

On the lack of evidence that non-human animals possess anything remotely resembling a ‘theory of mind’

Derek C. Penn and Daniel J. Povinelli*

Cognitive Evolution Group, University of Louisiana, Louisiana, Lafayette, LA 70504, USA

After decades of effort by some of our brightest human and non-human minds, there is still little consensus on whether or not non-human animals understand anything about the unobservable mental states of other animals or even what it would mean for a non-verbal animal to understand the concept of a ‘mental state’. In the present paper, we confront four related and contentious questions head-on: (i) What exactly would it mean for a non-verbal organism to have an ‘understanding’ or a ‘representation’ of another animal’s mental state? (ii) What should (and should not) count as compelling empirical evidence that a non-verbal cognitive agent has a system for understanding or forming representations about mental states in a functionally adaptive manner? (iii) Why have the kind of experimental protocols that are currently in vogue failed to produce compelling evidence that non-human animals possess anything even remotely resembling a theory of mind? (iv) What kind of experiments could, at least in principle, provide compelling evidence for such a system in a non-verbal organism?

Keywords: theory of mind; folk psychology; mental state attribution; parsimony; chimpanzees; corvids

1. INTRODUCTION

Are humans alone in their capacity to reason about unobservable mental states, such as perceptions, intentions, emotions, desires and beliefs? Over a quarter-century ago, Premack & Woodruff (1978) launched a multinational industry dedicated to answering this question and coined the term, ‘theory of mind’ (hereafter, ToM) to refer to this distinctive capacity: ‘a system of inferences of this kind’, they observed, ‘may properly be regarded as a theory because such [mental] states are not directly observable, and the system can be used to make predictions about the behavior of others’ (p. 515).

Unfortunately, after decades of effort by some of our brightest human and non-human minds, there is still little consensus on whether or not non-human animals understand anything about unobservable mental states or even what it would mean for a non-verbal animal to understand the concept of a ‘mental state’. Nearly 10 years ago, Heyes (1998) observed that there had been ‘no substantial progress’ (p. 101) on Premack & Woodruff’s (1978) original question for many years. It is debatable whether there has been any more agreement on the matter since then (for the latest version of these ongoing and seemingly intractable debates, see Povinelli & Vonk 2003, 2004; Tomasello *et al.* 2003a,b; Tomasello & Call 2006).

Povinelli & Vonk (2004) pointed out one glaring reason for the impasse, namely comparative

researchers have never specified ‘the unique causal work’ that representations about mental states do above and beyond the work that can be done by representations of the observable features of other agents’ past and occurrent behaviours. As a result, almost all of the experimental protocols that have been used to test the ToM capabilities of non-human animals over the past quarter-century, including those that are currently in vogue today, are incapable, even in principle, of validating or falsifying the hypotheses being tested. One does not need to hold a Popperian view of science to acknowledge that arguments among unfalsifiable hypotheses are likely to be of little or no value to practicing scientists.

There seems to be a dire need, then, to focus more attention on the basic definitional and evidential issues confronting comparative researchers and less time spent arguing over ambiguous experimental results. In this paper, we will confront four related and contentious questions head-on:

- (i) What exactly would it mean for a non-verbal organism to have an ‘understanding’ or a ‘representation’ of another animal’s mental state?
- (ii) What should (and should not) count as compelling empirical evidence that a non-verbal cognitive agent has a system for understanding or forming representations about mental states in a functionally adaptive manner?
- (iii) Why have the kind of experimental protocols that are currently in vogue failed to produce compelling evidence that non-human animals possess anything even remotely resembling a theory of mind?

* Author for correspondence (djp3463@louisiana.edu).

One contribution of 19 to a Discussion Meeting Issue ‘Social intelligence: from brain to culture’.

- 129 (iv) What kind of experiments could, at least in
130 principle, provide compelling evidence for such a
131 system in a non-verbal organism?
132

133 Only after we have addressed these fundamental
134 issues in a formal, principled fashion will we be in a
135 position to attempt to answer the fascinating question
136 that Premack & Woodruff (1978) first posed so many
137 years ago.

138 Theory of mind, *sensu* Premack & Woodruff (1978),
139 entails the capacity to make lawful inferences about the
140 behaviour of other agents on the basis of abstract,
141 theory-like representations of the causal relation
142 between unobservable mental states and observable
143 states of affairs. This is certainly not the only way
144 to construe the capacity in question (for an overview
145 of the possibilities, see Davies & Stone 1995a,b;
146 Carruthers & Smith 1996). Many researchers have
147 argued, for example, that the ability to take the
148 causal role of mental states into account does not
149 involve theory-like inferences at all, but is grounded
150 in practical, sensorimotor, simulative abilities
151 (e.g. Gordon 1986, 1996; Goldman 1993).

152 For the purposes of the present essay, we wish to
153 remain rigorously agnostic as to *how* the capacity to
154 take other agents' mental states into account is
155 implemented. We will henceforth use the acronym
156 ToM, to refer to *any* cognitive system, whether theory-
157 like or not, that predicts or explains the behaviour of
158 another agent by postulating that unobservable inner
159 states particular to the cognitive perspective of that
160 agent causally modulate that agent's behaviour. We
161 believe this construal of ToM *sensu lato* is about as
162 broad and minimalist as possible without losing the
163 distinctive character of the capacity in question.

164 In our opinion, the major impediment that has
165 stood in the way of understanding whether or not
166 other species employ a ToM has been our species'
167 inveterate intuitions about how our own ToM works.
168 Appeals to folk psychological assumptions and
169 reasoning by analogy to introspective intuitions have
170 played an inordinate role in comparative researchers'
171 claims over the last quarter-century (see Povinelli &
172 Giambone 1999; Povinelli *et al.* 2000; Povinelli &
173 Vonk 2003, 2004). Thus, to undermine the insidious
174 role that introspective intuitions and folk psychology
175 play in the comparative debate, we propose to treat the
176 ToM explanandum here in more formalistic terms than
177 is typical among comparative researchers. Our
178 approach is as follows:

- 179
180 (i) present a simple formalism to clarify exactly
181 what is (and is not) at stake with respect to the
182 comparative ToM explanandum,
183 (ii) use the formalism in (i) to specify what should
184 (and should not) count as evidence for a ToM
185 system in a non-verbal organism,
186 (iii) take a prominent experimental result with
187 chimpanzees as a case study for exposing why
188 the kind of protocols currently in vogue do not
189 satisfy the conditions set out in (ii),
190 (iv) show why the analysis in (iii) applies, *mutatis*
191 *mutandis*, to the protocols currently being
192 employed with corvids as well, and

- 193 (v) propose two sample experimental protocols that
194 could, at least in principle, provide compelling
195 positive evidence for a ToM system in a non-
196 human species.
197

2. A SIMPLE FORMALISM

200 To begin, let us agree without too much argument that
201 cognitive agents—biological or otherwise—can learn
202 from their past experience, in part because they have
203 dynamic internal states that are decoupled from any
204 immediate physical connection to the external world.
205 Some of these internal states carry information about
206 what the agent has learned about the world that is
207 distinct from the information immediately available to
208 the system's perceptual inputs. And some of these
209 internal states describe goal states against which actual
210 states of the organism can be compared so that the
211 organism's behaviours can be dynamically adjusted in
212 order to close the gap. Let us denote all these internal
213 goal states by the variable, *g*, and all the informational
214 states that affect and/or mediate the goal-directed
215 behaviour of a cognitive agent by the variable, *r*.
216

217 Our rough-and-ready definition of *r*- and *g*-states is
218 meant to be as ecumenical as possible. For example, we
219 are entirely agnostic (for our present purposes anyway)
220 about whether an organism's *r*- and *g*-states are modal or
221 amodal, discrete or distributed, symbolic or connec-
222 tionist or even about how they come to have their
223 representational or informational qualities to begin
224 with. And we make no judgment about whether *r*- and
225 *g*-states as we have defined them here bear any
226 resemblance to the mental state concepts putatively
227 posited by our commonsense folk psychology. We do not
228 pretend that this definition of *g*- and *r*-states puts to rest
229 the entire (or even a small part of the) controversy over
230 what counts as goal-directed behaviour or internal
231 mental representations (see Markman & Dietrich
232 2000 for a better start); but it is good enough for our
233 present purposes.

234 Of course, there are innumerable other factors that
235 also contribute to shaping a biological organism's
236 behaviour, including information from sensory inputs,
237 feedback from perception-action loops, autonomic-
238 visceral states, the physical structure and capabilities of
239 the organism's body and all the other many variables
240 that influence the actions of situated, embodied,
241 biological agents in the wild. But for our present
242 purpose, these many multifarious influences can be
243 reduced to two additional variables and an ellipsis. We
244 will use the variable, *p*, to denote any dynamic,
245 occurrent information obtained through perceptual
246 inputs (including autonomic and proprioceptive
247 channels); and we will use the variable, *q*, to denote
248 feedback from the organism's sensorimotor loops
249 (including online and offline emulators). Using this
250 notation, any cognitive behaviour, *b*, can be described
251 formally (albeit simplistically) as follows:

$$252 \quad b = f(g, r, p, q, \dots). \quad (2.1) \quad 253$$

254 In other words, any cognitive behaviour is some
255 function of the system's *g*- and *r*-states plus any
256 occurrent information from perceptual inputs and

257 sensorimotor emulators at the time the function is
 258 computed—plus any other cognitive variables not
 259 incorporated in the present model. The reason we are
 260 unconcerned with unpacking such broad variables as g ,
 261 r , p and q , or with what falls under the ellipsis, is
 262 because we are only concerned, herein, with the
 263 question of whether or not a given cognitive agent
 264 possesses a ToM. And the question of whether or not a
 265 given cognitive agent possesses a ToM boils down to
 266 the question of whether or not that agent is able to treat
 267 other agents as if their behaviour is a function of the
 268 kind of variables described in equation (2.1). The only
 269 condition that must be met in order to qualify as a
 270 ToM, by our minimalist standards, is that the system
 271 must be able to produce and employ a particular class
 272 of information, namely information about the state of
 273 these cognitive variables from the perspective of that
 274 agent *as distinct from the perspective of the system itself*. We
 275 will refer to this special class of information by the
 276 variable, ms .

277 What exactly does it mean for one cognitive
 278 information state to be ‘about’ some other state of
 279 affairs? Much greater minds than ours have tried to
 280 answer this question (for example, Dretske 1988); and
 281 the complexities of taking this question seriously would
 282 take us far beyond the scope of the present essay. So
 283 here is a simple stop-gap answer that will suffice for our
 284 present purposes: let us agree that an ms variable carries
 285 information about some other cognitive state iff the
 286 state of the ms variable covaries with the state of the
 287 other cognitive state in a generally reliable manner such
 288 that, *ceteris paribus*, variations in the ms variable can be
 289 used by the consuming cognitive system to infer
 290 corresponding variations in the other cognitive state.

291 In a genuine mind-reader, the function describing
 292 the informational relation between one agent’s ms
 293 variables and another agent’s cognitive state variables
 294 might be something like the following:

$$295 \quad ms = f_{mr}(g^*, r^*, p^*, \dots), \quad (2.2)$$

296 where $*$ denotes the state of the corresponding variable
 297 for the other agent and f_{mr} denotes a cognitive function
 298 capable of intuiting the state of these unobservable
 299 variables directly, for example, telepathically.

300 Of course, there are no genuine mind-readers on this
 301 planet and all the relevant cognitive variables are,
 302 strictly speaking, unobservable from the point of view
 303 of the aspiring mind-reader. Hence, any purported
 304 mind-reading being performed on this planet is, in fact,
 305 a trick. A very good trick, to be sure, but a trick
 306 nevertheless. The trick is to be able to infer the state of
 307 the unobservable cognitive variables that will influence
 308 the behaviour of another agent using information
 309 observed from the perspective of the system itself:

$$310 \quad ms = f_{ToM}(r, p, \dots), \quad (2.3)$$

311 where f_{ToM} denotes a special function that computes an
 312 ms variable based on the inputs available to sentient,
 313 situated, embodied but non-telepathic organisms.

314 There is a burgeoning debate over how f_{ToM} might
 315 be implemented (for examples of the debate, see
 316 Davies & Stone 1995a,b; Carruthers & Smith 1996;
 317 Hurley & Chater 2005). Traditionally, f_{ToM} has been

318 construed as a kind of inferential function that uses a
 319 database of law-like generalizations to make logical
 320 inferences about other agents’ g - and r -states in a
 321 theory-like manner. This is certainly the kind of f_{ToM}
 322 that Premack & Woodruff (1978) had in mind when
 323 they coined the term that started the debate. But, as we
 324 noted previously, there are many alternative hypotheses
 325 at play today, some of which propose that f_{ToM} is
 326 implemented via offline simulation capabilities that
 327 encode ms variables about other subjects’ internal
 328 states using the same mechanisms that are used to
 329 encode ms variables about the subject’s own internal
 330 states. Still other researchers advocate hybrid functions
 331 between theory and simulation (e.g. Nichols & Stich
 332 2003; Meltzoff *in press*). For our present purposes, we
 333 are agnostic as to how the f_{ToM} is implemented; we
 334 simply note that a cognizer that has a ToM system of
 335 any kind must have an f_{ToM} of some kind. And any f_{ToM}
 336 must take information from the system’s own inputs
 337 and produce (or enact) a special class of information,
 338 i.e. information that is postulated to be from the
 339 cognitive perspective of another agent and relevant to
 340 predicting the behaviour of that agent.

341 The simple formalism we have proposed here leaps
 342 over innumerable details and complex, unresolved
 343 issues; but it nevertheless helps to keep track of, what
 344 is and what is not at stake with respect to the question
 345 of whether or not chimpanzees or any other non-
 346 human animal have a ToM. Our definition of an f_{ToM}
 347 does not require the agent to have any insight into the
 348 subjective phenomenological experience of others. Nor
 349 does our definition require ms variables to have an
 350 isomorphic relationship with the content or structure of
 351 the mental state that is being represented. Metarepre-
 352 sentations are one way of implementing ms variables.
 353 But they are certainly not the only way. Some theorists,
 354 for example, have argued that apes’ representations of
 355 mental states might simply involve ‘intervening vari-
 356 ables’ (aka ‘secondary representations’) rather than
 357 explicitly structured metarepresentations (Whiten
 358 1996, 1997, 2000; Suddendorf & Whiten 2001;
 359 Whiten & Suddendorf 2001). We believe Whiten and
 360 Suddendorf are right in this sense: being able to recode
 361 perceptually disparate behavioural patterns resulting
 362 from the same underlying cognitive state as instances of
 363 the same abstract equivalence class is a bona fide
 364 example of postulating an ms variable in the sense
 365 defined hereinabove (we differ from Whiten &
 366 Suddendorf (2001), however, in that we do not see
 367 any compelling evidence of this ability in non-human
 368 animals; see discussion below).

369 We particularly want to point out that the debate
 370 concerning whether or not non-human animals possess
 371 an f_{ToM} should not be concerned with whether or not
 372 they are cognitive creatures capable of reasoning about
 373 general classes of past and occurrent behaviours (e.g.
 374 ‘threat posture’, ‘eye or face direction’, ‘body position’
 375 or ‘eye-direction-in-relation-to-objects-in-the-world’).
 376 Indeed, they *must* be able to do so if they are potential
 377 candidates for a ToM at all. The theory of mind debate
 378 among comparative researchers should turn only
 379 around the question of whether, in addition to the
 380 representational abilities that any cognitive agent
 381 possesses as defined in equation (2.1), some particular

cognitive system in the agent in question also produces information that is specific to the cognitive perspective of another agent and uses this information to predict the behaviour of that agent.

3. WHAT SHOULD COUNT AS EVIDENCE OF AN f_{ToM} ?

We hope that our simplistic formalism will also help define more clearly what should and should not count as compelling evidence for an f_{ToM} . The subtle confounding problem, from an experimentalist's point of view, is that all organisms with the potential to have an f_{ToM} are also, necessarily, cognitive agents in the sense defined by equation (2.1) above.¹ The unavoidable null hypothesis is that any agent capable of possessing an f_{ToM} must already be employing the information provided by g , r , p and q in their cognitive behaviours. Thus, in order to produce experimental evidence for an f_{ToM} one must first falsify the null hypothesis that the agents in question are simply using their normal, first-person cognitive state variables as defined by equation (2.1). One must, in other words, create experimental protocols that provide compelling evidence for the cognitive (i.e. causal) necessity of an f_{ToM} in addition to and distinct from the cognitive work that could have been performed without such a function.

The last qualification is crucial. Imagine an organism, **A**, that always manifests some determinate set of observable cues, **C**₁, whenever it is in a given r -state, **r-state**₁, such that $P(\mathbf{r}\text{-state}_1|\mathbf{C}_1)=1$ and $P(\mathbf{r}\text{-state}_1|\sim\mathbf{C}_1)=0^2$. And suppose that **r-state**₁ causes **A** to emit behaviour **b**₁. A second cognitive agent having perceptual access to organism **A** and its observable traits, **C**₁, would have no need to infer the presence of **r-state**₁ in order to predict the occurrence of **b**₁; simply observing **C**₁ suffices. Thus, a researcher observing that a given experimental subject is able to reliably predict the occurrence of **b**₁ in **A** after observing **C**₁ would have no basis for concluding that the subject possesses an f_{ToM} dedicated to inferring **r-state**₁ (even though she, herself, may know that **r-state**₁ causes **b**₁) unless she can also show that possessing information directly about **r-state**₁ does some special causal work for **A** in addition to predicting **b**₁. Although this is rarely noted by experimentalists, we believe this point to be indisputable (see Povinelli & Vonk 2003, 2004). Curiously, though, it is nevertheless, often disputed, or completely ignored (see Tomasello *et al.* 2003a,b; Tomasello & Call 2006).

When framed in formalistic terms, the point appears obvious. But a simple real-life example will illustrate how easy it is to be duped by commonsense. A chimpanzee (the subject) observes a second chimpanzee, turn her head and look-off in the distance. In response, the subject turns his head in the same direction. From a folk psychological point of view (i.e. from the point of view of any normal adult human observer), it is tempting to conclude that the subject's act of turning his head is mediated by an internal representation of the second chimpanzee's belief that there is something interesting to look at and an implicit understanding that 'seeing' leads to a change in the internal, epistemic state of the looker. In other words, our commonsense intuitions assume that

the subject's behaviour was mediated by an ms variable (i.e. the subject had some understanding of the second chimpanzee's g - and r -states). Indeed, many comparative researchers have been tempted to attribute ms variables to their subjects under similar experimental circumstances (Call *et al.* 1998; Tomasello *et al.* 1999; Bugnyar & Heinrich 2005; Flombaum & Santos 2005; Santos *et al.* 2006; Tomasello & Call 2006).

What commonsense intuition overlooks, however, is that it is also possible for the same behaviour to be produced without an f_{ToM} of any kind. The set of perceptual cues available to the subject (i.e. 'eye or face direction', 'body position', 'eye-direction-in-relation-to-objects-in-the-world', etc.) are sufficient to explain the subject's behaviour. Any socially intelligent subject like a chimpanzee must possess a rich database of r -states based on what he has learned about perceptually similar situations in the past and the conditional dependencies that tend to hold between these observable cues and other animals' subsequent behaviour. Thus, the subject may have turned his head in the direction of the other chimp's head simply because it learned from past experience (or was born with the propensity to learn) that the given pattern of perceptual cues is a reliable indicator of something worth looking at in the direction inferred by the other agent's eyes and head. There is no need for the subject to reason in terms of an ms state variable, and positing an ms state variable does no additional explanatory work in the given situation.

The evidential case for an ms variable is no better simply because the second chimpanzee looks behind a barrier and the subject adjusts his position to see behind the barrier as well (e.g. Povinelli & Eddy 1996b). Barriers are, of course, visible entities. Subjects who have learned (or are born knowing) that they must alter their own position in order to see behind a barrier if a conspecific's eyes are directed towards a location behind a barrier do not necessarily need, in addition, to form representations postulating the hypothetical content of the conspecific's perceptual field or to understand that 'seeing' leads to any change at all in the looker's r -states (see also Povinelli *et al.* 2002).

And the evidential case is still no better just because the subject 'checks back' with the looker if he does not find anything interesting behind the barrier. Chimpanzees check back with moving objects all the time in order to update their internal representation of the object's location and projected trajectory without thereby postulating that all moving objects have mental states.

Following the gaze of a conspecific, checking behind barriers and checking back with the looker when nothing is found certainly *seem* to be compelling evidence for reasoning in terms of unobservable mental states when interpreted from a commonsense point of view. And it is easy to understand why normal adult human beings reflexively make this assumption when they interpret the behaviour of animals (Dennett 1987). From a scientific stance, however, we are only warranted in attributing an ms variable to the subject if we can specify why an f_{ToM} of some kind is computationally necessary in order to perform the given behaviour and why the information provided by the resulting ms variable is not redundant with the

information provided by the r , p , g and q variables which we have already posited to exist. The role of an experimentalist (as opposed to the folk observer) is to construct situations or protocols in which the unique cognitive work performed by the ms variables can be distinguished from the work that could be performed by r , p , g and q inputs alone.

Here is the crux of the matter then, and possibly the most important point we will make in this essay: in almost all experimental procedures reported to date, purported ms variables appear to be causally superfluous re-descriptions of the other observable inputs and representations that are logically required by the experimental design. No special f_{ToM} is required. The problem with existing protocols is that they fail to create situations in which the information purportedly carried by the ms variables is not causally redundant with the information already carried by the r , p , g and q variables.

Now, we are ready to evaluate the evidence with respect to the formalism we have outlined.

4. AN EXPERIMENTAL PROTOCOL THAT CANNOT, EVEN IN PRINCIPLE, PROVIDE EVIDENCE FOR f_{ToM}

This is not the forum for an exhaustive examination of all claims for theory of mind in chimpanzees (let alone other species). Our strategy, therefore, will be to examine what has come to be seen as the ‘strongest’ case for the existence of theory of mind in chimpanzees: the work of Hare *et al.* (2000, 2001). To be perfectly clear, we do not believe these studies have any bearing whatsoever, positive or negative, on the question of whether chimpanzees reason about mental states. However, because many other scholars believe they do, we shall use this protocol as a case study to expose the conceptual confusion that dominates this area of research.

We will take the ‘most significant’ experiment reported by Hare *et al.* (2001) as our example, but it must be noted that our analysis applies with equal force to all the experiments in this series (see also Povinelli & Vonk 2004). Two chimpanzees, one subordinate to the other, were kept in separate chambers on either side of a middle area. Two cloth bags in the middle chamber served as hiding places for small food items. Opaque doors on each side chamber prevented the respective chimpanzees from entering the middle chamber and retrieving the food until the doors were raised. On each trial, the subordinate’s door was partially raised while the food was being hidden, allowing the subordinate to peek out and see where the food items were placed and whether or not the dominant was present and looking. On each trial, the dominant’s door was either partially raised or completely closed while the food items were placed in one of the two containers. Once the food had been placed, the dominant’s door was closed and the subordinate was released into the middle chamber and given a slight headstart before the dominant was released as well.

Hare *et al.* (2001) reported a number of experimental conditions based on this protocol. In only one of these experiments, however, was the critical metric statistically significant³. In the uninformed condition of experiment 1, the dominant’s door was kept closed

while the food was hidden and the subordinate could see that the dominant’s door was closed; in the control condition, the dominant could see where the reward was hidden and the subordinate could see that the dominant was watching. The subordinate ‘approached’ the hidden food more often in the uninformed condition than in the control condition. On the basis of this result, Hare *et al.* (2001) concluded that ‘chimpanzees know what individual groupmates do and do not know’ (p. 148). Reversing their previous opinion on the matter (see Tomasello & Call 1997; Visalberghi & Tomasello 1998), Tomasello *et al.* (2003a) cite these experiments as ‘breakthrough’ (p. 154) evidence that chimpanzees ‘understand some psychological states in others’ (p. 156). Tomasello *et al.* are hardly alone. The Hare *et al.* (2000, 2001) results are now widely cited as supporting evidence for the idea that chimpanzees possess some kind of f_{ToM} .

Unfortunately, as our research group has pointed out that (see Karin-D’Arcy & Povinelli 2002; Povinelli & Vonk 2003, 2004), the protocol employed by Hare *et al.* (2001) lacks the power, even in principle, to distinguish between responses by the subordinate that could have been produced simply by employing observable information and representations of past behavioural patterns (i.e. p - and r -states) from responses that must have required computations involving information about the dominant’s unobservable mental states (i.e. ms states). For example, Povinelli & Vonk (2003) point out that the behaviour of the subordinates might result from a simple strategy glossed by ‘Don’t go after food if a dominant who is present has oriented towards it.’ The additional claim that the chimpanzees adopted this strategy because they understood that ‘The dominant knows where the food is located’ is intuitively appealing but causally superfluous.

Let us re-examine the problem with Hare *et al.*’s protocol using the formalism we developed above. Imagine an organism, **A**, that manifests some determinate set of observable cues, **C**₁, when it is in a given r -state, **r-state**₁, where **C**₁ = (‘eyes of **A** oriented towards food’, ‘uninterrupted visual access between **A** and placement of food’, ‘food is placed in location **X**’, ...) and **r-state**₁ = (‘**A** knows that food is in location **X**’). And suppose further that **r-state**₁ causes **A** to emit behaviour **b**₁, where **b**₁ = (‘**A** tries to retrieve food in location **X**’). A second cognitive agent having perceptual access to organism **A** and its observable traits, **C**₁, would have no need to infer the presence of **r-state**₁ in order to predict the occurrence of **b**₁; simply observing **C**₁ suffices. Thus, a researcher observing that a given experimental subject is able to reliably predict the occurrence of **b**₁ in **A** after observing **C**₁ would have no basis for concluding that the subject possesses an f_{ToM} dedicated to inferring **r-state**₁ (even if she herself knows that **r-state**₁ causes **b**₁), unless she can also show that possessing information directly about **r-state**₁ does some special causal work in addition to predicting **b**₁. Once again, we believe this point to be indisputable—though, as in the case of Hare *et al.* (2001), persistently (and inexplicably) disputed (see Tomasello *et al.* 2003a,b; Tomasello & Call 2006).

5. WHAT ABOUT CORVIDS?

Chimpanzees, of course, are not the only non-human species which might be potential candidates for an f_{ToM} . And, indeed, some of the most well-controlled results and provocative claims in recent years have not come from experiments with primate subjects at all, but from experiments with corvids (for general reviews of the literature, see Clayton *et al.* 2001; Emery 2004; Emery & Clayton 2004, 2005; Clayton & Emery 2005; see also Clayton *et al.* 2006). Corvids are quite adept at pilfering the food caches of other birds and will adjust their own caching strategies in response to the potential risk of pilfering by others. Indeed, not only do they remember which food caches were observed by competitors, but also they appear to remember the specific individuals who were present when specific caches were made and modify their re-caching behaviour accordingly (Dally *et al.* 2006). Corvids' cognitive prowess is not limited to caching and pilfering. In many tool-use tasks, their cognitive abilities also seem to be superior to those of non-human primates in certain respects (for example, Hunt 1996, 2004; Seed *et al.* 2006; Tebbich *et al.* in press). What is at issue here, however, is not whether or not corvids are cognitively sophisticated creatures, but whether or not, *in addition*, any of their sophisticated cognitive abilities require the possession of an f_{ToM} .

Many comparative researchers clearly feel the answer to this question is yes. For example, Emery & Clayton (2001, 2004, 2005) suggest that corvids discriminate between competitors who possess knowledge of cache sites from those that do not by attributing specific, contentful r -states to knowledgeable competitors. Moreover, Emery and Clayton suggest that corvids may be able to understand the internal mental experience of their conspecifics by analogy to their own first-hand experience (see also Emery 2004). Similarly, Bugnyar & Heinrich (2006) showed that ravens delay pilfering from cache sites when confronted by the individuals who made those caches and suggest that this is consistent with the hypothesis that corvids possess a sophisticated understanding of others' visual perception as well as the ability to tactically manipulate competitors' mental states (see also Bugnyar & Heinrich 2005).

While we certainly agree with these researchers that it is *possible* that corvids are capable of reasoning in terms of the r -states of their competitors, we nevertheless must point out that none of the evidence to date provides convincing evidence for this hypothesis. One of the defining characteristics of *ms* variables, as defined above, is that they are construed from the cognitive perspective of the other agent as distinct from the cognitive perspective of the subject itself. Unfortunately, none of the reported experiments with corvids require the subjects to infer or encode any information that is unique to the cognitive perspective of the competitor. For example, none of the reported experiments require the subjects to reason in terms of the *counterfactual* content of their competitors' r -states. As Dennett (1987) pointed out a long time ago, without evidence that a subject is able to reason in terms of counterfactual as well as factual r -states in another agent, it is very difficult, if not impossible, to

provide evidence that they are cognizing the other agent's r -states qua r -states at all.

In all of the experiments with corvids cited above, it suffices for the birds to associate specific competitors with specific cache sites and to reason in terms of the information they have observed from their own cognitive perspective: e.g. 'Re-cache food if a competitor has oriented towards it in the past', 'Attempt to pilfer food if the competitor who cached it is not present', 'Try to re-cache food in a site different from the one where it was cached when the competitor was present', etc.⁴ The additional claim that the birds adopt these strategies because they understand that 'The competitor knows where the food is located' does no additional explanatory or cognitive work.

The case for 'experience projection' is no stronger than the case for 'knowledge attribution'. Emery & Clayton (2001) showed that scrub jays who had had previous experience pilfering food from others were more likely to re-cache food that had been observed by competitors than birds who had had no previous experience pilfering from others. 'This result raises the exciting possibility,' Emery (2004, p. 21) writes, 'that birds with pilfering experience can project their own experience of being a thief onto the observing bird, and so counter what they would predict a thief would do in relation to their hidden food' (see also Emery & Clayton 2004).

The fact that only birds with previous pilfering experience re-cache observed food sites is an interesting result but sheds no light on the internal mental representations or cognitive processes being employed by the birds in question. This experimental result certainly does not demonstrate that ex-pilferers understand anything about the internal, subjective experience of their potential competitors. Monkeys, after all, often initiate aggressive acts against innocent third parties after they themselves have been attacked but this hardly means that they are projecting their own subjective experience of being attacked onto the potential victims. There are any number of much lower-level explanations for this redirected aggression (see Silk (2002) for a review)—as there are for the connection between pilfering and re-caching in corvids.

To be sure, many researchers explicitly acknowledge that an explanation based on reasoning about observed cues alone is sufficient to account for the existing data. Dally *et al.* (2006), for example, acknowledge, that scrub jays' ability to keep track of which competitors have observed which cache sites 'need not require a humanlike 'theory of mind' in terms of unobservable mental states, but [...] may result from behavioral predispositions in combination with specific learning algorithms or from reasoning about future risk.' Similarly, Bugnyar & Heinrich (2006) acknowledge that a representation of 'states in the physical world' would be sufficient for explaining the available evidence concerning the manipulative behaviours of ravens. Notwithstanding the foregoing, these researchers continue to hold out the 'possibility' that the birds' behaviour could be consistent with a more generous, mentalistic interpretation and suggest that more generous interpretations might be more 'parsimonious' (see also Tomasello & Call 2006).

Admittedly, explanations in terms of folk psychological abilities do appear more ‘parsimonious’ at first blush. But the fact that such explanations are ‘simpler for us’ to understand does not mean, as Heyes (1998) pointed out, that they are ‘simpler for them’ to implement (see also Dennett 1987). The cognitive mechanisms that would be required to actually implement these purported f_{ToM} abilities at a subpersonal, causal level are hardly simple at all—they only seem simple because folk psychological explanations gloss over all the devilish details. Comparing the simplicity of a folk psychological explanation, e.g. ‘chimpanzees understand seeing’, ‘corvids know what others do and do not know’, to the complexity of a subpersonal cognitive explanation is like comparing a marketing description of Microsoft Word, e.g. ‘prints, saves and edits complex documents’, to a detailed functional specification of the underlying application architecture. The fact that the detailed functional specification of Microsoft Word runs to thousands of pages, and the marketing pitch takes one sentence is not a reasonable metric for comparing the merits of the two descriptions. Likewise, while folk psychological descriptions may be invaluable heuristics for ethologists in the field (Dennett 1987), they should not be confused or compared with cognitive hypotheses framed at a subpersonal, functional level of explanation.

Our position is that chimpanzees and corvids (like many other non-human animals) possess representational architectures of enormous sophistication and flexibility. We also believe that they employ both inferential and simulative mechanisms for forming abstractions about classes of behaviours and environmental conditions that are relevant to their goal-directed actions. Furthermore, we believe that non-human animals are able to generalize the lessons learned from these abstractions to novel scenarios.

Thus, unlike the motley collection of learning experiences that might be required in an associationist model, our hypothesis is that non-human animals are able to respond intelligently to novel situations based on general, abstract representations (i.e. r -states) they have formed about similar situations in the past and specific, concrete representations they have formed about the events leading up to the present moment (including, at least in the case of corvids, the ‘what’, ‘when’ and ‘where’ information associated with those events).

Our principle disagreement with those who explain non-human behaviours in terms of an f_{ToM} is not about the inferential or learning abilities that non-human animals possess (at least for our present purposes; but see Penn & Povinelli in press). Our principle disagreement is about the kind of representations over which these inferential and learning processes operate. The available evidence suggests that chimpanzees, corvids and all other non-human animals only form representations and reason about *observable* features, relations and states of affairs from their own cognitive perspective. We know of no evidence that non-human animals are capable of representing or reasoning about *unobservable* features, relations, causes or states of affairs or of construing information from the cognitive

perspective of another agent. Thus, positing an f_{ToM} , even in the case of corvids, is simply unwarranted by the available evidence.

6. TWO EXPERIMENTAL PROTOCOLS THAT COULD, IN PRINCIPLE, PROVIDE EVIDENCE FOR f_{ToM}

In response to the kind of critiques that our research group has levelled, some scholars have claimed that the distinctions we are proposing are experimentally intractable and/or empirically vacuous. For example, Andrews (2005) worries that ‘any success in a predictive paradigm can be explained as the result of a behavioristic psychological system that relies on behavioral, rather than mental, intervening variables’ (p. 528 and see also Leavens *et al.* 2004; Hurley 2006). Tomasello *et al.* (2003b) worry that our extreme stinginess in attributing mentalistic abilities to chimpanzees is an example of ‘derived behaviourism’ and will only lead to ‘despair’ (p. 239).

To forestall any worry that a theoretically rigorous stance towards the interpretation of comparative experimental results will lead only to despair, we will now propose two separate experimental protocols that could, in fact, provide principled evidence for an f_{ToM} in chimpanzees or corvids and could be easily adapted for other non-verbal cognitive organisms as well. The first tests a non-verbal subject’s ability to reason from first- to third-person mental states. The second tests a subject’s ability to use *ms* variables to solve prediction problems that would be computationally unsolvable otherwise. We hope these two proposals will demonstrate that our stringent criteria for attributing an f_{ToM} to a non-human animal are neither empirically vacuous nor experimentally intractable.

(a) *The opaque visor experiment*

Building on previous suggestions, Povinelli & Vonk (2003, 2004) highlighted (in a version appropriate for chimpanzees) one protocol that could provide principled positive evidence for f_{ToM} in a non-verbal organism. Since this proposal has now been critiqued, we briefly summarize its logic, and show why the critiques are invalid.

During an initial training session, subjects are given first-hand experience wearing two mirrored visors. One of the visors is see-through; the other is not. The visors themselves are of markedly different colours (and/or shape). During the subsequent test session, the subjects are given the opportunity to use their species-typical begging gesture to request food from one of the two experimenters, one wearing the see-through visor and the other wearing the opaque visor. Subjects who beg significantly more often from an experimenter wearing the see-through visor have manifested evidence of possessing an f_{ToM} in the sense defined herein.

This protocol has been tested on highly human-enculturated chimpanzees (Vonk *et al.* in press), who failed. A functionally equivalent variation of the protocol (using trick blindfolds) has been tested on 18-month-old human infants (Meltzoff in press), who passed. These results would seem to provide positive confirmatory evidence that even very young human

infants possess some sort of f_{ToM} whereas even highly enculturated adult chimpanzees do not.

There have been several criticisms of the experimental protocol, ranging from the claim that it is formally inadequate (Andrews 2005; Hurley 2006) to the claim that it has ‘very low ecological validity’ (Tomasello *et al.* 2003b). We will first defend why the proposed experiment does, in fact, provide principled evidence for an f_{ToM} and, secondly, why the charge of ‘low ecological validity’ is misplaced.

Both Hurley (2006) and Andrews (2005) argue that a subject could pass the proposed experiment simply by reasoning about the analogy between first-person manifest physical behaviours and third-person manifest behaviours. As Andrews (2005) puts it:

...the chimp might make the behavioral connection between wearing the opaque bucket and *not being able to do things* [emphasis in the original]. From whom should he beg? Certainly not the person who isn’t able to do things (p. 530).

It is certainly true that reasoning from first- to third-person behaviours forms a crucial part of the human cognitive tool-kit (for example, Meltzoff & Moore 1997; Meltzoff *in press*). And there is substantial evidence that neural systems, such as ‘mirror neurons’, in both human and non-human animals register correspondences between first- and third-person behaviours (for reviews of the literature, see Hurley & Chater 2005). Thus, it is possible (though certainly not proven) that the capacity to find behavioural equivalences between self and other is, as Hurley (2006) argues, developmentally and phylogenetically prior to the capacity to find mentalistic equivalences between self and other.

However, the ability to form first- to third-person equivalences in terms of manifest physical behaviours is not sufficient to solve the protocol proposed by Povinelli & Vonk. The reason the bucket protocol works as a test of mental state reasoning is because there is, in fact, no way (i.e. no computationally tractable way) to draw the necessary correspondences based purely on representations of observable information and manifest behaviours.

In this context, let us examine more closely the data available to a subject lacking an f_{ToM} . Such a subject would be limited to r -states about his own manifest behaviour while wearing the opaque visor (e.g. ‘I stumbled around while wearing the red visor’) and occurrent p -states about the experimenter (e.g. ‘she is wearing a red visor’). However, a subject lacking an f_{ToM} would not have access to r -states about his own internal cognitive states while wearing the visors (e.g. ‘I was unable to see while wearing the red visor’). Nor would such a subject have any information concerning his own propensity to respond to begging gestures while wearing the opaque visor, since he never attempted to respond to begging gestures while wearing the visor.

Thus, a subject capable of cognizing analogies between first- to third-person physical behaviours, but incapable of cognizing analogies between unobservable mental states, might be able to infer that the experimenter will stumble around and bump into things while wearing the red visor; but there would be no basis for this subject to infer that wearing the red

visor will necessarily preclude the experimenter from *physically* producing the actions necessary to respond to begging gestures. Indeed, the subject would have every reason to believe that wearing the red visor will have no effect at all on the experimenter’s ability to respond to begging gestures.

In the proposed protocol, the only manifest physical actions required for the experimenter to respond to begging gestures are the ability to sit still, move her arm and keep her eyes open and directed straight ahead. The subject has first-hand experience that he is perfectly capable of sitting still, of freely moving his arms and of keeping his eyes open while wearing the red visor. Thus, based on the manifest behavioural evidence, a subject without an f_{ToM} would have no reason to suspect any limitation on the experimenter’s ability to perform the physical acts required to respond to begging gestures. In order to infer that the experimenter is not likely to respond to begging gestures while wearing the red visor, the subject must realize that responding to begging gestures requires more than a set of manifest physical actions and observable conditions. To be precise, the subject must realize (by logical inference or embodied simulation, or some combination of the two) the following:

- (i) wearing the opaque visor results in an inability to ‘see-what-is-going-on’ (i.e. a general epistemic condition applicable to any subsequent behaviour not just a particular manifest physical effect of bumping-into-things),
- (ii) this general epistemic condition will be experienced, analogously, by the other subject when she wears the red visor but not the blue visor, and
- (iii) a subject who experiences this general epistemic condition will not respond to begging gestures.

The preceding three steps are a paradigmatic example of encoding an ms variable about a first-person internal state (i.e. the general epistemic condition of not-being-able-to-see) that results from a given manifest contingency (i.e. wearing the red visor) and then using these representations to predict the behaviour of another cognitive agent to a novel situation (i.e. responding to begging gestures). We contend that without the ms variable, the subject could not immediately solve the problem presented.

Some (e.g. Andrews 2005) might still object that during the initial, first-person familiarization phase, the chimpanzee could form a general aversion to red visors or might make the blanket inference that since ‘I can’t do anything with the red visor on’, others will not be able to do anything either.

We should first point out that no such generalized aversion to the opaque bucket was observed in the familiarization phase of this experiment with chimpanzees (Vonk *et al.* *in press*). More importantly, the protocol calls for the subjects to learn that they can do many things while wearing the opaque visors: they run about, reach out, feel objects and their body, and they themselves engage in acts that look very much like begging gestures (Vonk *et al.* *in press*). Thus, it is simply false that the subjects learn that ‘I can’t do anything with the red visor on’.⁵

1025 We now turn to Tomasello *et al.*'s (2003b) objection
 1026 that the visor test lacks 'ecological validity' because it
 1027 involves a 'cooperative–communicative' rather than a
 1028 'competitive' paradigm (Hare 2001) and because it
 1029 involves strange artefacts like visors.

1030 Several things need to be noted about this objection.
 1031 First, it is simply false to claim that chimpanzees are
 1032 more likely to reveal their true cognitive potential under
 1033 'competitive' situations rather than 'cooperative/com-
 1034 municative' ones (Hare 2001). Certainly, they may
 1035 exhibit different cognitive abilities in competitive versus
 1036 cooperative/communicative situations, but there is no
 1037 empirical or theoretical basis for claiming that the
 1038 abilities revealed under competitive paradigms are
 1039 either more fundamental or more sophisticated than
 1040 those revealed under cooperative ones.

1041 For example, consider the chimpanzees' natural
 1042 food-begging gesture (Goodall 1990), a gesture that
 1043 has been observed in all captive and free-ranging
 1044 populations of chimpanzees. In a simple experimental
 1045 setting, if a chimpanzee is confronted with two
 1046 caretakers who could potentially give them food, but
 1047 one is facing towards them and the other is facing away,
 1048 the chimpanzee will immediately (from trial one
 1049 forward) gesture to the one facing them (Povinelli &
 1050 Eddy 1996a–c). Chimpanzees are even capable of
 1051 selectively employing auditory rather than visual
 1052 behaviours as a function of specific perceptual/behav-
 1053 ioural cues exhibited by the caretaker from whom they
 1054 are begging (Hostetter *et al.* 2001; Leavens *et al.* 2004).
 1055 It is only when more subtle experimental manipula-
 1056 tions are employed, that chimpanzees display their lack
 1057 of understanding of the specific causal relation between
 1058 the disposition of the eyes or face of the caretaker and
 1059 the caretaker's mental state (see Povinelli (2003)
 1060 chapter 3 for a review).⁶ Of course, this cooperative–
 1061 communicative act—gesturing to the front (as opposed
 1062 to the back) of a communication partner—is part of the
 1063 natural social behaviour of chimpanzees (see Tomasello
 1064 *et al.* 1994), as is competition over food resources
 1065 (Karin-D'Arcy & Povinelli 2002). In other cooperative
 1066 experimental settings, where a chimpanzee needs help
 1067 in obtaining a just-out-of-reach food item, chimpan-
 1068 zeas will robustly modulate their gestures to fit the
 1069 locations to where their cooperative partner is looking
 1070 (Povinelli & Vonk 2004). Thus, we are just as
 1071 impressed by the sophistication of chimpanzee social
 1072 cognition in cooperative–communicative situations as
 1073 we are by their sophistication in competitive ones.

1074 Claiming that visors are ecologically 'unnatural'
 1075 (Hare 2001, p. 276) is a disingenuous argument. When
 1076 chimpanzees pass tests involving ecologically bizarre
 1077 artefacts, such as blindfolds, locked boxes, transparent
 1078 tubes and mirrors, the same experimenters are quick
 1079 to claim victory. When chimpanzees fail, the visors are
 1080 to blame.

1081 In any case, the point of the proposed protocol is not
 1082 the visors. The point of the proposed protocol is the
 1083 functional, informational challenge it poses. There are
 1084 certainly many species for whom having a visor
 1085 covering their eyes is not a species-typical experience.
 1086 It suffices to find an alternative implementation of the
 1087 experiment that retains the same informational and
 1088 functional challenge in a more species-acceptable form.

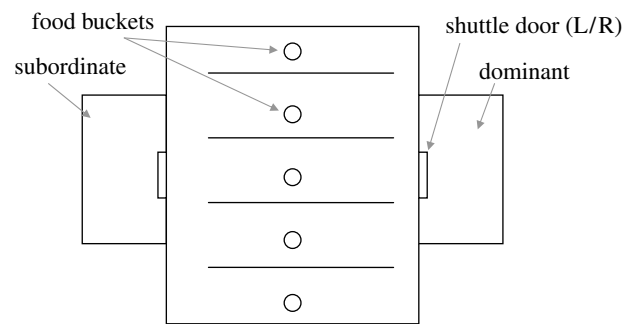


Figure 1. General experimental set-up for five-bucket protocol (see §6b for details).

Meltzoff (in press) provides an exemplary case study: he cleverly adapted the proposed protocol for human infants using blindfolds and recorded whether or not infants were more likely to track the gaze of an adult wearing an opaque blindfold than one wearing a see-through blindfold. Notably, although tracking the gaze of blindfolded adults has pretty low ecological validity for human children as well, the 18-month-old children, nevertheless, passed.

To be sure, it is true that failure on any experimental test of this sort is not demonstrative evidence of a lack of f_{ToM} : false negatives are a fact of life in comparative research as they are in ToM research in general (see Birch & Bloom 2004). *Ceteris paribus*, ecological validity is often (but not always) a desirable feature of comparative experimentation. But the more critical issue is to isolate experimental procedures that are capable, at least in principle, of providing positive (or negative) evidence for the specific cognitive skills. Unfortunately, most of the 'ecologically valid' protocols currently in vogue cannot provide principled evidence for or against the presence of f_{ToM} . The proposed visor protocol is simply one example of an experiment that can.

For those who nonetheless insist that only competitive paradigms will reveal the true nature of chimpanzee cognition, we propose a second experimental protocol below that retains the purported 'ecological validity' of Hare *et al.*'s (2001) competitive paradigm while, nevertheless, proffering the possibility of positive evidence for an f_{ToM} .

(b) A systematic version of Hare *et al.*'s competitive food protocol

As in Hare *et al.*'s (2001) experiment described above (see §4), a subordinate and a dominant chimp are kept in separate compartments on opposite sides of a middle chamber and each side chamber is separated from the middle chamber by an opaque shuttle door (see figure 1). The doors are raised and lowered and the two subjects released into the middle chamber. Unlike Hare *et al.*'s set-up, however, the middle chamber has n stalls (e.g. 5) spaced evenly across the width of the compartment, divided from each other by Plexiglas walls. There are five buckets on the floor at the centre of each stall in full view of the subjects. The contents of the bucket, however, are not directly visible to the subjects. On each trial, the experimenter places two different amounts of food into two different buckets:

1153 a larger amount of food is placed in one bucket and a
1154 visibly smaller amount of food is placed in another. The
1155 order in which the amounts are placed is randomized
1156 (i.e. on one-half of the trials, the larger amount is
1157 placed first).

1158 The experiment is carried out in a series of
1159 incrementally more challenging steps. In the step 1,
1160 subjects are exposed to a series of non-competitive
1161 trials. There is no rival present during these trials and
1162 both rewards are placed in full view of the subject.
1163 When the subject is released, it is only allowed to
1164 approach and retain the contents of one bucket. Trials
1165 continue until the subject learns to reliably approach
1166 and retain the more desirable reward.

1167 In step 2, chimpanzees are paired in dominant/sub-
1168 ordinate dyads. In each dyad, both chimpanzees have
1169 full visual access to the placement of both rewards.
1170 Only dyads in which subordinates learn to retrieve the
1171 less desirable reward and dominants retrieve the more
1172 desirable reward in a reliable fashion are allowed to
1173 continue to the third and final session.

1174 In the step 3, the following conditions are randomly
1175 presented (Note that in all conditions, the subordinate
1176 has complete visual access to the activities of the
1177 experimenter. Only the dominant's visual access is
1178 manipulated as described.):

1180 — *Informed control*. Both chimpanzees have full visual
1181 access to the placement of both food rewards.

1182 — *Partially uninformed*. One reward is placed while
1183 the dominant chimp is looking and the other reward
1184 is placed while dominant's door is down. Whether
1185 or not the dominant's door is down during the
1186 initial placement or the subsequent placement is
1187 randomized.

1188 — *Removed informed*. Both rewards are placed while the
1189 dominant subject is looking. Then, one of the
1190 rewards is removed from the middle chamber and
1191 replaced with an empty bucket while the dominant is
1192 looking.

1193 — *Removed uninformed*. Both rewards are placed while
1194 the dominant subject is looking. Then, one of the
1195 rewards is removed from the middle chamber and
1196 replaced with an empty bucket while the dominant's
1197 door is down.

1198 — *Moved*. The dominant's door is down during the
1199 initial placement of two rewards; then the domi-
1200 nant's door is open and both rewards are moved to
1201 new locations while the dominant is watching.

1202 — *Replaced*. The dominant witnesses the placement of
1203 one of the two rewards and then the dominant's door
1204 is closed while that reward is moved to a new
1205 location and the amount not witnessed is placed in
1206 the previously occupied bucket.

1207 — *Misinformed*. Both rewards are placed while the
1208 dominant is looking; then, while the dominant's
1209 door is down, one of the buckets (which may or may
1210 not have food in it) is moved to the location occupied
1211 by one of the rewards, that bucket and its reward are
1212 moved to a new location and the bucket at that
1213 location is put back in the stall originally occupied by
1214 the first bucket.

1215 — *Swapped*. Both rewards are placed while the domi-
1216 nant is looking, then the locations of the two buckets

are swapped while the dominant's door is down. 1217

— *Other variations*. Note that the conditions described 1218
above only represent a subset of the systematic 1219
variations which could be employed. 1220

Q7 The initial two steps can be mastered using simple 1221
heuristics based on observable contingencies. 1222
However, if the subject learns to pass the initial sessions 1223
using only observable contingencies, and does not have 1224
access to an f_{ToM} , the final test session presents an 1225
intractable mess. 1226

For example, the response rule 'Don't go after food 1228
if the dominant has oriented towards it in its present 1229
location' (Povinelli & Vonk 2003), which worked 1230
perfectly in the original protocol proposed by Hare 1231
et al. (2001), no longer suffices. The relational rule 1232
'always retrieve the less desirable of two rewards when 1233
there's a dominant present' only works consistently 1234
under the *informed control* and *moved* conditions. Even 1235
the higher-order relational strategy, 'Go after the less 1236
desirable reward unless the dominant has previously 1237
oriented towards it in its current location' fails any 1238
condition in which it would be optimal for the 1239
subordinate to retrieve the larger food item (e.g. the 1240
Swapped condition). Based purely on patterns of 1241
observable cues, each condition requires a different 1242
response rule; and there is no way to systematically 1243
generalize from familiar to novel conditions. 1244

For the purposes of testing whether or not a subject 1245
possesses an f_{ToM} , the critical conditions are those 1246
which require the subject to formulate an *ms* variable 1247
that keeps track of where the dominant believes the 1248
food rewards are located as distinct from where they are 1249
actually located, e.g. the *removed uninformed*, *replaced*, 1250
misinformed and *swapped* conditions. In the context of 1251
the present protocol, i.e. randomly interspersed among 1252
the other conditions, there is no way for a subject to 1253
reliably pass these critical conditions without the ability 1254
to keep track of the counterfactual state of affairs from 1255
the dominant's cognitive perspective while simultane- 1256
ously keeping track of the occurrent state of affairs 1257
from the subject's own perspective. The subject must 1258
not only understand that the competitor was present 1259
and oriented; but he must also cognize the specific 1260
content of the competitor's counterfactual *r*-states and 1261
relate these counterfactual *r*-states to the competitor's 1262
subsequent behaviour. Success on these conditions is 1263
thus functionally (though not necessarily psychologi- 1264
cally) equivalent to reasoning in terms of a competitor's 1265
'false beliefs' and would provide compelling evidence 1266
for an f_{ToM} . 1267

Failure, however, is no less instructive than 1269
success. A subject who has passed the first two 1270
training steps has clearly understood the procedural 1271
aspects of the task, and the protocol retains the 1272
competitive food paradigm advocated so vigorously by 1273
Hare (2001) and others. Thus, unlike previous non- 1274
verbal 'false belief' tests (e.g. Call & Tomasello 1999) 1275
or even the protocol proposed by Hare *et al.* (2001), 1276
failure on this one cannot be blamed on interspecific 1277
misunderstandings, ecological implausibility or the 1278
subjects' inability to understand the procedural 1279
aspects of the task. 1280

1281 Indeed, it is the *pattern* of successes and failures on
1282 different conditions in our protocol that is likely to
1283 provide the most interesting evidence concerning the
1284 cognitive strategy being employed by a given non-
1285 human subject. For example, a subject who employs a
1286 ‘Don’t go after a food reward if the dominant has
1287 oriented towards it’ strategy will pass a different set of
1288 conditions than a subject employing a ‘Always retrieve
1289 the less desirable of the two food amounts’ strategy.
1290 Similarly, a subject who passes the *removed informed*
1291 condition but not the *removed uninformed* condition (or
1292 vice versa) has revealed something significant about the
1293 characteristics and the limitations of the cognitive
1294 strategy he is employing.

1295 It might be objected that the complexity of the
1296 conditions in our version of Hare *et al.*’s protocol is too
1297 great for chimpanzees or corvids to handle and that the
1298 processing capacity limitations of these subjects are
1299 orthogonal to the question of whether or not they
1300 possess an f_{ToM} . The conditions in our five-bucket
1301 protocol do, indeed, pose a significant degree of
1302 ‘relational complexity’ (Halford *et al.* 1998), but we
1303 disagree with the claim that this invalidates the protocol
1304 as a test of a subject’s ability to reason about what their
1305 conspecifics do and do not know.

1306 While our five-bucket protocol poses an intractable
1307 computational challenge to a subject without an f_{ToM}
1308 of any kind, our protocol would be much less daunting
1309 to a subject who is able to encode the appropriate *ms*
1310 state variables. As Whiten and Suddendorf pointed
1311 out, one function of an f_{ToM} is to reduce the
1312 complexity of social interactions by positing abstract
1313 hidden variables that encode abstract, relational
1314 similarities between perceptually disparate behavioural
1315 patterns (Whiten 1996, 1997, 2000; Suddendorf &
1316 Whiten 2001; Whiten & Suddendorf 2001). For
1317 example, a subject endowed with the appropriate
1318 simulational abilities should be able to significantly
1319 reduce the relational complexity of the task by first
1320 simulating what they would do from the perspective of
1321 the dominant competitor. (Indeed, we suspect that
1322 many readers did exactly this while reading the
1323 description of each condition.)

1324 Furthermore, we would argue that the ability to
1325 perceive relational similarities between perceptually
1326 disparate behavioural patterns (i.e. to form ‘abstract
1327 equivalence classes’; in Whiten’s (1996) terms) and to
1328 postulate the existence of unobservable causes like
1329 mental states are paradigmatic examples of higher-
1330 order relational reasoning (see Gentner *et al.* 2001 for
1331 an overview of the current literature; see Penn &
1332 Povinelli *in press* for a relational analysis of non-
1333 human causal cognition). Consistent with this
1334 hypothesis, Andrews *et al.* (2003) have shown that
1335 children’s ability to reason relationally and their ability
1336 to reason about unobservable mental states is closely
1337 linked, both computationally and ontogenetically (see
1338 also Halford *et al.* 1998; Zelazo *et al.* 2002). Thus, the
1339 ability to encode *ms* variables via an f_{ToM} is probably
1340 inseparable, both computationally and phylogeneti-
1341 cally, from the ability to reason about the relational
1342 similarity between complex behavioural patterns and
1343 higher-order causal relations.
1344

(c) Take-home lessons from the proposed experimental protocols

1345 The key point to be taken from the two protocols
1346 proposed herein is not that they constitute an acid test
1347 for an f_{ToM} in a chimpanzee or corvid, or that failure on
1348 these tests would be demonstrative evidence of an
1349 absence of an f_{ToM} . Rather, they are a direct response to
1350 the concern that success in any predictive paradigm can
1351 be explained as the result of a behaviouristic psycho-
1352 logical system rather than mental, intervening variables
1353 (e.g. Andrews 2005). If this concern were true, then the
1354 entire project of testing non-human animals’ ability to
1355 use an f_{ToM} to predict the behaviour of their
1356 conspecifics would be experimentally intractable and
1357 otiose. While this concern applies to virtually all other
1358 experimental protocols to date, the present proposals
1359 are existence proofs that experimental protocols can be
1360 constructed that could provide positive, principled
1361 evidence for the predictive function of an f_{ToM} in non-
1362 verbal organisms.

Q8 We hope our proposed protocols also put to rest the
1363 worry that an f_{ToM} has no functional, adaptive value or,
1364 worse, may be a figment of our folk psychological
1365 imagination. Regardless of our doubts concerning the
1366 ontological status of the hypothetical entities posited by
1367 our folk psychology, it is clear to us that the ability to
1368 cognize the world from the cognitive perspective of
1369 another agent would provide an animal with enormous
1370 advantages over and above the ability to reason in terms
1371 of observable first-person relations alone. Our pro-
1372 posed experiments set forth two artificial examples of
1373 how the value of such an f_{ToM} might manifest itself.
1374 Hundreds of experimental studies with young children
1375 have shown that they are able to solve the kind of tasks
1376 that require an f_{ToM} in the sense defined herein (e.g.
1377 Meltzoff *in press*); and see Wellman *et al.* (2001) for a
1378 review and meta-analysis). And there are good reasons
1379 for believing that the traditional hallmarks of human
1380 cognition, language and culture, are intimately depen-
1381 dent on f_{ToM} systems of various kinds (for example,
1382 Bloom 2000, 2002; Tomasello *et al.* 2005). The
1383 problem is not that a ToM system has no value or is
1384 experimentally intractable; the problem is that there is
1385 still no evidence that non-human animals possess
1386 anything remotely resembling one.
1387
1388
1389
1390
1391
1392

8. UNCITED REFERENCES

Q9 Churchland & Churchland (1996) and Godfrey-Smith
1393 (2000).
1394

The theoretical work developed in this essay was generously
1395 supported by a James S. McDonnell Foundation Centennial
1396 Fellowship to DJP.
1397
1398
1399

ENDNOTES

1400
1401
1402 ¹Of course, not all comparative researchers believe that non-human
1403 animals are cognitive agents in the sense defined by equation (2.1).
1404 But all comparative researchers who believe that non-human animals
1405 are potentially capable of possessing an f_{ToM} must necessarily believe
1406 that these same animals are cognitive agents in the sense defined by
1407 equation (2.1) above.

²NB: it is not necessary for there to be a deterministic relation between
1408 the observable and the unobservable variables. Our argument holds,
1409 *mutatis mutandis*, whenever $P(\mathbf{b}_1|C_1) > P(\mathbf{b}_1|\sim C_1)$ or, indeed,
1410

anytime a probabilistic model (e.g. Bayesian) can predict b_1 on the basis of observable cues and past conditional dependencies without taking the value of r -state_i into account.

³Hare *et al.* used two metrics, 'retrieve' and 'approach', to measure the animals' performance on these tests. The first recorded the percentage of food items actually retained by the subordinate. The second recorded the percentage of trials on which the subordinate left its own chamber and crossed into the middle chamber prior to the dominant being released. As Karin-D'Arcy & Povinelli (2002) note, given the fact that the dominant chimp often did not know where the food was located and given the fact that the subordinate was given a sizeable headstart, it is hardly meaningful that the subordinate retrieved more food. Thus, as an important and overlooked point of scholarship, the approach metric was not statistically significant in the Misinformed condition of experiment 1, or in any of the other experiments reported in Hare *et al.* (2001).

⁴These glosses are not meant to suggest that corvids are constrained to simple conditional rules. We believe that corvids, like many other non-human animals, are perfectly capable of reasoning about the world in a flexible manner, albeit only with respect to observable first-person relations.

⁵Andrews' (2005) objection nevertheless suggests an interesting modification to the visor protocol. First, train the chimpanzees to (i) make a begging gesture in front of experimenters who can see them and (ii) to produce an auditory cue (e.g. stomping) in front of any experimenter who cannot see them (using the kind of seeing/not-seeing conditions developed by Povinelli & Eddy (1996b), such as bucket-over-head, blindfold on and back turned). In the transfer session, present the subject with a single experimenter wearing either the opaque or see-through visor and test whether or not the subject stomps or begs in front of that experimenter. Chimpanzees who have simply learned to stomp in response to an arbitrary set of perceptual cues (e.g. bucket-over-head, blindfold on, back turned), without any understanding of the underlying epistemic states involved will stomp regardless of the kind of bucket being worn. Chimpanzees who have cognized the physical conditions that result in 'seeing' and physical conditions that result in 'not-seeing' will beg and/or stomp from the experimenter with the see-through visor, but stomp in front of the experimenter with the opaque visor.

⁶One might ask why, given that chimpanzees do preferentially gesture to someone facing them as opposed to someone facing away, this is not *prima facie* evidence for an understanding of the perceptual state of seeing. The point to be clarified by the formalism of this paper is that immediate knowledge of how to respond to a social context is completely orthogonal to the question of whether the chimpanzee's underlying representation of the situation is comprised of r , p and ms variables, or r and p variables alone.

REFERENCES

Andrews, K. 2005 Chimpanzee theory of mind: looking in all the wrong places? *Mind Lang.* 20, 521–536.

Andrews, G., Halford, G. S., Bunch, K. M., Bowden, D. & Jones, T. 2003 Theory of mind and relational complexity. *Child Dev.* 74, 1476–1499. (doi:10.1111/1467-8624.00618)

Birch, S. A. J. & Bloom, P. 2004 Understanding children's and adults' limitations in mental state reasoning. *Trends Cogn. Sci.* 8, 255–260. (doi:10.1016/j.tics.2004.04.011)

Bloom, P. 2000 *How children learn the meaning of words*. Cambridge, MA: MIT Press.

Bloom, P. 2002 Mindreading, communication and the learning of names for things. *Mind Lang.* 17, 37–54.

Bugnyar, T. & Heinrich, B. 2005 Ravens, *Corvus corax*, differentiate between knowledgeable and ignorant competitors. *Proc. R. Soc. B* 272, 1641–1646. (doi:10.1098/rspb.2005.3144)

Bugnyar, T. & Heinrich, B. 2006 Pilfering ravens, *Corvus corax*, adjust their behaviour to social context and identity of competitors. *Anim. Cogn.* 9, 369–376. (doi:10.1007/s10071-006-0035-6)

Call, J. & Tomasello, M. 1999 A nonverbal false belief task: the performance of children and great apes. *Child Dev.* 70, 381–395. (doi:10.1111/1467-8624.00028)

Call, J., Hare, B. & Tomasello, M. 1998 Chimpanzee gaze following in an object-choice task. *Anim. Cogn.* 3, 23–34. (doi:10.1007/s100710050047)

Carruthers, P. & Smith, P. K. (eds) 1996 *Theories of theory of mind*. New York, NY: Cambridge University Press.

Churchland, P. M. & Churchland, P. S. 1996 The future of psychology, folk and scientific. In *The churchlands and their critics* (ed. R. N. McCauley). Cambridge, UK: Blackwell.

Clayton, N. S. & Emery, N. J. 2005 Corvid cognition. *Curr. Biol.* 15, R80–R81. (doi:10.1016/j.cub.2005.01.020)

Clayton, N. S., Griffiths, D. P., Emery, N. J. & Dickinson, A. 2001 Elements of episodic-like memory in animals. *Phil. Trans. R. Soc. B* 356, 1483–1491. (doi:10.1098/rstb.2001.0947)

Clayton, N. S., Dally, J. M. & Emery, N. J. 2006 Social cognition by food-caching corvids The western scrub-jay as a natural psychologist. *Phil. Trans. R. Soc. B* 362. (doi:10.1098/rstb.2006.1992)

Dally, J. M., Emery, N. J. & Clayton, N. S. 2006 Food-caching western scrub-jays keep track of who was watching when. *Science* 312, 1662–1665. (doi:10.1126/science.1126539)

Davies, M. & Stone, T. (eds) 1995 *Folk psychology*. Oxford, UK: Blackwell Publishers.

Davies, M. & Stone, T. (eds) 1995 *Mental simulation*. Oxford, UK: Blackwell.

Dennett, D. C. 1987 *The intentional stance*. Cambridge, MA: MIT Press.

Dretske, F. I. 1988 *Explaining behavior*. Cambridge, MA: The MIT Press.

Emery, N. J. 2004 Are corvids 'feathered apes'? Cognitive evolution in crows, jays, rooks and jackdaws. In *Comparative analysis of minds* (ed. S. Watanabe). Tokyo, Japan: Keio University Press.

Emery, N. J. & Clayton, N. S. 2001 Effects of experience and social context on prospective caching strategies by scrub jays. *Nature* 414, 443–446. (doi:10.1038/35106560)

Emery, N. J. & Clayton, N. S. 2004 The mentality of crows: convergent evolution of intelligence in corvids and apes. *Science* 306, 1903–1907. (doi:10.1126/science.1098410)

Emery, N. J. & Clayton, N. S. 2005 Evolution of the avian brain and intelligence. *Curr. Biol.* 15, R946–R950. (doi:10.1016/j.cub.2005.11.029)

Flombaum, J. I. & Santos, L. R. 2005 Rhesus monkeys attribute perceptions to others. *Curr. Biol.* 15, 447–452. (doi:10.1016/j.cub.2004.12.076)

Gentner, D., Holyoak, K. J. & Kokinov, B. K. (eds) 2001 *The analogical mind: perspectives from cognitive science*. Cambridge, MA: MIT Press.

Godfrey-Smith, P. 2000 On folk psychology and mental representation. In *Representation in mind: new approaches to mental representation* (ed. H. Clapin, P. Staines & P. Slezak), pp. 147–162. Amsterdam, The Netherlands: Elsevier.

Goldman, A. 1993 The psychology of folk psychology. *Behav. Brain Sci.* 16, 15–28.

Goodall, J. 1990 *Through a window*. Boston, MA: Houghton Mifflin.

Gordon, R. 1986 Folk psychology as simulation. *Mind Lang.* 1, 158–171.

Gordon, R. 1996 'Radical' simulationism. In *Theories of theories of mind* (ed. P. Carruthers & P. K. Smith), pp. 11–21. Cambridge, UK: Cambridge University Press.

Halford, G. S., Wilson, W. H. & Phillips, S. 1998 Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. *Behav. Brain Sci.* 21, 803–864. (doi:10.1017/S0140525X98001769)

- 1537 Hare, B. 2001 Can competitive paradigms increase the
1538 validity of experiments on primate social cognition?
1539 *Anim. Cogn.* **4**, 269–280. (doi:10.1007/s100710100084)
- 1540 Hare, B., Call, J., Agnetta, B. & Tomasello, M. 2000
1541 Chimpanzees know what conspecifics do and do not see.
1542 *Anim. Behav.* **59**, 771–785. (doi:10.1006/anbe.1999.1377)
- 1543 Hare, B., Call, J. & Tomasello, M. 2001 Do chimpanzees
1544 know what conspecifics know? *Anim. Behav.* **61**, 771–785.
1545 (doi:10.1006/anbe.2000.1518)
- 1546 Heyes, C. M. 1998 Theory of mind in nonhuman primates.
1547 *Behav. Brain Sci.* **21**, 101–148. (doi:10.1017/S0140525X
1548 98000703)
- 1549 Hostetter, A. B., Cantero, M. & HONKINS, W. D. 2001
1550 Differential use of vocal and gestural communication by
1551 chimpanzees (*Pan troglodytes*) in response to the atten-
1552 tional status of a human (*Homo sapiens*). *J. Comp. Psychol.*
1553 **115**, 337–343.
- 1554 Hunt, G. R. 1996 Manufacture and use of hook-tools by New
1555 Caledonian crows. *Nature* **379**, 249–251. (doi:10.1038/
1556 379249a0)
- 1557 Hunt, G. R. 2004 The crafting of hook tools by wild New
1558 Caledonian crows. *Proc. R. Soc. B* **271**, S88–S90.
- 1559 Hurley, S. 2006 Introduction. In *Rational animals?* (ed
1560 M. Nudds & S. Hurley). Oxford, UK: Oxford University
1561 Press.
- 1562 Hurley, S. & Chater, N. (eds) 2005 *Perspectives on imitation:
1563 from neuroscience to social science*. Cambridge, MA: MIT
1564 Press.
- 1565 Karin-D'Arcy, M. R. & Povinelli, D. J. 2002 Do chimpanzees
1566 know what each other see? A closer look. *Int. J. Comp.
1567 Psychol.* **15**, 21–54.
- 1568 Leavens, D. A., Hostetter, A. B., Wesley, M. J. & Hopkins, W. D.
1569 2004 Tactical use of unimodal and bimodal communication
1570 by chimpanzees, *Pan troglodytes*. *Anim. Behav.* **67**, 467–476.
1571 (doi:10.1016/j.anbehav.2003.04.007)
- 1572 Markman, A. B. & Dietrich, E. 2000 In defense of
1573 representation. *Cogn. Psychol.* **40**, 138–171. (doi:10.
1574 1006/cogp.1999.0727)
- 1575 Meltzoff, A. & Moore, M. K. 1997 Explaining facial
1576 imitation: a theoretical model. *Early Dev. Parenting* **6**,
1577 179–192. (doi:10.1002/(SICI)1099-0917(199709/12)6:3/
1578 4<179::AID-EDP157>3.0.CO;2-R)
- 1579 Meltzoff, A. In press. 'Like me': a foundation for social
1580 cognition. *Dev. Sci.*
- 1581 Nichols, S. & Stich, S. P. 2003 *Mindreading: an integrated
1582 account of pretence, self-awareness and understanding other
1583 minds*. Oxford, UK: Oxford University Press.
- 1584 Penn, D. & Povinelli, D. J. In press. Causal cognition in
1585 human and nonhuman animals: a comparative, critical
1586 review. *Annu. Rev. Psychol.* **58**.
- 1587 Povinelli, D. J. 2003 *Folk physics for apes*. Oxford, UK: Oxford
1588 University Press.
- 1589 Povinelli, D. J. & Eddy, T. J. 1996a Chimpanzees: joint visual
1590 attention. *Psychol. Sci.* **7**, 129–135. (doi:10.1111/j.1467-
1591 9280.1996.tb00345.x)
- 1592 Povinelli, D. J. & Eddy, T. J. 1996b Factors influencing young
1593 chimpanzees' (*Pan troglodytes*) recognition of attention.
1594 *J. Comp. Psychol.* **110**, 336–345. (doi:10.1037/0735-7036.
1595 110.4.336)
- 1596 Povinelli, D. J. & Eddy, T. J. 1996c What young chimpanzees
1597 know about seeing. *Monogr. Soc. Res. Child Dev.* **61**, i–vi.
1598 (doi:10.2307/1166159) 1–191.
- 1599 Povinelli, D. J. & Giambrone, S. 1999 Inferring other minds:
1600 flaws in the argument by analogy. *Phil. Top.* **27**, 167–201.
- 1601 Povinelli, D. J. & Vonk, J. 2003 Chimpanzee minds:
1602 suspiciously human? *Trends Cogn. Sci.* **7**, 157–160.
1603 (doi:10.1016/S1364-6613(03)00053-6)
- 1604 Povinelli, D. J. & Vonk, J. 2004 We don't need a microscope to
1605 explore the chimpanzee's mind. *Mind Lang.* **19**, 1–28.
- 1606 Povinelli, D. J., Bering, J. M. & Giambrone, S. 2000 Toward a
1607 science of other minds: escaping the argument by analogy.
1608 *Cogn. Sci.* **24**, 509–541. (doi:10.1016/S0364-0213(00)
1609 00023-9)
- 1610 Povinelli, D. J., Dunphy-Lelii, S., Reauxa, J. E. & Mazza,
1611 M. P. 2002 Psychological diversity in chimpanzees and
1612 humans: new longitudinal assessments of chimpanzees'
1613 understanding of attention. *Brain Behav. Evol.* **59**, 33–53.
1614 (doi:10.1159/000063732)
- 1615 Premack, D. & Woodruff, G. 1978 Does the chimpanzee have
1616 a theory of mind? *Behav. Brain Sci.* **4**, 515–526.
- 1617 Santos, L. R., Nissen, A. G. & Ferrugia, J. 2006 Rhesus
1618 monkeys, *Macaca mulatta*, know what others can and
1619 cannot hear. *Anim. Behav.* **71**, 1175–1181. (doi:10.1016/
1620 j.anbehav.2005.10.007)
- 1621 Seed, A. M., Tebbich, S., Emery, N. J. & Clayton, N. S.
1622 2006 Investigating physical cognition in rooks (*Corvus
1623 frugilegus*). *Curr. Biol.* **16**, 697–701. (doi:10.1016/j.cub.
1624 2006.02.066)
- 1625 Silk, J. B. 2002 The form and function of reconciliation in
1626 primates. *Annu. Rev. Anthropol.* **31**, 21–44. (doi:10.1146/
1627 annurev.anthro.31.032902.101743)
- 1628 Suddendorf, T. & Whiten, A. 2001 Mental evolution
1629 and development: evidence for secondary represen-
1630 tation in children, great apes and other animals.
1631 *Psychol. Bull.* **127**, 629–650. (doi:10.1037/0033-2909.
1632 127.5.629)
- 1633 Tebbich, S., Seed, A. M., Emery, N. J. & Clayton, N. S. In
1634 press. Non-tool-using rooks (*Corvus frugilegus*) solve the
1635 trap-tube task. *Anim. Cogn.*
- 1636 Tomasello, M. & Call, J. 1997 *Primate cognition*. New York,
1637 NY: Oxford University Press.
- 1638 Tomasello, M. & Call, J. 2006 Do chimpanzees know what
1639 others see- or only what they are looking at? In *Rational
1640 animals?* (ed S. Hurley & M. Nudds). Oxford, UK: Oxford
1641 University Press.
- 1642 Tomasello, M., Call, J., Nagell, K., Olguin, R. & Carpenter,
1643 M. 1994 The learning and use of gestural signals by young
1644 chimpanzees: a trans-generational study. *Primates* **35**,
1645 137–154. (doi:10.1007/BF02382050)
- 1646 Tomasello, M., Hare, B. & Agnetta, B. 1999 Chimpan-
1647 zees, pan troglodytes, follow gaze direction geo-
1648 metrically. *Anim. Behav.* **58**, 769–777. (doi:10.1006/
1649 anbe.1999.1192)
- 1650 Tomasello, M., Call, J. & Hare, B. 2003a Chimpanzees
1651 understand psychological states—the question is which
1652 ones and to what extent. *Trends Cogn. Sci.* **7**, 153–156.
1653 (doi:10.1016/S1364-6613(03)00035-4)
- 1654 Tomasello, M., Call, J. & Hare, B. 2003b Chimpanzees versus
1655 humans: it's not that simple. *Trends Cogn. Sci.* **7**, 239–240.
1656 (doi:10.1016/S1364-6613(03)00107-4)
- 1657 Tomasello, M., Carpenter, M., Call, J., Behne, T. & Moll, H.
1658 2005 Understanding and sharing intentions: the origins of
1659 cultural cognition. *Behav. Brain Sci.* **28**, 675–691. (doi:10.
1660 1017/S0140525X05000129)
- 1661 Visalberghi, E. & Tomasello, M. 1998 Primate causal
1662 understanding in the physical and psychological domains.
1663 *Behav. Processes* **42**, 189–203. (doi:10.1016/S0376-
1664 6357(97)00076-4)
- 1665 Vonk, J. *et al.* In press. Social and physical reasoning in
1666 human-reared chimpanzees.
- 1667 Wellman, H. M., Cross, D. & Watson, J. 2001 Meta-analysis
1668 of theory-of-mind development: the truth about false
1669 belief. *Child Dev.* **72**, 655–684. (doi:10.1111/1467-8624.
1670 00304)
- 1671 Whiten, A. 1996 When does behaviour-reading become
1672 mind-reading. In *Theories of theory of mind* (ed.
1673 P. Carruthers & P. K. Smith), pp. 277–292. New York,
1674 NY: Cambridge University Press.

1665	Whiten, A. 1997 The Machiavellian mindreader. In <i>Machiavellian intelligence II: extensions and evaluations</i> (ed. A. Whiten & R. W. Byrne), pp. 144–173. Cambridge, UK; New York, NY: Cambridge University Press.	1729
1666		1730
1667		1731
1668		1732
1669	Whiten, A. 2000 Chimpanzees and mental re-representation. In <i>Metarepresentations: a multidisciplinary perspective</i> (ed. D. Sperber), pp. 139–167. New York, NY: Oxford University Press.	1733
1670		1734
1671		1735
1672		1736
1673		1737
1674		1738
1675		1739
1676		1740
1677		1741
1678		1742
1679		1743
1680		1744
1681		1745
1682		1746
1683		1747
1684		1748
1685		1749
1686		1750
1687		1751
1688		1752
1689		1753
1690		1754
1691		1755
1692		1756
1693		1757
1694		1758
1695		1759
1696		1760
1697		1761
1698		1762
1699		1763
1700		1764
1701		1765
1702		1766
1703		1767
1704		1768
1705		1769
1706		1770
1707		1771
1708		1772
1709		1773
1710		1774
1711		1775
1712		1776
1713		1777
1714		1778
1715		1779
1716		1780
1717		1781
1718		1782
1719		1783
1720		1784
1721		1785
1722		1786
1723		1787
1724		1788
1725		1789
1726		1790
1727		1791
1728		1792

1793	Author Queries	1857
1794	<i>JOB NUMBER:</i> 20062023	1858
1795		1859
1796	<i>JOURNAL:</i> RSTB	1860
1797		1861
1798	Q1 We have inserted a short title. Please approve or	1862
1799	provide an alternative.	1863
1800	Q2 Additional keyword has been deleted as the journal	1864
1801	style permits only a maximum of six keywords. Please	1865
1802	approve.	1866
1803	Q3 We have inserted year for the reference Premack &	1867
1804	Woodruff's (1978). Please check and approve.	1868
1805	Q4 We have inserted year for the reference Whiten &	1869
1806	Suddendorf (2001). Please check and approve.	1870
1807	Q5 Please note that the symbols '<' and '>' have been	1871
1808	changed to single quotes through the article.	1872
1809	Q6 Please check the edit of the sentence 'On each trial,	1873
1810	the experimenter...'	1874
1811	Q7 Please check the edit of the sentence 'The initial	1875
1812	steps...'	1876
1813	Q8 Please check the sense of the sentence 'We hope our	1877
1814	proposed protocols...'	1878
1815	Q9 References Churchland & Churchland (1996) and	1879
1816	Godfrey-Smith (2000) are provided in the list but not	1880
1817	cited in the text. Please supply citation details or delete	1881
1818	the reference from the reference list.	1882
1819	Q10 Please provide page range for the reference Church-	1883
1820	land & Churchland (1996).	1884
1821	Q11 Please provide page range for the reference Hurley	1885
1822	(2006).	1886
1823	Q12 Please update the year of publication for the reference	1887
1824	Meltzoff (in press).	1888
1825	Q13 Please update the year of publication for the reference	1889
1826	Penn & Povinelli (in press).	1890
1827	Q14 Please update the year of publication for the reference	1891
1828	Tebbich et al. (in press).	1892
1829	Q15 Please provide page range for the reference Toma-	1893
1830	sello & Call (2006).	1894
1831	Q16 Please update the year of publication for the reference	1895
1832	Vonk et al. (in press).	1896
1833	Q17 Please check the inserted page range for the	1897
1834	reference Whiten (1996).	1898
1835		1899
1836		1900
1837		1901
1838		1902
1839		1903
1840		1904
1841		1905
1842		1906
1843		1907
1844		1908
1845		1909
1846		1910
1847		1911
1848		1912
1849		1913
1850		1914
1851		1915
1852		1916
1853		1917
1854		1918
1855		1919
1856		1920