cooperation: working memory, delay of gratification,

590 theory of mind, and social norms. Working memory allows for parallel processing of diverse

591 information, to properly assess others' actions and update their reputation scores. Delay of

592 gratification is useful for many types of cooperation, but may be particularly relevant for

593 reputation-based cooperation where the returns come from a future interaction with an observer

594 rather than an immediate reciprocation by one's current partner. Theory of mind makes it easier

595 to properly assess others' actions, and reduces the risk that spreading errors will undermine

596 cooperation. Finally, norms support theory of mind by giving individuals a benchmark of what

597 is right or wrong. The more developed that each of these building blocks is, the more complex

598 the interaction structure can become. We are aware that by picking these four socio-cognitive

599 mechanisms we leave out other processes that might be involved, e.g. long-term memory, yet

600 we think the ones we picked are more critical and better allow for comparison across species.

601

602 Reputation-based cooperation based on partner choice might often be less cognitively

603 demanding than that based on indirect reciprocity. On the one hand, reputation-based partner

604 choice might require a better ability to delay gratification (as it might take several acts of

605 investment to outcompete competitors and be chosen by third parties), while IR games are

606 typically set up in such a way that individuals alternate roles as helper and recipient. On the

607 other hand, reputation-based partner choice can exist in cognitively simple forms like "walk

608 away or reject partner if they seem uncooperative" (Aktipis, 2004; McNamara et al., 2008);

609 this does not require high working memory, theory of mind, or normative behaviour, though

610 these abilities can make reputation-based partner choice more efficient. In contrast, analyses

611 of indirect reciprocity games have shown that Kandori is the simplest strategy yielding stable

612 cooperation (Santos et al., 2018), and Kandori requires norms, theory of mind to identify errors,

613 and as a consequence more computational power (e.g. working memory). Therefore, the vast
614 majority of animal species may be cognitively constrained from implementing indirect
615 reciprocity, and hence be limited to simple forms of reputation-based partner choice. In line
616 with this hypothesis, the few non-human examples of reputation-based cooperation largely fit
617 the concept of reputation-based partner choice, not indirect reciprocity. Most of the examples
618 seem to be about one party gaining information about another, to know whom to cooperate or
619 mate with, or whom to avoid in fights – a type of reputation-based partner choice based on
620 eavesdropping (McGregor, 2005). As such, there is a clear evolutionary path for reputation-
621 based partner choice: start with cognitively simple eavesdropping, which then evolves into an
622 active signalling system (see Biernaskie et al., 2018 for cues evolving into signals), with more
623 complex abilities arising later in both signallers and receivers in order to perform better within
624 that signalling system.
625
626 Future work should further clarify the role of these cognitive mechanisms in reputation-based
627 cooperation in both humans and non-humans. Studies could investigate reputation-based
628 cooperation in humans when these cognitive mechanisms cannot function properly, such as
629 experimental paradigms that increase cognitive load (e.g., Milinski & Wedekind, 1998), special
630 populations that lack some of these cognitive mechanisms (e.g., Cage et al., 2013; Izuma et al.,
631 2011), or online networks where one cannot use these mechanisms. Non-human studies could
632 artificially grant these abilities to non-humans, for example by dissociating cooperative
633 investments from ability to delay gratification (c.f. Stephens et al., 2002). Other studies could
634 use other creative ways of outsourcing cognition to see how they affect reputation-based
635 cooperation. We look forward to seeing further tests of the cognitive building blocks of
636 reputation-based cooperation.
637
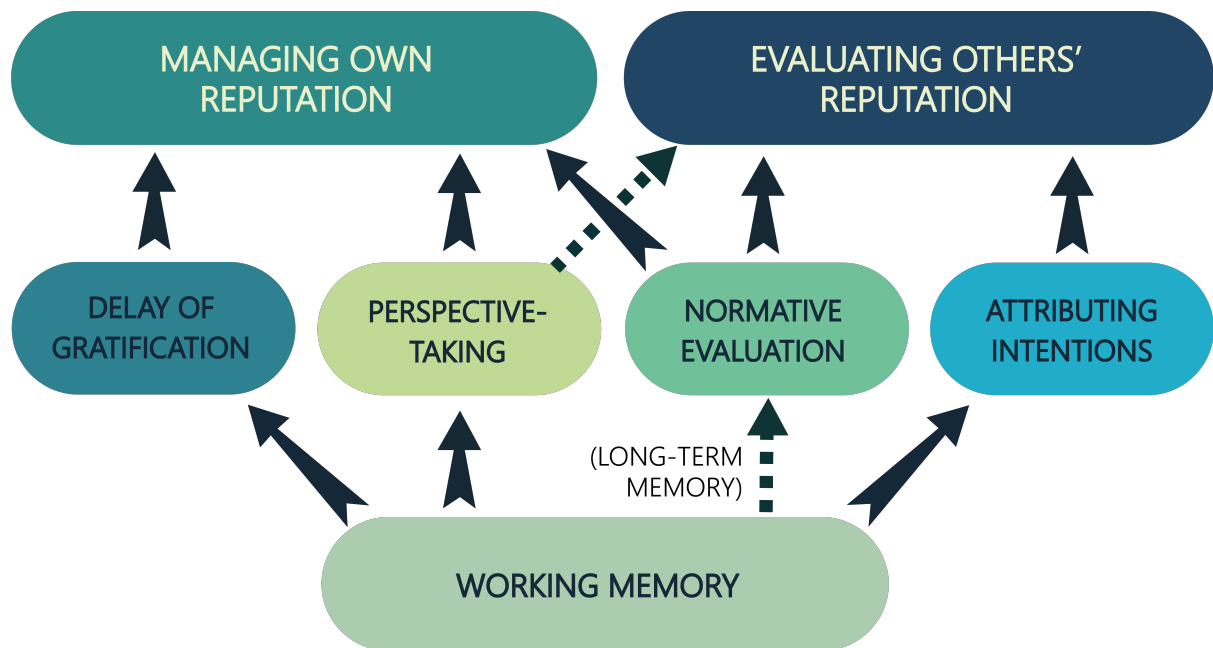
Figure 1. Depiction of how our four socio-cognitive mechanisms are recruited for the managing of one's own reputation as opposed to evaluating third-party reputations. No connecting lines indicate there is no need for the socio-cognitive mechanism in question to be recruited. Arrow continuity expresses 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

661

662 Aellen, M., Dufour, V., & Bshary, R. (2021). Cleaner fish and other wrasse match primates

663 in their ability to delay gratification. *Animal Behaviour, 176*, 125-143.

664 https://doi.org/10.1016/j.anbehav.2021.04.002

665

666 Aktipis, C. A. (2004). Know when to walk away: contingent movement and the evolution of

667 cooperation. *Journal of theoretical biology, 231*(2), 249-260.

668 http://dx.doi.org/10.1016/j.jtbi.2004.06.020

669

670 André, J.-B. (2010). The evolution of reciprocity: social types or social incentives? *The*

671 *American Naturalist, 175*(2), 197-210. http://dx.doi.org/10.1086/649597

672

673 Andreoni, J. (1990). Impure altruism and donations to public goods: A theory of warm-glow

674 giving. *The economic journal, 100*(401), 464-477. https://doi.org/10.2307/2234133

675

676 Aloise-Young, P. A. (1993). The development of self-presentation: Self-promotion in 6-to

677 10-year-old children. *Social Cognition, 11*(2), 201-222.

678 http://dx.doi.org/10.1521/soco.1993.11.2.201

679

680 Apfelbaum, E. P., Sommers, S. R., & Norton, M. I. (2008). Seeing race and seeming racist?

681 Evaluating strategic colorblindness in social interaction. *Journal of personality and social*

682 *psychology, 95*(4), 918-932. https://doi.org/10.1037/a0011990

683

684 Aplin, L. M., Farine, D. R., Morand-Ferron, J., Cockburn, A., Thornton, A., & Sheldon, B. C.

685 (2015). Experimentally induced innovations lead to persistent culture via conformity in wild

686 birds. *Nature, 518*(7540), 538-541. http://dx.doi.org/10.1038/nature13998

687

688 Arnott, G., & Elwood, R. W. (2009). Assessment of fighting ability in animal contests.

689 *Animal Behaviour, 77*(5), 991-1004. http://dx.doi.org/10.1016/j.anbehav.2009.02.010

690

691 Arseneau-Robar, T. J. M., Taucher, A. L., Müller, E., van Schaik, C., Bshary, R., & Willems,

692 E. P. (2016). Female monkeys use both the carrot and the stick to promote male participation

693 in intergroup fights. *Proceedings of the Royal Society B: Biological Sciences, 283*(1843),

694 20161817. http://dx.doi.org/10.1098/rspb.2016.1817

695

696   Barclay, P. (2013). Strategies for cooperation in biological markets, especially for humans.

697   *Evolution & Human Behavior*, *34*(3), 164-175.

698   http://dx.doi.org/10.1016/j.evolhumbehav.2013.02.002

699

700   Barclay, P. (2015). Reputation. In D. Buss (Ed.) *Handbook of Evolutionary Psychology (2nd*

701   *Ed.)*, pp. 810-828. Hoboken, NJ: J. Wiley & Sons.

702   http://dx.doi.org/10.1002/9781119125563.evpsych233

703

704   Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and

705   friendship. *Current Opinion in Psychology*, *7*, 33-38.

706   http://dx.doi.org/10.1016/j.copsyc.2015.07.012

707

708   Barclay, P., & Barker, J.L. (2020). Greener than thou: people who protect the environment

709   are more cooperative, compete to be environmental, and benefit from reputation. *Journal of*

710   *Environmental Psychology*, *72*, 101441. http://dx.doi.org/10.1016/j.jenvp.2020.101441

711

712   Barrett, H.C., Bolyanatz, A., Crittenden, A.N., Fessler, D.M.T., Fitzpatrick, S., Gurven, M.,

713   Henrich, J., Kanovsky, M., Kushnick, G., Pisor, A., Scelza, B.A., Stich, S., von Rueden, C.,

714   Zhao, W., & Laurence, S. (2016). Small-scale societies exhibit fundamental variation in the

715   role of intentions in moral judgement. *Proceedings of the New York Academy of Sciences*,

716   *113*(17), 4688-4693. http://dx.doi.org/10.1073/pnas.1522070113

717

718   Baumeister, R. F., Maranges, H. M., & Sjåstad, H. (2018). Consciousness of the future as a

719   matrix of maybe: Pragmatic prospection and the simulation of alternative possibilities.

720   *Psychology of Consciousness: Theory, Research, and Practice*, *5*(3), 223–238.

721

722   Bebbington, K., MacLeod, C., Ellison, T. M., & Fay, N. (2017). The sky is falling: evidence

723   of a negativity bias in the social transmission of information. *Evolution and Human Behavior*,

724   *38*(1), 92-101. http://dx.doi.org/10.1016/j.evolhumbehav.2016.07.004

725

726   Behne, T., Carpenter, M., Call, J., & Tomasello, M. (2005). Unwilling versus unable: infants'

727   understanding of intentional action. *Developmental psychology*, *41*(2), 328-337.

728   http://dx.doi.org/10.1037/0012-1649.41.2.328

729

730 Beran, M. J., & Evans, T. A. (2006). Maintenance of delay of gratification by four

731 chimpanzees (Pan troglodytes): The effects of delayed reward visibility, experimenter

732 presence, and extended delay intervals. *Behavioural Processes*, *73*(3), 315-324.

733 http://dx.doi.org/10.1016/j.beproc.2006.07.005

734

735

736 Biernaskie, J. M., Perry, J. C., & Grafen, A. (2018). A general model of biological signals,

737 from cues to handicaps. *Evolution Letters, 2*(3), 201-209. https://doi.org/10.1002/evl3.57

738

739 Bradley, A., Lawrence, C., & Ferguson, E. (2018). Does observability affect prosociality?

740 *Proceedings of the Royal Society of London B*, *285*, 20180116.

741 http://dx.doi.org/10.1098/rspb.2018.0116

742

743 Bshary, R. (2002). Biting cleaner fish use altruism to deceive image–scoring client reef fish.

744 *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *269*(1505), 2087-

745 2093. http://dx.doi.org/10.1098/rspb.2002.2084

746

747 Bshary, R., & Grutter, A. S. (2006). Image scoring and cooperation in a cleaner fish

748 mutualism. *Nature*, *441*(7096), 975-978. http://dx.doi.org/10.1038/nature04755

749

750 Bueno-Guerra, N., Colell, M., & Call, J. (2020). Effects of indirect reputation and type of

751 rearing on food choices in chimpanzees (Pan troglodytes). *Behavioral Ecology and*

752 *Sociobiology, 74*(6), 1-10. https://doi.org/10.1007/s00265-020-02861-w

753

754 Bugnyar, T., Reber, S. A., & Buckner, C. (2016). Ravens attribute visual access to unseen

755 competitors. *Nature Communications*, *7*(1), 1-6. http://dx.doi.org/10.1038/ncomms10506

756

757 Bugnyar, T., Stöwe, M., & Heinrich, B. (2004). Ravens, Corvus corax, follow gaze direction

758 of humans around obstacles. *Proceedings of the Royal Society of London. Series B:*

759 *Biological Sciences*, *271*(1546), 1331-1336. http://dx.doi.org/10.1098/rspb.2004.2738

760

761 Cage, E., Pellicano, E., Shah, P., & Bird, G. (2013). Reputation management: evidence for
762 ability but reduced propensity in autism. *Autism Research, 6*, 433-442.
763 http://dx.doi.org/10.1002/aur.1313
764
765 Call, J., Hare, B., Carpenter, M., & Tomasello, M. (2004). `Unwilling' versus `unable':
766 chimpanzees' understanding of human intentional action. *Developmental Science,*
767 *7*(4), 488-498. http://dx.doi.org/10.1111/j.1467-7687.2004.00368.x
768
769 Call, J., & Tomasello, M. (1998). Distinguishing intentional from accidental actions in
770 orangutans (Pongo pygmaeus), chimpanzees (Pan troglodytes) and human children (Homo
771 sapiens). *Journal of Comparative Psychology, 112*(2), 192-206.
772 http://dx.doi.org/10.1037/0735-7036.112.2.192
773
774 Canteloup, C., Piraux, E., Poulin, N., & Meunier, H. (2016). Do Tonkean macaques
775 (Macaca tonkeana) perceive what conspecifics do and do not see? *PeerJ,* 4:e1693
776 http://dx.doi.org/10.7717/peerj.1693
777
778 Carruthers, P., (2013). Evolution of working memory. *Proceedings of the National Academy*
779 *of Sciences of the USA, 110*(Supplement 2), 10371-10378.
780 http://dx.doi.org/10.1073/pnas.1301195110
781
782 Clutton-Brock, T. H., O'riain, M. J., Brotherton, P. N., Gaynor, D., Kansky, R., Griffin, A. S.,
783 & Manser, M. (1999). Selfish sentinels in cooperative mammals. *Science, 284*(5420), 1640-
784 1644. http://dx.doi.org/10.1126/science.284.5420.1640
785
786 Clayton, N. S., & Dickinson, A. (1998). Episodic-like memory during cache recovery by
787 scrub jays. *Nature, 395*(6699), 272-274. http://dx.doi.org/10.1038/26216
788
789 Coricelli, G., Joffily, M., Montmarquette, C., & Villeval, M.C. (2010). Cheating, emotions,
790 and rationality: an experiment on tax evasion. *Experimental Economics, 13*(2), 226-247.
791
792 Correia, S. P., Dickinson, A., & Clayton, N. S. (2007). Western scrub-jays anticipate future
793 needs independently of their current motivational state. *Current Biology, 17*(10), 856-861.
794 http://dx.doi.org/10.1016/j.cub.2007.03.063

795

796 Costes-Thiré, M., Levé, M., Uhlrich, P., Pasquaretta, C., De Marco, A., & Thierry, B. (2015).

797 Evidence that monkeys (Macaca tonkeana and Sapajus apella) read moves, but no evidence

798 that they read goals. *Journal of Comparative Psychology*, *129*(3), 304-310.

799 http://dx.doi.org/10.1037/a0039294

800

801 Côté, I. M. (2000). Evolution and ecology of cleaning symbioses in the sea. *Oceanography*

802 *and Marine Biology*, *38*, 311–355. Taylor & Francis, CRS Press, United Kingdom.

803

804 da Cunha, F. C. R., Fontenelle, J. C. R., & Griesser, M. (2017). The presence of conspecific

805 females influences male-mobbing behavior. *Behavioral Ecology and Sociobiology*, *71*(3), 52.

806 http://dx.doi.org/10.1007/s00265-017-2267-7

807

808 Curry, O.S., Price, M.E., & Price, J.G. (2008). Patience is a virtue: cooperative people have

809 lower discount rates. *Personality and Individual Differences*, *44*(3), 778-783.

810 http://dx.doi.org/10.1016/j.paid.2007.09.023

811

812 Dere, D., Zlomuzica, A., Dere, E. (2019). Fellow travellers in cognitive evolution: Co-

813 evolution of working memory and mental time travel? *Neuroscience & Biobehavioral*

814 *Reviews*, *105*, 94-105. https://doi.org/10.1016/j.neubiorev.2019.07.016.

815

816 Diamond, A., & Doar, B. (1989). The performance of human infants on a measure of frontal

817 cortex function, the delayed-response task. *Developmental Psychobiology*, *22*(3), 271-294.

818 http://dx.doi.org/10.1002/dev.420220307

819

820 Dunfield, K. A., & Kuhlmeier, V. A. (2010). Intention-mediated selective helping in infancy.

821 *Psychological science*, *21*(4), 523-527. http://dx.doi.org/10.1177/0956797610364119

822

823 Dunfield, K., Kuhlmeier, V. A., O'Connell, L., & Kelley, E. (2011). Examining the diversity

824 of prosocial behavior: Helping, sharing, and comforting in infancy. *Infancy*, *16*(3), 227-247.

825

826 Engelmann, J. M., Herrmann, E., & Tomasello, M. (2012). Five-year olds, but not

827 chimpanzees, attempt to manage their reputations. *PLoS On*e 7:e48433.

828 http://dx.doi.org/10.1371/journal.pone.0048433

829

830 Engelmann, J. M., Herrmann, E., & Tomasello, M. (2016). The effects of being watched on

831 resource acquisition in chimpanzees and human children. *Animal Cognition, 19*(1), 147-151.

832 http://dx.doi.org/10.1007/s10071-015-0920-y

833

834 Evans, T. A., & Beran, M. J. (2007). Chimpanzees use self-distraction to cope with

835 impulsivity. *Biology Letters*, 3, 599-602. http://dx.doi.org/10.1098/rsbl.2007.0399

836

837 Flavell, J. H. (1977). The development of knowledge about visual perception. Nebraska

838 Symposium on Motivation, 25, 43–76.

839

840 Flavell, J. H., Everett, B. A., Croft, K., & Flavell, E. R. (1981). Young children's knowledge

841 about visual perception: Further evidence for the Level 1–Level 2 distinction. *Developmental*

842 *Psychology*, *17*(1), 99-103. http://dx.doi.org/10.1037/0012-1649.17.1.99

843

844 Flombaum, J. & Santos, L. (2005). Rhesus monkeys can assess the visual perspective of

845 others when competing for food. *Current Biology*, *15*(5), 447-452.

846 http://dx.doi.org/10.1016/j.cub.2004.12.076

847

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

849 319-333.

850

851 Gathercole, S.E., S.J. Pickering, B. Ambridge, B. & Wearing, H. (2004). The structure of

852 working memory from 4 to 15 years of age. *Development Psychology*, *40*(2), 177-190.

853 http://dx. doi.org/10.1037/0012-1649.40.2.177

854

855 Gerber, A.A., Green, D.P., & Larimer, C.W. (2008). Social pressure and voter turnout:

856 evidence from a large-scale field experiment. *American Political Science Review*, *102*(1), 33-

857 48. http://dx.doi.org/10.1017/S000305540808009X

858

859 Gergely, G., Bekkering, H., & Király, I. (2002). Rational imitation in preverbal infants.

860 *Nature*, *415*(6873), 755. http://dx.doi.org/10.1038/415755a

861

862 Grueneisen, S. & Tomasello, M. (2017). Children coordinate in a recurrent social dilemma by

863 taking turns and along dominance asymmetries. *Developmental Psychology*, *53*(2), 265-273.

864 http://dx.doi.org/10.1037/dev0000236

865

866 Grutter, A. S. (1995). The relationship between cleaning rates and ectoparasites loads in coral

867 reef fishes. *Marine Ecology Progress Series, 118*, 51–58.

868 http://dx.doi.org/10.3354/meps118051

869

870 Grutter, A.S., & Bshary, R. (2003). Cleaner wrasse prefer client mucus: support for partner

871 control mechanisms in cleaning interactions. *Proceedings of the Royal Society of London.*

872 *Series B: Biological Sciences*, *270* (Supplement 2), S242–S244.

873 http://dx.doi.org/10.1098/rsbl.2003.0077

874

875 Hare, B., Call, J., Agnetta, B., & Tomasello, M. (2000). Chimpanzees know what

876 conspecifics do and do not see. *Animal Behaviour*, *59*(4), 771-785.

877 http://dx.doi.org/10.1006/anbe.1999.1377

878

879 Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-control in decision-making involves

880 modulation of the vmPFC valuation system. *Science*, *324*(5927), 646–648.

881 http://dx.doi.org/10.1126/science.1168450

882

883 Harris, A.C., & Madden, G.J. (2002). Delay discounting and performance on the Prisoner's

884 Dilemma Game. *Psychological Record*, *52*(4), 429-440.

885 http://dx.doi.org/10.1007/BF03395196

886

887 Herrmann, E., Keupp, S., Hare, B., Vaish, A., & Tomasello, M. (2013). Direct and indirect

888 reputation formation in nonhuman great apes (Pan paniscus, Pan troglodytes, Gorilla gorilla,

889 Pongo pygmaeus) and human children (Homo sapiens). *Journal of Comparative Psychology,*

890 *127*(1), 63–75. https://doi.org/10.1037/a0028929

891

892 Heyes, C.M. (1998). Theory of mind in nonhuman primates. *Behavioral and Brain Sciences,*

893 *21*(1), 101-148. http://dx.doi.org/10.1017/S0140525X98000703

894

895   Heyman, G., Barner, D., Heumann, J., & Schenck, L. (2014). Children's sensitivity to ulterior

896   motives when evaluating prosocial behavior. *Cognitive science*, *38*(4), 683-700.

897   http://dx.doi.org/10.1111/cogs.12089

898

899   Inoue, S., & Matsuzawa, T. (2007). Working memory of numerals in chimpanzees. *Current*

900   *Biology*, *17*(23), R1004-R1005. http://dx.doi.org/10.1016/j.cub.2007.10.027

901

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

903   288.

904

905   Izuma, K., Matsumoto, K., Camerer, C.F., & Adolphs, R. (2011). Insensitivity to social

906   reputation in autism. *Proceedings of the National Academy of Sciences*, *108(42)*, 17302-

907   17307. http://dx.doi.org/10.1073/pnas.1107038108

908

909   Johnstone, R. A., & Bshary, R. (2004). Evolution of spite through indirect reciprocity.

910   *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *271*(1551), 1917-

911   1922. http://dx.doi.org/10.1098/rspb.2003.2581

912

913   Johnstone, R. A., & Bshary, R. (2007). Indirect reciprocity in asymmetric interactions: when

914   apparent altruism facilitates profitable exploitation. *Proceedings of the Royal Society B:*

915   *Biological Sciences*, *274*(1629), 3175-3181. http://dx.doi.org/10.1098/rspb.2003.2581

916

917   Kandori, M. (1992). Social norms and community enforcement. *The Review of Economic*

918   *Studies*, *59*(1), 63-80. http://dx.doi.org/10.2307/2297925

919

920   Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2015). The goggles experiment: can

921   chimpanzees use self-experience to infer what a competitor can see? *Animal Behaviour*, *105*,

922   211-221. http://dx.doi.org/10.1016/j.anbehav.2015.04.028

923

924   Karg, K., Schmelz, M., Call, J., & Tomasello, M. (2016). Differing views: Can chimpanzees

925   do Level 2 perspective-taking? *Animal Cognition*, *19*, 555-564.

926   http://dx.doi.org/10.1007/s10071-016-0956-7

927

928   Kawai, N., Nakagami, A., Yasue, M., Koda, H., & Ichinohe, N. (2019). Common marmosets

929   (Callithrix jacchus) evaluate third-party social interactions of human actors but Japanese

930   monkeys (Macaca fuscata) do not. *Journal of Comparative Psychology, 133*(4), 488–495.

931   https://doi.org/10.1037/com0000182

932

933   Kazim, A. J. N., & Aureli, F. (2005). Redirection of aggression: multiparty signalling within

934   a network? In P. McGregor (Ed) Animal Communication Networks, pp. 191-218.

935   Cambridge, UK: Cambridge University Press.

936   http://dx.doi.org/10.1017/CBO9780511610363.013

937

938   Kelsey, C., Grossmann, T., & Vaish, A. (2018). Early reputation management: Three-year-

939   old children are more generous following exposure to eyes. *Frontiers in psychology*, *9*, 698.

940   http://dx.doi.org/10.3389/fpsyg.2018.00698

941

942   Kern, J. M., & Radford, A. N. (2018). Experimental evidence for delayed contingent

943   cooperation among wild dwarf mongooses. *Proceedings of the National Academy of*

944   *Sciences*, *115*(24), 6255-6260. http://dx.doi.org/10.1073/pnas.1801000115

945

946   Koomen, R., Grueneisen, S., & Herrmann, E. (2020). Children delay gratification for

947   cooperative ends. *Psychological science*, *31*(2), 139-148.

948   http://dx.doi.org/10.1177/0956797619894205

949

950   Lacetera, N., & Macis, M. (2010). Social image concerns and prosocial behavior: field

951   evidence from a nonlinear incentive scheme. *Journal of Economic Behavior & Organization*,

952   *76*(2), 225-237.

953

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

955

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

957

958   Langer, J. (2000). The heterochronic evolution of Primate cognitive development. In: Parker,

959   S.T., Langer, J., McKinney, M.L.(Eds.), Biology, Brains and Behavior. The Evolution of

960   Human Development. Santa Fe, American School of Research Press, "Advanced Seminar

961   Series", and Oxford, James Currey, pp. 215-235. http://dx.doi.org/10.1162/biot.2006.1.1.41

962

963 Leimgruber, K. L., Shaw, A., Santos, L. R., & Olson, K. R. (2012). Young children are more

964 generous when others are aware of their actions. *PloS one, 7*(10), e48292. http://dx.

965 doi.org/10.1371/journal.pone.0048292

966

967 Luerssen, A., Gyurak, A., Ayduk, O., Wendelken, C., & Bunge, S. A. (2015). Delay of

968 gratification in childhood linked to cortical interactions with the nucleus accumbens. *Social*

969 *cognitive and affective neuroscience, 10*(12), 1769-1776.

970 http://dx.doi.org/10.1093/scan/nsv068

971

972 Ma, F., Zeng, D., Xu, F., Compton, B. J., & Heyman, G. D. (2020). Delay of Gratification as

973 Reputation Management. *Psychological Science, 31*(9), 1174-1182.

974 http://dx.doi.org/10.1177/0956797620939940

975

976 Manrique, H. M., Walker, M. J. (2017). *Early Evolution of Human Memory*. London:

977 Palgrave Macmillan. http://dx.doi.org/10.1007/978-3-319-64447-9

978

979 Matos, R., & Schlupp, I. (2005). Performing in front of an audience: Signallers and the social

980 environment. In P. McGregor (Ed.), *Animal Communication Networks* (pp. 63-83).

981 Cambridge: Cambridge University Press. http://dx.doi.org/10.1017/CBO9780511610363

982

983 McAuliffe, K., Blake, P. R., & Warneken, F. (2020). Costly fairness in children is influenced

984 by who is watching. *Developmental Psychology, 56*(4), 773–782.

985 https://doi.org/10.1037/dev0000888

986

987 McAuliffe, K., Drayton, L.A., Royka, A., Aellen, M., Santos, L.R. & Bshary, R. (submitted)

988 Do cleaner fish know what others can and cannot see?

989

990 McGregor, P. (Ed.). (2005). *Animal Communication Networks*. Cambridge University Press.

991 http://dx.doi.org/10.1017/CBO9780511610363

992

993 McNamara, J. M., Barta, Z., Frohmage, L., & Houston, A. I. (2008). The coevolution of

994 choosiness and cooperation. *Nature, 451*(7175), 189-192.

995 http://dx.doi.org/10.1038/nature06455

996

997   Meltzoff, A. N. (1995). Understanding the intentions of others: re-enactment of intended acts

998   by 18-month-old children. *Developmental psychology*, *31*(5), 838-850.

999   http://dx.doi.org/10.1037/0012-1649.31.5.838

1000

1001   Milinski, M., & Wedekind, C. (1998). Working memory constrains human cooperation in the

1002   Prisoner's Dilemma. *Proceedings of the National Academy of Sciences*, *95*(23), 13755-

1003   13758.

1004   http://dx.doi.org/10.1073/pnas.95.23.13755

1005

1006   Milinski, M., Semmann, D., Bakker, T. C., & Krambeck, H. J. (2001). Cooperation through

1007   indirect reciprocity: image scoring or standing strategy? *Proceedings of the Royal Society of*

1008   *London. Series B: Biological Sciences*, *268*(1484), 2495-2501.

1009   http://dx.doi.org/10.1098/rspb.2001.1809

1010

1011   Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our

1012   capacity for processing information. *Psychological review*, *63*(2), 81-97.

1013   http://dx.doi.org/10.1037/h0043158

1014

1015   Miller, J. G. (2007). Cultural psychology of moral development. In S. Kitayama & D. Cohen

1016   (Eds.), Handbook of cultural psychology (p. 477–499). The Guilford Press.

1017

1018   Miller, R., Boeckle, M., Jelbert, S. A., Frohnwieser, A., Wascher, C. A., & Clayton, N. S.

1019   (2019). Self-control in crows, parrots and nonhuman primates. *Wiley Interdisciplinary*

1020   *Reviews: Cognitive Science*, *10*(6), e1504. http://dx.doi.org/10.1002/wcs.1504

1021

1022   Mischel, W., & Ebbesen, E.B. (1970). Attention in delay of gratification. *Journal of*

1023   *Personality and Social Psychology*, 16, 329–337. http://dx.doi.org/10.1037/h0029815

1024

1025   Mischel, H. N., & Mischel, W. (1983). The development of children's knowledge of self-

1026   control strategies. *Child Development*, *54*(3), 603-619. http://dx.doi.org/10.2307/1130047

1027

1028 Moll, H., & Meltzoff, A.N. (2011). How does it look? Level 2 perspective-taking at 36

1029 months of age. *Child development*, 82(2), 661-673. http://dx.doi.org/10.1111/j.1467-

1030 8624.2010.01571.x

1031

1032 Moll, H., & Tomasello, M. (2006). Level 1 perspective-taking at 24 months of age. *British*

1033 *Journal of Developmental Psychology*, 24(3), 603-613.

1034 http://dx.doi.org/10.1348/026151005X55370

1035

1036 Moriguchi, Y., Tanaka, M., & Itakura, S. (2011). Executive function in young children and

1037 chimpanzees (Pan troglodytes): evidence from a non-verbal dimensional change card sort

1038 task. *Journal of Genetic Psychology, 172*(3), 252-265.

1039 http://dx.doi.org/10.1080/00221325.2010.534828

1040

1041 Nettle, D., Cronin, K. A., & Bateson, M. (2013). Responses of chimpanzees to cues of

1042 conspecific observation. *Animal Behaviour, 86*(3), 595-602.

1043 https://doi.org/10.1016/j.anbehav.2013.06.015

1044

1045 Ohtsuki, H. & Iwasa, Y. (2007). Global analyses of evolutionary dynamics and exhaustive

1046 search for social norms that maintain cooperation by reputation. *Journal of Theoretical*

1047 *Biology*, 24(3), 518–531. http://dx.doi.org/10.1016/j.jtbi.2006.08.01

1048

1049 Okamoto-Barth, S., Call, J., & Tomasello, M. (2007). Great apes' understanding of other

1050 individuals' line of sight. *Psychological Science*, 18(5), 462-468.

1051 http://dx.doi.org/10.1111/j.1467-9280.2007.01922.x

1052

1053 Péron, F., Rat-Fischer, L., Nagle, L., & Bovet, D. (2010). 'Unwilling'versus 'unable': do

1054 grey parrots understand human intentional actions?. *Interaction studies*, 11(3), 428-441.

1055 http://dx.doi.org/10.1075/is.11.3.06per

1056

1057 Phillips, W., Barnes, J. L., Mahajan, N., Yamaguchi, M., & Santos, L. R. (2009).

1058 'Unwilling'versus 'unable': capuchin monkeys'(Cebus apella) understanding of human

1059 intentional action. *Developmental science*, 12(6), 938-945. http://dx.doi.org/10.1111/j.1467-

1060 7687.2009.00840.x

1061

1062 Pinto, A.I., Oates, J., Grutter, A.S., & Bshary, R. (2011). Cleaner wrasse Labroides

1063 dimidiatus are more cooperative in the presence of an audience. *Current Biology, 21(13),*

1064 1140–1144. http://dx.doi.org/10.1016/j.cub.2011.05.021

1065

1066 Potì, P., Langer, J., Savage-Rumbaugh, S., & Brakke, K. E. (1999). Spontaneous

1067 logicomathematical constructions by chimpanzees (Pan troglodytes, P. paniscus). *Animal*

1068 *Cognition*, *2*(3), 147-156. http://dx.doi.org/10.1007/s100710050035

1069

1070 Povinelli, D.J., & Eddy, T. (1996). What young chimpanzees know about seeing.

1071 *Monographs of the Society for Research in Child Development, 61*(3), 1-189.

1072 http://dx.doi.org/10.2307/1166159

1073

1074 Povinelli, D. J., Perilloux, H. K., Reaux, J. E., & Bierschwale, D. T. (1998). Young and

1075 juvenile chimpanzees'(Pan troglodytes) reactions to intentional versus accidental and

1076 inadvertent actions. *Behavioural Processes*, *42*(2-3), 205-218.

1077 http://dx.doi.org/10.1016/S0376-6357(97)00077-6

1078

1079 Raihani, N., & Power, E. A. (2021). No Good Deed Goes Unpunished: the social costs of

1080 prosocial behaviour. PsyArXiv. http://dx.doi.org/10.31234/osf.io/ebfrg

1081

1082 Rakoczy, H., & Schmidt, M. F. (2013). The early ontogeny of social norms. *Child*

1083 *Development Perspectives*, *7*(1), 17-21. http://dx.doi.org/10.1111/cdep.12010

1084

1085 Read, D. W. (2008). Working memory: A cognitive limit to non-human primate recursive

1086 thinking prior to hominid evolution. *Evolutionary Psychology*, *6*(4), 147470490800600413.

1087 http://dx.doi.org/10.1177/147470490800600413

1088

1089 Read, D., & Van Der Leeuw, S. (2008). Biology is only part of the story…. *Philosophical*

1090 *Transactions of the Royal Society B: Biological Sciences*, *363*(1499), 1959-1968.

1091

1092 Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle.

1093 *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 427-

1094 431. http://dx.doi.org/10.1098/rspb.1998.0312

1095

1096   Roberts, W. A. (2002). Are animals stuck in time? *Psychological Bulletin, 128*(3), 473–489.

1097

1098   Roberts, G., Raihani, N., Bshary, R., Manrique, H.M., Fariña, A., Samu, F., & Barclay, P.

1099   (submitted). The benefits of being seen to help others: indirect reciprocity and reputation-

1100   based partner choice. Manuscript submitted to this special issue of *Philosophical*

1101   *Transactions of the Royal Society*.

1102

1103   Russell, Y. I., Call, J., & Dunbar, R. I. (2008). Image scoring in great apes. *Behavioural*

1104   *processes, 78*(1), 108-111. https://doi.org/10.1016/j.beproc.2007.10.009

1105

1106   Rutland, A., Cameron, L., Bennett, L., & Ferrell, J. (2005). Interracial contact and racial

1107   constancy: A multi-site study of racial intergroup bias in 3–5 year old Anglo-British children.

1108   *Journal of Applied Developmental Psychology*, *26*(6),699-713.

1109   http://dx.doi.org/10.1016/j.appdev.2005.08.005

1110

1111   Salvadori, E., Blazsekova, T., Volein, A., Karap, Z., Tatone, D., Mascaro, O., & Csibra, G.

1112   (2015). Probing the strength of infants' preference for helpers over hinderers: Two replication

1113   attempts of Hamlin and Wynn (2011). *PloS one, 10*(11), e0140570.

1114   https://doi.org/10.1371/journal.pone.0140570

1115

1116   Santos, L. R., Nissen, A. G., & Ferrugia, J. A. (2006). Rhesus monkeys, Macaca mulatta,

1117   know what others can and cannot hear. *Animal Behaviour*, *71*(5), 1175–1181.

1118   http://dx.doi.org/10.1016/j.anbehav.2005.10.007

1119

1120   Santos, F. P., Santos, F. C., & Pacheco, J. M. (2018). Social norm complexity and past

1121   reputations in the evolution of cooperation. *Nature*, *555*(7695), 242-245.

1122   http://dx.doi.org/10.1038/nature25763

1123

1124   Schacter, D. L., Addis, D. R., & Buckner, R. L. (2007). Remembering the past to imagine the

1125   future: The prospective brain. *Nature Reviews Neuroscience*, *8*(9), 657–661.

1126

1127   Schino, G., Boggiani, L., Mortelliti, A., Pinzaglia, M., & Addessi, E. (2021). Testing the two

1128   sides of indirect reciprocity in tufted capuchin monkeys. *Behavioural Processes, 182*,

1129   104290. https://doi.org/10.1016/j.beproc.2020.104290

1130

1131 Schmidt, M.F.H., & Rakoczy, H. (2019) On the uniqueness of human normative attitudes. In:

1132 Bayertz, K., Roughley, N. (eds) *The normative animal? On the anthropological significance*

1133 *of social, moral and linguistic norms*. Oxford University Press, Oxford

1134

1135 Schnell, A.K., Boeckle, M., Rivera, M., Clayton, N.S., & Hanlon, R.T. (2021). Cuttlefish

1136 exert self-control in a delay of gratification task. *Proceedings of the Royal Society of London.*

1137 *Series B: Biological Sciences*. *288*(1946), 20203161. https://doi.org/10.1098/rspb.2020.3161

1138

1139 Sebastián-Enesco, C., & Warneken, F. (2015). The shadow of the future: 5-year-olds, but not

1140 3-year-olds, adjust their sharing in anticipation of reciprocation. *Journal of Experimental*

1141 *Child Psychology*, *129*, 40–54. http://dx.doi.org/10.1016/j.jecp.2014.08.007

1142

1143 Seligman, M. E. P., Railton, P., Baumeister, R. F., & Sripada, C. (2013). Navigating into the

1144 future or driven by the past. *Perspectives on Psychological Science*, *8*(2), 119–141.

1145 https://doi.org/10.1177/1745691612474317

1146

1147 Shaw, A., Montinari, N., Piovesan, M., Olson, K.R., Gino, F., & Norton, M.I. (2014).

1148 Children Develop a Veil of Fairness. *Journal of Experimental Psychology: General*, *143*(1),

1149 363–375. https://doi.org/10.1037/a0031247

1150

1151 Sjåstad, H. (2019). Short-sighted greed? Focusing on the future promotes reputation-based

1152 generosity. *Judgment and Decision Making*, *14*(2), 199–213.

1153

1154 Smith, E. A., & Bliege Bird, R. (2000). Turtle hunting and tombstone opening public

1155 generosity as costly signaling. *Evolution and Human Behavior*, *21*(4), 245–261.

1156 https://doi.org/10.1016/S1090-5138(00)00031-3

1157

1158 Skinner, B. F. (1953). *Science and Human Behavior*. Free Press.

1159

1160 Spinozzi, G., Natale, F., Langer, J. & Brakke, K.E. (1999). Spontaneous class grouping

1161 behavior by bonobos (*Pan paniscus*) and common chimpanzees (*P. troglodytes*). *Animal*

1162 *Cognition*, *2*(3), 157-170. http://dx.doi.org/10.1007/s100710050036

1163

1164   Stephens, D.W., McLinn, C.M., & Stevens, J.R. (2002). Discounting and reciprocity in an

1165   iterated Prisoner's Dilemma. *Science*, *298*, 2216-2218.

1166   http://dx.doi.org/10.1126/science.1078498

1167

1168   Subiaul, F., Vonk, J., Okamoto-Barth, S. et al. (2008). Do chimpanzees learn reputation by

1169   observation? Evidence from direct and indirect experience with generous and selfish

1170   strangers. *Animal Cognition, 11*(4), 611–623. https://doi.org/10.1007/s10071-008-0151-6

1171

1172   Suddendorf, T. (2013). *The gap: The science of what separates us from other animals*. New

1173   York, NY, US: Basic Books.

1174

1175   Susini, I., Safryghin, A., Hillemann, F., & Wascher, C. A. F. (2020). Delay of gratification in

1176   non-human animals: A review of inter- and intra-specific variation in performance. BioRxiv.

1177   https://doi.org/10.1101/2020.05.05.078659

1178

1179   Tomasello, M. (2019). The moral psychology of obligation. *Behavioral and Brain Scien*ces,

1180   *43*, 1–33. http://dx.doi.org/10.1017/S0140525X19001742

1181

1182   Triki, Z., Wismer, S., Levorato, E., & Bshary, R. (2018). A decrease in the abundance and

1183   strategic sophistication of cleaner fish after environmental perturbations. *Global Change*

1184   *Biology, 24*, 481–489. http://dx.doi.org/10.1111/gcb.13943

1185

1186   Triki, Z., Emery, Y., Teles, M.C., Oliveira, R.F., & Bshary, R. (2020). Brain morphology

1187   predicts social intelligence in wild cleaner fish. *Nature Communications, 11*, 6423.

1188   http://dx.doi.org/10.1038/s41467-020-20130-2

1189

1190   Trösch, M., Bertin, E., Calandreau, L., Nowak R. & Lansade, L. (2020). Unwilling or willing

1191   but unable: can horses interpret human actions as goal directed? *Animal Cognition, 23*, 1035–

1192   1040. http://dx.doi.org/10.1007/s10071-020-01396-x

1193

1194   Van de Vondervoort, J. W., & Hamlin, J. K. (2018). The early emergence of sociomoral

1195   evaluation: infants prefer prosocial others. *Current Opinion in Psychology, 20*, 77-81.

1196   https://doi.org/10.1016/j.copsyc.2017.08.014

1197

1198   Van de Waal, E., Borgeaud, C., & Whiten, A. (2013). Potent social learning and conformity

1199   shape a wild primate's foraging decisions. *Science, 340*(6131), 483-485.

1200   http://dx.doi.org/10.1126/science.1232769

1201

1202   Völter, C. J., Mundry, R., Call, J., & Seed, A. M. (2019). Chimpanzees flexibly update

1203   working memory contents and show susceptibility to distraction in the self-ordered search

1204   task. *Proceedings of the Royal Society B: Biological Sciences, 286*(1907): 20190715.

1205   https://doi.org/10.1098/rspb.2019.0715

1206

1207   Vaish, A., Carpenter, M., & Tomasello, M. (2009). Sympathy through affective perspective

1208   taking and its relation to prosocial behavior in toddlers. *Developmental psychology, 45*(2),

1209   534-543. https://doi.org/10.1037/a0014322

1210

1211   Vaish, A., Missana, M., & Tomasello, M. (2011). Three-year-old children intervene in third-

1212   party moral transgressions. *British Journal of Developmental Psychology, 29*(1), 124-130.

1213   http://dx.doi.org/10.1348/026151010X532888

1214

1215   Vonasch, A. J., Reynolds, T., Winegard, B. M., & Baumeister, R. F. (2018). Death before

1216   dishonor: Incurring costs to protect moral reputation. *Social Psychological and Personality*

1217   *Science, 9*(5), 604–613. https://doi.org/10.1177/1948550617720271

1218

1219   Warneken, F., & Tomasello, M. (2007). Helping and cooperation at 14 months of age.

1220   *Infancy, 11*(3), 271-294. https://doi.org/10.1111/j.1532-7078.2007.tb00227.x

1221

1222   Washburn, D. A., Gulledge, J. P., James, F., & Rumbaugh, D. M. (2007). A species

1223   difference in visuospatial working memory: Does language link "what" with "where"?.

1224   *International Journal of Comparative Psychology, 20*(1), 55-64.

1225

1226   Wood, J. N., Glynn, D. D., Phillips, B. C., & Hauser, M. D. (2007). The perception of

1227   rational, goal-directed action in nonhuman primates. *Science, 317*(5843), 1402-1405.

1228   http://dx.doi.org/10.1126/science.1144663

1229

1230 Wu, J., Balliet, D., & Van Lange, P. A. (2016). Reputation management: Why and how

1231 gossip enhances generosity. Evolution and Human Behavior, 37(3), 193-201.

1232 http://dx.doi.org/10.1016/j.evolhumbehav.2015.11.001

1233

1234 Wu, J., Balliet, D., Tybur, J. M., Arai, S., Van Lange, P. A., & Yamagishi, T. (2017). Life

1235 history strategy and human cooperation in economic games. *Evolution and Human Behavior*,

1236 *38*(4), 496-505. http://dx.doi.org/10.1016/j.evolhumbehav.2017.03.002

1237

1238 Yoeli, E., Hoffman, M., Rand, D.G., & Nowak, M.A. (2013). Powering up indirect

1239 reciprocity with a large-scale field experiment. *Proceedings of the National Academy of*

1240 *Sciences of the USA*, *110*(Supplement 2), 10424-10429.

1241 https://doi.org/10.1073/pnas.1301210110

1242

1243 Zahavi, A. (1995). Altruism as a handicap: the limitations of kin selection and reciprocity.

1244 *Journal of Avian Biology*, *26*(1), 1-3. http://dx.doi.org/10.2307/3677205

1245

Author Note

1263   Correspondence concerning this article should be addressed to Héctor Marín Manrique,

1264   Department of Psychology and Sociology, Universidad de Zaragoza, Campus Universitario

1265   de Teruel, Ciudad Escolar, s/n. 44003 Teruel (Spain). E-mail: manrique@unizar.es; Tel: +34

1266   – 978 – 645343; Fax: +34 – 978 – 618103

1267   Héctor M. Manrique https://orcid.org/0000-0002-1943-340X

1268

1269

1270

1271

1272

1273

1274

1275

1276

1277