1	RUNNING HEAD: Mental state reference in an inter-group context
2	
3	
4	Representing other minds: Mental state reference is moderated by group membership
5	
6	Jennifer Susan McClung 1,2, & Stephen David Reicher 2
7	1 Institute of Biology, University of Neuchâtel, 2 School of Psychology and Neuroscience,
8	University of St Andrews
9	
10	
11	
12	
13	
14	
15	Authors' Address:
16	
17	J. S. McClung
18	University of Neuchâtel, Centre for Cognitive Science, Institute of Biology
19	Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland
20	
21	S. Reicher
22	School of Psychology and Neuroscience, University of St Andrews, South Street, St Andrews
23	KY16 9JP, Scotland, UK
24	
25	
26	
27	
28	Corresponding Author:
29	
30	J. S. McClung
31	University of Neuchâtel, Centre for Cognitive Science, Institute of Biology
32	Rue Emile-Argand 11, 2000 Neuchâtel, Switzerland
33	Email: jennifer.mcclung@unine.ch; Phone: +41.32.718.3128

34 Abstract

35

36

37

38

39

40

41

42

43

44

45

46

47

48

49

50

51

52

53

54

The ability to infer the psychological forces that drive others' behavior is a cornerstone of human cognition. This 'theory of mind' (ToM) we have has been extensively studied in its developmental stages and non-human forms. However, how the fully developed theory of mind functions on a daily basis is still the focus of ongoing research. One capacity stemming from theory of mind involves overt linguistic mental state reference. We propose that, rather than being a capacity that those with a fully developed ToM use consistently, mental state reference is a function of our social relationship to others: specifically, whether the other is perceived as an in-group or out-group member. We therefore examined spontaneous mental state reference during casual conversation as a function of group membership. Participants were divided into 'in-group' or 'out-group' pairs using a classic minimal group paradigm. Next, they were allowed to converse casually with their partner without the experimenter present and then subsequently asked to describe their partner in a written format after interactions. We scored participants' conversations and their written descriptions of each other for frequency and complexity of mental state reference. Results showed that, when interacting with presumed out-group members, participants referenced their partners' mental states significantly less often than when interacting with presumed ingroup members. This effect was found both during conversations and in subsequent descriptions of the partner. Spontaneous mental state reference is apparently not a consistent psychological process but instead subject to social constructs, specifically group membership.

55

56

57 58

Key Words: theory of mind, inter-group, mental state attribution, mental state reference

Theory of mind, or the ability to infer unobservable mental states and to use these mental states to predict future behaviour, has long been investigated in its incomplete or premature forms. Both from a developmental perspective (Alison & Astington, 1988; e.g. Ensink & Mayes, 2010; Wellman, Cross, & Watson, 2001) and from an inter-species comparative perspective (e.g. Call & Tomasello, 2008; Heyes, 1998; Premack & Woodruff, 1978) the 'non-normal' theory of mind has been thoroughly canvassed, somewhat to the detriment of the study of the actual mechanism itself (Apperly et al., 2010; Apperly, Riggs, Simpson, Chiavarino, & Samson, 2006). The normally functioning adult theory of mind has received attention more recently in roughly the last decade. To date, evidence suggests that the use of a normally developed theory of mind (and a host of related processes) is heavily influenced by cognitive, cultural, and social factors.

First, in terms of the impact other cognitive processes have on theory of mind, people have difficulty interpreting another person's perspective without using their own knowledge as a template (e.g. Fussell & Krauss, 1991; Keysar, Barr, Balin, & Brauner, 2000; Keysar, Lin, & Barr, 2003; Nickerson, Baddeley, & Freeman, 1987; Royzman, Cassidy, & Baron, 2003). This process, sometimes termed epistemic egocentrism, can lead to misjudgments about other's knowledge and occurs even when people are motivated to make accurate inferences (Keysar, Ginzel, & Bazerman, 1995). People also encounter difficulties in interpreting others' visual perspective in the face of high executive demands and distractions (Epley, Morewedge, & Keysar, 2004; Lin, Keysar, & Epley, 2010), as well as with lower moods (Converse, Lin, Keysar, & Epley, 2008). Furthermore, Apperly and colleagues have shown that questions requiring theory of mind usage are answered less quickly than nonmentalistic, reality-matching questions (Apperly et al., 2006) an indication that theory of mind processes may be a function of cognitive processing demands.

Second, cultural differences also seem to play a role in how effectively people take another's perspective, as shown by a study in which Chinese and American participants were asked to infer a partner's visual perspective (Wu & Keysar, 2007). The Chinese participants inferred their partner's visual perspective more accurately than their American counterparts.

Third, social factors, specifically group membership, may also alter perception, making individuals less attendant to minds perceived as 'other' (Haslam, 2006) and more likely to stereotype those perceived as less similar (Ames, 2004). On the more extreme end,

people categorized as 'other', or out-group, may be infrahumanized and attributed fewer uniquely human emotions (Leyens et al., 2001). Dehumanization research shows similar effects, in that out-group members are attributed fewer human values and traits and more animalistic qualities than are in-group members (for a review see Haslam, 2006). Furthermore, Hackel and colleagues showed that out-group members are required to be more human to be perceived as having a mind, in that shared group membership impacted how participants perceived the presence of mind in not only actual humans but dolls as well (Hackel, Looser, & Van Bavel, 2014). Group membership also affects how people empathise with others: considerable research has shown that empathic responses are lowered when observing out-group compared to in-group members (for a review see Cikara, Bruneau, & Saxe, 2011).

The previously discussed evidence suggests that normal processing of others' mental states is neither automatic nor consistent, and that social factors play a role in how people attribute emotions, perceive the presence of minds, and empathically respond. If group membership can affect these processes related to theory of mind, we wanted to address whether it would also impact the mechanism in its most basic and original form, the attribution of mental states to others (Premack & Woodruff, 1978).

The present study attempts to address these gaps by investigating whether group membership affects one aspect of theory of mind usage - mental state reference - in cognitively normal adults during typical, daily interactions. Specifically, we were interested in whether group membership plays a role in how people spontaneously reference others' mental states. The aim of this study was to gather data from the most natural contexts possible: unlike previous studies, we did not want to prompt theory of mind usage but rather to gauge one of its natural manifestations. We therefore examined natural social interactions for evidence of one manifestation of theory of mind usage, spontaneous mental state reference during casual conversation. We began by categorizing pairs of people using a classic minimal group paradigm based on estimation abilities into in-group and out-group pairs (Tajfel, Billig, Bundy, & Flament, 1971). After categorization participants were allowed to freely converse with each other, after which we asked participants to describe their partner in a written format.

Our aim was to compare participants' conversations and their descriptions of each other to assess the impact of group membership on spontaneous mental state reference.

Linguistic reference to mental states has long been considered an indication of developmental

processes children go through as they learn to form and use mental representations of others' mental states (Piaget & Inhelder, 1966). Even the usage of simple mental state verbs, such as 'to remember' or 'to hope', requires that the user form a mental representation of the target's mental state, specifically, what is being remembered or hoped (Antonietti, Liverta-Sempio, Marchetti, & Astington, 2006). As such, this type of reference has been used to study children's developing theory of mind (e.g. Meins, Fernyhough, Johnson, & Lidstone, 2006; Meins et al., 2002). This link between language and theory of mind has been exploited by researchers to develop the 'Spontaneous Theory of Mind Protocol' (STOMP), which measures spontaneous descriptions of the mental states of characters in videos to show that it correlates with thickness of certain cortical areas of the brain (Rice & Redcay, 2014). The STOMP approach used trained coders to differentiate between physical and mental state reference, whereas our design used a finite list of mental state reference words, based on and including words used to study mental-reference in children (Jenkins, Turrell, Kogushi, Lollis, & Ross, 2003) as well as the 'state verbs' used in the linguistic category model (LCM) approach (Semin & Fiedler, 1988). We used this list to examine participants' conversations and their subsequent written descriptions of each other to compare how participants referenced the mental states of in-group and out-group partners.

143

126

127

128

129

130

131

132

133

134

135

136

137

138

139

140

141

142

144 Method

145146

Participants

147148

149

150

151

152

153

154

155

156

157

158

Participants were recruited using noticeboards around the University of St Andrews after the study was approved by the university's Ethics Committee. Participants from all departments except psychology were accepted to take part in the study. 86 female undergraduates (age range 17-20) took part in the study to form a total of 43 pairs. In this way we controlled for the gender of our participants in order to avoid gender effects, or the possibility that participants would use gender to categorise themselves on top of our group manipulation (Ito & Urland, 2003).

Two pairs of participants were discarded from analysis: one because one of the participants had previous experience with minimal group paradigms, and another because the recording equipment did not function during the trial. In total this produced 41 pairs: 21 in the out-group condition and 20 in the in-group condition. All participants were tested in a

single session lasting approximately 30 minutes. All participants were naïve to the experimental hypothesis, told that their data would be treated confidentially and used anonymously in publication, gave informed consent, were fully debriefed at the end of each experiment, and received £3 for participation.

Procedure

The experiment was conducted in the Social Immersion Lab in the Psychology Department at the University of St Andrews. Prior to the experiment it was confirmed that participants did not know each other in any way. Participants arrived at the lab at the same time and were given instructions before any social chatting could take place. Participants were given information forms describing the experiment and then asked to complete a consent form.

As the minimal group paradigm, participants were told the cover story that the experiment was designed to study the link between cognitive style and social interaction. To that end, the experimenter would first assess their cognitive style and then ask them to complete a social interaction task. Their cognitive style, they were told, would be assessed using a test called the 'Dot Estimation Task' (DET), which was in reality the minimal group paradigm used to categorize participants into out-group and in-group conditions (adapted from Howard & Rothbart, 1980). Participants were told that using the DET the experimenter would be able to tell whether they were over- or under-estimators, and that this categorization was significant since estimation abilities correlated with such abilities as spatial computation and mathematical skills. The DET itself involved estimating the amount of dots present on three consecutive pictures (made using Power Point, see Fig. 1 for example below). Dot pictures were presented for 3 seconds each using Microsoft Power Point and a projector.

Figure 1 about here

Figure 1. Representative illustration of a 'Dot Estimation Task' picture

Participants were asked to write down their estimates for each picture and to do the task alone in order to 'get a clear and true read-out' of each of their cognitive styles. In actual fact this request was designed to keep participants from discussing their answers and thereby

realizing that there was no actual correlation between estimates and assigned category. The experimenter then made a brief show of calculating the average of their estimations, and then arbitrarily assigned each participant to be either an over- or under-estimator. Participants were asked to wear a badge with their estimation type displayed on it (two 'over' or 'under' estimators in the in-group condition, and one of each in the out-group condition). Participants were told this was so that 'the experimenter would not forget who was what for future analysis' whereas it was in fact done to maintain the salience of the categorization. Again, this categorization was in reality arbitrary.

Once categorized, participants were told that they would take a short break from the experiment to allow the experimenter to set up the rest of the experiment before the social interaction. Participants were told they were allowed to chat to pass the time before the supposed upcoming social task if they wanted. The experimenter then left the room and allowed the participants 7 minutes to freely converse, during which time the CCTV camera system in the lab was recording. After the conversation, participants were separated and asked to 'solidify their impressions' of each other by completing a free-form written description of their partner before the supposed social interaction task. Participants were told that their written descriptions would remain anonymous and that their only purpose was to allow each participant to collect and focus their impressions of the other before the interaction task. After completing the written descriptions the experiment was concluded and participants were told that there was in actual fact no social interaction task. Participants were then debriefed in full and paid.

214 Coding

Previous research has shown that the use of mental state verbs provides a useful metric of theory of mind functioning (Bretherton, McNew, & Beeghly-Smith, 1981; Dunn, Bretherton, & Munn, 1987). Mental state verb usage is correlated with children's theory of mind development as measured by false belief tasks (Brown, Donelan-McCall, & Dunn, 1996). Mental state talk has been coded in a variety of ways, usually tailored to the study's particular aims. The aim of the current study was to quantify differences in linguistic manifestation of mental state reference as a function of our group manipulation. To this end, a master list of mental state words was devised to identify all references participants made to their partner's mental states.

To devise the master word list, we began with previous research which analysed spoken language to identify instances of mental state reference (e.g. Bartsch & Wellman, 1995; Jenkins et al., 2003; Shatz, Wellman, & Silber, 1983). The Jenkins list (2003) has come to be the standard list used in developmental research. However, the terms that constitute the Jenkins list are not exhaustive of all possible ways in which adults reference mental states. We therefore added to the Jenkins list the 'state verbs' identified by the Linguistic Category Model (Semin & Fiedler, 1988), except for those that did not fit the following rule: the word was required to make sense if and only if the concept of a mental state was invoked. The 'state verbs' that did not pertain solely to a mental state (but in some cases could describe a behaviour or a personality) and that were therefore excluded from our mental state term list were the following: *aggressive, charismatic, impulsive, moody, outgoing, reliable*, and *reserved*. For example, 'aggressive' could be used to describe an 'aggressive behaviour' without reference to the person as such. Hence, these aforementioned words were excluded from our coding system to avoid ambiguity.

The current corpus of transcribed speech showed that participants in the current study used many more expressions to reference mental states than those listed by both Jenkins et al (2003) and the LCM model. The first author and a second coder therefore submitted the transcribed corpus to a two-stage analysis to form a more complete master list. First, all terms that pertained to any mental state were identified by both the first coder (first author) and a second coder separately. This involved each coder combing the manuscript for statements that adhered to the aforementioned rule: the statement was required to make sense if and only if the concept of a mental state was invoked. When this rule was met, the term used was then added to the master list. We identified another 71 verbs used to reference mental states (e.g. to be interested in, to be pleased to, to learn., etc). See table 1 for complete master list. The two coders identified the same terms in the corpus except for three, 'to reckon', 'to find' (in the context of a sentiment or thought as opposed to an object), and 'to be only joking'. After discussion the two coders agreed that these three terms each fit the rule of necessitating the concept of a mental state to be understood and so were subsequently added to the master list.

Some terms are more ambiguous, in that they can be used to actively refer to a mental state or simply as conversation fillers that hold some social function. For instance, although conversation analysts previously treated the phrase 'you know' that exists without an object

as an ambiguous filler (G. Brown, 1977), more recent work with adult discourse analysis treats 'you know' as a referent to either shared knowledge (Edwards, 1997; Holmes, 1986; Potter, Hepburn, & Tileagă, 2011) or knowledge the recipient alone holds (P. Brown & Levinson, 1987). Therefore, in order to parallel recent content analysis techniques we have coded every use of the term 'to know' as reference to a mental state.

to want	to feel	to think	to mean
to hope	to be hurt	to know	to be serious
to wish	to be angry	to believe	to realise
to care	to be happy	to wonder	to recognise
to be pleased to	to be excited	to remember	to learn
to be tempted to	to love	to forget	to have an idea
to be interested in	to like	to guess	to be conscious of
to be bothered to	to dislike	to pretend	to imagine
to be keen on	to be afraid	to understand	to reckon
to look forward to	to enjoy	to expect	to fathom
to be bored of	to have fun	to have a clue	to figure (out)
to have a crush on	to be glad	to be confused	to plan to
to be mad about	to be mad	to notice	to lie
to desire	to be scared	to assume	to be sorry
to fancy	to be upset	to find out	to decide
to miss	to be surprised	to underestimate	to choose
to need	to fear	to agree	to trust
to enjoy	to be disgusted	to be sure	to be intelligent
to be fond of	to worry	to make sense	to be
to hate	to be anxious	to disagree	pessimistic/optimistic
to abhor	to be relieved	to be able to relate	to esteem
to detest	to be shocked	to judge	to admire
to loath	to be disappointed	to be determined	to find (in the sense of a
to hold in contempt	to be nervous	to be only joking	cognitive act, i.e. without
to prefer	to be sad	to accept	a physical object)
1			

to be glad to	to commiserate	to respect	to be into (not in the	
to dread	to envy	to suspect	physical sense)	
to pity	to mourn for	to intend		

Table 1. Master list of words used to reference mental states.

266

267

268

269

270

271

272

273

274

275

276

277

278

279

280

281

282

283

284

285

286

287

288

289

290

291

Once the master list was formed, coding was carried out on the entire corpus. The first coder coded the entire corpus of transcribed conversations and written descriptions while the second coder coded 20% of each. Both coders were blind to experimental condition and the second coder was blind to the hypothesis. First, each instance of mental state reference using any of the terms on the master list was identified in the corpus. This consisted of using a Microsoft Word XP 'find' function to locate every occurrence of each word on the master list (and all related grammatical forms). All grammatical forms were located by inputting the stem of a verb into the search function, or each grammatical form individually for irregular verbs. For example, to find all references to the mental state 'to want', the word 'want' was inserted into the search function and every instance of all grammatical forms, including to want, wanted, wanting, and will want, were all highlighted. The same procedure was applied to participants' written impressions of each other. To be considered an instance of mental state reference, the same rule was invoked requiring that the statement could be made sense if and only if the concept of a mental state was invoked. For example, in one written impression a participant wrote the following: 'She was nervous. Lots of nervous laughter'. The phrase 'She was nervous' was coded as a mental state reference (as it referred to the partner's mental state), while 'Lots of nervous laughter' was not coded as mental state reference since here the word 'nervous' did not refer to the partner's mental state but to the laughter itself.

Next, once all usages of the mental state reference terms were located, each was coded for complexity of mental state reference (as either 'basic', 'complex', or 'highly complex') along with whether the object of the mental state attribution in the utterance was the speaker or the partner (e.g. Baron-Cohen, 1989; Flavel, Botkin, Fry, Wright, & Jarvis, 1968; Wimmer & Perner, 1983). 'Basic' mental state reference is equivalent to what is sometimes called 'first order' mental state reference in that it involves only one degree of

intentionality and references one mental state alone (Dennett, 1987). For example 'You like chocolate?' contains reference to only one mental state and would qualify as a basic reference. Basic mental state references included such statements as 'What do you think of it?' (partner as object, trial 22) and 'and I'm not like a hideously messy person, I like to have a bit of mess' (self as object, trial 35). 'Complex' mental state reference is equivalent to what is also called 'second order' mental state reference: it involves a statement with two degrees of intentionality realised by using two mental states in reference to each other within the same expression (Dennett, 1987). This included such utterances as, 'ok, you have to think back to your first thoughts of me' (partner as object, trial 37) and 'I think I wasn't quite so sure before I came' (self as object, trial 1). 'Highly-complex' mental state reference (equivalent to Dennett's 'third order intentionality') combines three mental state terms in one utterance. Highly-complex mental state reference only occurred when participants were speaking about themselves, such as in the statement, 'I wish I knew what I wanted to do' (self as object, trial 5). Lastly, all references to both the self and the partner that did not involve the use of one of the mental state terms in table 1 were located and coded. These 'nonmentalistic' references were required to make sense if and only if either the speaker or the speaker's partner was invoked as the referent of the statement without the use of a mental state term from the master list.

Correlation between the two coders was assessed by calculating cohen's kappa for each of the categories of mental state and non-mentalistic reference in the conversations and in the written descriptions. Inter-rater reliability was high with all kappas, > 0.86. Given the small sample size, all of the complex and highly-complex instances of mental state reference were then coded by the second coder (still blind to condition and hypothesis) which produced a high inter-rater reliability (kappa = 1.00).

316

292

293

294

295

296

297

298

299

300

301

302

303

304

305

306

307

308

309

310

311

312

313

314

315

317 Results

318319

Data Analysis

320321

322

323

324

To assess the impact of group membership on spoken conversation and written descriptions we conducted generalized linear mixed models with the individual as the unit of analysis, the fixed effect set as group membership, and the random effect set as the pair that participants conversed in. All analysis was conducted using SPSS 24. Analyses were divided

into the following sections 3.1) Natural conversations: 3.2) Writen impressions, and 3.3) Correlation between mental state reference in conversations and written descriptions.

Natural Conversation Results

Analysis of participants' spontaneous conversations is divided into the following sections: A) 'Overall conversation', that is, amount of total speech as a function of group membership, B) 'Non-theory of mind reference', including all reference that does not involve attributing mental states to the partner, and C) 'Theory of mind like reference', including all references to the partner's mental states.

A) Overall conversation.

First, in comparing the overall amount of spoken conversation between conditions, we see that group membership had no effect on total words spoken: out-group condition (estimated marginal mean = 676.159) vs in-group condition (estimated marginal mean = 660.106; F(1, 80) = .46, p = .5). This is important because it indicates that any differences in mental state referencing cannot be ascribed to absolute differences in the overall amount of speech produced or the motivation to converse. However, we nevertheless analyse all subsequent differences in conversation as a function of percentage of total utterances for the sake of accurate comparability across conditions.

Also, to ensure that there was no effect of the specific categories used in our minimal group paradigm (over- and under-estimators) we also conducted GLZMs on in-group pairs alone to determine whether category type affected any of the referencing behaviours, including mental state reference and non-mental state reference. We found no effect of category type in any of the following analysis, in that over-estimators did not significantly differ from under-estimators along any dimension. We therefore pooled both types of estimators for the in-group condition in all analyses.

B) Non-'theory of mind' reference.

Regarding statements which require no theory of mind processing, we looked at the effect of group membership on non-mentalistic reference to the self and the partner, as well

as mentalistic reference to the self (using mental state verbs as defined above in the coding section). For example, non-mentalistic references to the self included things like, 'I tend to leave things 'til the last minute so...' (trial 40), or, 'I never saw a person faint in front of me' (trial 34). Non-mentalistic references to the partner included both statements and questions that referenced the partner without using the mental state verbs listed in table 1. For example, a non-mentalistic reference to the partner was, 'Do you have brothers and sisters that you left behind?' (trial 10), or 'You're not from the UK, though' (trial 36). Mentalistic reference to the self included statements like 'I *miss* not having animals around' (trial 10, basic reference), and 'I *knew* that like at 18 or whatever I couldn't *fathom* being in a different country from my parents' (trial 29, complex reference). Table 2 summarises non-theory-of-mind reference as a function of group membership.

	Estimated		
	Marginal Means (%	F(1,80) =	P value
	of total utterances)		
Non-mentalistic self-reference			
Out-group	18.00	6.838	.011 *
In-group	22.40		
Basic mentalistic self-reference			
Out-group	13.2	.442	.508
In-group	12.6		
Complex mentalistic self-			
<u>reference</u> Out-group	0.3	.405	.526
In-group	0.2		
Non-mentalistic partner-			
<u>reference</u> Out-group	6.7	4.067	.047*
In-group	8.7		

Table 2. Frequency of non-theory of mind reference to the self and partner.

Of the references participants made that required no theory of mind abilities, only the non-mentalistic references to the self and the partner differed significantly between in-group and out-group conditions. That is, shared group membership increased both non-mentalistic reference to the self and the partner. No significant differences were found regarding mentalistic reference to the self.

C) 'Theory of mind-like' reference to the partner.

We next analyzed the impact of group membership on how participants referenced their partners' basic mental states, indicated by the presence of any one of the words on the master list used in reference to the partner's mental state. First, participants talked about their partners' basic mental states less if their partner was an out-group member (estimated marginal mean = 1.2%) than a presumed in-group member (estimated marginal mean = 3.8%; F(1,80) = 205.634, p = .000, fig. 2). That is, statements that referred to a partner's basic mental states, such as 'So, would you like to be a lecturer?' (Trial 12) were more common when participants spoke with in-group than out-group members.

Figure 2 about here

Figure 2. Percentage of references to a partner's basic mental states in natural conversations (estimated marginal means \pm SE) as a function of group membership.

Given the infrequency of complex reference to the partner's mental states (eight times in our entire data set with no pair using this type of reference more than once) we used a simple Fisher's exact test to assess whether this type of reference was more likely to occur in the in-group or out-group conditions. In fact, participants referenced their partners' complex mental states less frequently when speaking with an out-group member (1 times total) than with an in-group member (7 times total; p < .02, Fisher's exact test). That is, statements such as, 'What you might *find* is that you *enjoy* it more' (trial 39) were more likely to be used between in-group members than between out-group members.

Figure 3 about here

Figure 3. Total references to a partner's complex mental states in natural conversations as a function of group membership.

Written Impressions Results

A) Overall amount of written words.

Questionnaire data from 6 pairs was unavailable as these pairs requested to leave early due to time constraints. There was no statistical difference between the total number of words participants wrote when describing an out-group partner (estimated marginal mean = 88.294) compared to describing an in-group partner (estimated marginal mean = 74.583; F(1, 68) = 2.578, p > .113). Again this is important as it shows that our minimal group manipulation did not affect participants' motivation to write about their partners.

B) Basic mental state reference to the partner.

Sentences written by participants were coded as either no mental state reference, basic mental state reference (one mental state verb in the sentence), or complex mental state reference (two mental state verbs in the sentence) according to the same coding scheme described above for conversation. Amount of both basic and complex mental state reference was calculated as a percentage of total sentences written. Out-group members were described less frequently in terms of their basic mental states (estimated marginal mean = 9.70%) than were in-group members (estimated marginal mean = 27.60%), F(1, 68) = 14.948, p = .000, fig. 4). For example, statements such as 'She doesn't *want* to fly into a conversation quickly' (Trial 4) were more likely to be written about in-group than out-group members.

Figure 4 about here

Figure 4. Percentage of reference to the partner's basic mental states in written impressions (estimated marginal means \pm SE) as a function of group membership.

C) Complex partner mental state reference.

Although participants used more complex forms of mental state reference more rarely, the same pattern was found. That is, participants writing about presumed out-group members referenced their complex mental states less frequently (estimated marginal mean = .5%) than those writing about presumed in-group members (estimated marginal mean = 3.6%; F (1, 68) = 8.457, p = .005; fig. 5). For example, participants were significantly more likely to write complex statements referring to a partner's mental states, such as 'She's not *afraid* to do what she *wants*' (Trial 28), about in-group than out-group members.

Figure 5 about here

Figure 5. Percentage of reference to the partner's complex mental states in the written impressions (estimated marginal means \pm SE) as a function of group membership.

Correlation Between Mental State Reference in Conversation and Written Descriptions

It is possible that the level of mental state reference participants manifest during conversation influenced their subsequent descriptions of their partner. That is, if people are provided with more mental state information about a person they may then be more prone to use that information preferentially over other information when describing a person, or they may simply have had less non-mentalistic information to hand with which to describe a person. Whichever is the case, it seems logical to assume that the conversations participants engaged in might have influenced the way in which they considered and described their partner afterward. However, we found no correlation between amount of total mental state reference in conversation and amount of total mental state reference in participants' subsequent descriptions of each other (R = .03, p = .97).

Discussion

In this study we aimed to determine whether people reference other's mental states automatically and consistently, or whether different social contexts, for example group membership, produce different referential behavior. To do this, we analyzed how people reference another person's mental states in both natural conversations and in subsequent

written descriptions as a function of artificially manipulated group membership. Broadly, our results showed that group membership affects how people overtly refer to another person's mental states, both immediately during an interaction and after it.

473

474

475

476

477

478

479

480

481

482

483

484

485

486

487

488

489

490

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

We first examined participants' casual conversations in order to examine the effect of group membership on spontaneous mental state reference as well as reference that would not require theory of mind processes. For non-theory-of-mind reference, it seems that shared group membership increases references to both the self and the partner that do not involve use of mental state verbs. So, for example, in-group participants discussed their current and past lives than out-group participants did. This could be because such references were aimed at uncovering shared interests to solidify or expand their shared identity, or simply due to participants in the in-group condition being more at ease with each other and hence more likely to discuss more personal information. Regarding reference that required mentalising, results revealed that, when interacting with an out-group member, participants referenced their partner's mental states less than when interacting with an in-group member. This effect held for both conversations during immediate interactions and for descriptions participants wrote about their partners even after interacting with them. Group membership also had the same effect on different levels of referential complexity, in that participants talking with an out-group member referenced both their basic and complex mental states less frequently than participants talking with an in-group member. Similarly, in their written descriptions, participants wrote less about out-group member's basic and complex mental states than they did about in-group members' basic and complex mental states. Importantly, as we show there is no correlation between mental state reference in conversation and mental state reference in written descriptions, this suggests that group membership is independently impacting these two processes. That is, we can rule out two alternative causes of increased mental-state terms in in-group participants' descriptions: a) that some individuals are simply more prone to overtly reference others' mental states regardless of the mode of reference, or b) that group membership impacts only spoken mental state reference which could have then primed in-group participants to increase their usage of mental state terms when describing their partner.

These results suggest that people are less likely to overtly reference other people's mental states if they perceive those others as out-group members. Importantly, this effect occurred despite the fact that groups were artificially created using arbitrary characteristics. That is, the groups in our study had no real-life group histories or prejudices between them

which could have led participants to stereotype or react to partners based on implicit associations (Greenwald, McGhee, & Schwartz, 1998; Stott & Reicher, 1998). This suggests that the difference in mental state reference shown here is a result of categorization based on group membership itself and not any other social or individual factors. However, it is worth noting a caveat of our design, that our minimal group paradigm could have impacted perceived similarity (as a function of their assigned 'cognitive style'). There is currently conflicting evidence regarding the impact of perceived similarity on processes related to theory of mind. For instance, while the accepted notion was that perceived similarity increases empathy (Davis, 1994) more recent studies have shown that, in fact, basic drives to nurture and protect have a bigger impact on empathy than perceived similarity (Batson, Lishner, Cook, & Sawyer, 2005). Future research will be able to dissociate the impact of shared group membership and perceived similarity on empathy and other processes linked to theory of mind (as well as extend the sample to males to examine any possible gender differences of this specific manifestation of theory of mind).

Our results also build on and extend research based on the Linguistic Category Model (LCM) (Semin & Fiedler, 1991). First, we have extended the LCM 'state verb' list to provide a more complete list of terms which people use to reference mental states. Secondly, we have extended the application of this type of natural conversation and description analysis. The LCM has been used to elucidate the typical inter-group biases shown in a vast variety of behavioural studies. For example, Maass and colleagues (Maass, Salvi, Arcuri, & Semin, 1989) showed that people used more abstract and dispositional terms to describe positive in-group behaviour and negative out-group behaviour (which was interpreted as furthering in-group favouritism and out-group discrimination). Along the same lines, Fiedler and colleagues (Fiedler, Semin, & Finkenauer, 1993) showed that people used more stereotypic language across all five of the LCM's defined categories when describing a gender out-group than a gender in-group, again suggesting a mode by which out-group discrimination is perpetuated. The current results are the first, however, to show that the actual quantity of mental state reference used in interaction and description differs as a function of the partner's group membership. As such our results expand on the LCM literature by suggesting both cognitive and behavioural mechanisms by which out-group discrimination and, conversely, in-group favouritism are facilitated.

While our results have shown an effect of group membership on mental state reference as a behavior, we cannot conclusively say which cognitive process this stems from. That is, even though mental state reference is commonly used as a proxy to gauge both mental state attribution and theory of mind usage (Schwanenflugel, Henderson, & Fabricius, 1998; Schwanenflugel, Martin, & Takahashi, 1999), strictly speaking, a difference in this type of referential behavior does not necessarily stem from a difference in underlying mental state attribution. That is, participants may or may not be attributing mental states to their partners regardless of how they overtly reference them in conversation. For example, participants may be truly unconcerned with an out-grouper's mental states and not even 'turn on' their theory of mind when interacting with them. Alternatively, theory of mind processes may be alive and active even with out-group members but people may choose not to show evidence of such during interaction, specifically with out-groupers, by actively suppressing reference to such processes. The explicit link between linguistic reference and actual mental representation, along with the degree to which participants consciously engage (or disengage) their theory of mind, remains to be shown by future research. However, taking other recent research into account we would speculate that theory of mind cognition is curtailed at a more fundamental level during interactions with out-group members, prior to any linguistic reference in conversation. For example, a recent study has shown that during joint interaction people fail to spontaneously form mental representations of out-group members (McClung, Jentzsch, & Reicher, 2013). In this study, people who did a computerized joint action task with a perceived in-group member showed altered reaction times due to the computational demands of mentally representing their partner, whereas people who did the task with a perceived out-group member reacted as if they were doing the task alone and evidenced no alteration in reaction times. This result suggests that the less socially relevant out-group member does not warrant mental representation even on a subconscious, unintentional level, which has implications for the current study. While speculative, a lack of basic mental representation of the out-group may be the source of the diminished linguistic reference to out-group members' mental states: without initial representation of a person one could hardly use their theory of mind to develop representations of their mental states. That is, if out-groupers are not even perceived as potentially intentional beings in the first place, then this may consequently affect more complex psychological processes, including the inhibition of theory of mind processes (Tomasello, Carpenter, Call, Behne, & Moll, 2005).

537

538

539

540

541

542

543

544

545

546

547

548

549

550

551

552

553

554

555

556

557

558

559

560

561

562

563

564

565

566

567

By showing that people reference mental states differently depending on the group membership of their target, our results extend the growing body of literature which suggests that actual theory of mind usage is not automatic or consistent. For instance, recent research suggests that understanding another's perspective or feelings does not happen spontaneously but instead requires effortful cognitive adjustment (e.g. I. A. Apperly et al., 2010; Epley, Keysar, Van Boven, & Gilovich, 2004; Nickerson, 1999). One benefit of the current study is that the differential mental state reference we showed occurred totally spontaneously in actual social interactions. Much of the previous research on theory of mind-related processes is based on artificial paradigms in which participants are required to mentalise about characters in a cartoon strip (e.g. Converse et al., 2008) or simply allowed to mentalise about a cartoon character (Apperly et al., 2006). In the current study participants interacted naturally and spontaneously with actual people without any prompts as to the direction of their attention.

This decrease in reference to out-group members' mental states may also have a range of consequences for 'real-world' social interactions. First, we would speculate that the effects we show as a result of minimal categorisation would also carry over to real-life groups. Given that such groups would be based on more relevant identities than minimal groups, it would seem a logical consequence that they would also be more motivated to maintain their bonds, which sharing more intimate knowledge of each other could facilitate. Conversely, between members of different groups, decreased mental state reference may facilitate out-group discrimination, even at its extremes of infrahumanization and dehumanization (Haslam & Levy, 2006; Levens et al., 2001). For example, even if we assume that theory of mind is functioning normally, without overt reference to a person's mental states, it may become easier to perceive such a person as less than human, or at least to treat them as such. However, the causal direction of such a link is as yet unclear: further research is needed to clarify the link between overt mental state reference and different aspects of out-group discrimination. For instance, the simple perception of a common identity during an interaction may be sufficient to increase mental state reference, which may in turn prohibit out-group discrimination entirely.

In sum, our results identify one social factor – group membership – as a major factor in how cognitively normal adults manifest their theory of mind usage, specifically in how they overtly refer to another person's mental states during natural conversation. However, this study does more than simply add to a list of contextual influences on different theory of mind-related processes. These results highlight the importance of group membership in the

cognition and behavior within our social world. We do not wish to claim that people never mentalize about the out-group. Our argument that mental state reference is contextual is just that: there are obvious contexts in which it would behoove a person to understand and reference the mental states of the out-group. Future research may identify other social factors (such as extreme power differentials) that may interact with categorization to impact mental state reference.

In conclusion, using the minimal group paradigm – a stripped down manipulation of social group context – the current study shows that people are less willing to talk about the mental states of anyone categorized 'other', and less willing to consider them even afterward in mentalistic terms. In sum, whether a person merits overt mental state reference seems to be a function of group membership.

Funding Sources

- This work was supported by a departmental stipend to the first author from the University of
- 622 St Andrews, St Andrews, UK.

Open Practices

- The research in this article earned Open Materials and Open Data badges for transparent
- 626 practices. Materials and data for the experiment are available at
- 627 https://figshare.com/articles/McClung_and_Reicher_Representing_Other_Minds_data/54016
- 628 <u>33</u>

646

647

648

649

654

655 656

657

658 659

- Alison, G., & Astington, J. W. (1988). Children's Understanding of Representational Change and Its Relation to the Understanding of False Belief and the Appearance-Reality Distinction. *Child Development*, *59*(1), 26-37.
- Ames, D. R. (2004). Inside the Mind Reader's Tool Kit: Projection and Stereotyping in Mental State Inference. *Journal of Personality and Social Psychology*, 87(3), 340-353. doi:10.1037/0022-3514.87.3.340
- Antonietti, A., Liverta-Sempio, O., Marchetti, A., & Astington, J. W. (2006). Mental language and understanding of epistemic and emotional mental states *Theory of mind and language in developmental contexts* (pp. 1-30): Springer.
 - Apperly, Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G. (2010). Why are there limits on theory of mind use? Evidence from adults' ability to follow instructions from an ignorant speaker. *The Quarterly Journal of Experimental Psychology*, 63(6), 1201-1217.
- Apperly, I. A., Carroll, D. J., Samson, D., Humphreys, G. W., Qureshi, A., & Moffitt, G.
 (2010). Why are there limits on theory of mind use? Evidence from adults' ability to
 follow instructions from an ignorant speaker. *Quarterly Journal of Experimental Psychology*, 63(6), 1201-1217.
 - Apperly, I. A., Riggs, K. J., Simpson, A., Chiavarino, C., & Samson, D. (2006). Is Belief Reasoning Automatic? *Psychological Science*, 17(10), 841-844. doi:10.1111/j.1467-9280.2006.01791.x
 - Baron-Cohen, S. (1989). The Autistic Child's Theory of Mind: A Case of Specific Developmental Delay. *Journal of Child Psychology and Psychiatry*, 30(2), 285-297.
 - Bartsch, K., & Wellman, H. (1995). *Children talk about the mind*. New York: Oxford University Press.
- Batson, C. D., Lishner, D. A., Cook, J., & Sawyer, S. (2005). Similarity and nurturance: Two possible sources of empathy for strangers. *Basic and applied social psychology*, 27(1), 15-25.
- Bretherton, I., McNew, S., & Beeghly-Smith, M. (1981). Early person knowledge as expressed in gestural and verbal communication: when do infants acquire a 'theory of mind'? In M. L. a. L. Sherrod (Ed.), *Social Cognition in Infancy*. Hillsdale, NJ: Erlbaum.
- Brown, Donelan-McCall, N., & Dunn, J. (1996). Why Talk about Mental States? The
 Significance of Children's Conversations with Friends, Siblings, and Mothers. *Child Development*, 67(3), 836-849. doi:10.1111/j.1467-8624.1996.tb01767.x
- Brown, G. (1977). *Listening to spoken English*: Longman London.
- Brown, P., & Levinson, S. (1987). Politeness: Universals in language usage. *Cambridge:* CUP.
- Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later.
 Trends in Cognitive Sciences, 12(5), 187-192.
- 676 Cikara, M., Bruneau, E. G., & Saxe, R. R. (2011). Us and them intergroup failures of empathy. *Current Directions in Psychological Science*, *20*(3), 149-153.
- Converse, B. A., Lin, S., Keysar, B., & Epley, N. (2008). In the mood to get over yourself: Mood affects theory-of-mind use. *Emotion*, 8(5), 725-730. doi:10.1037/a0013283
- Davis, M. H. (1994). *Empathy: A social psychological approach*: Westview Press.
- Dennett, D. C. (1987). The intentional stance. Massachusetts: MIT Press.

- Dunn, J., Bretherton, I., & Munn, P. (1987). Conversations about feeling states between mothers and their young children. *Developmental Psychology*, 23(1), 132-139. doi:10.1037/0012-1649.23.1.132
- 685 Edwards, D. (1997). Discourse and cognition. London and Beverly Hills, CA: Sage.
- Ensink, K., & Mayes, L. C. (2010). The Development of Mentalisation in Children From a Theory of Mind Perspective. *Psychoanalytic Inquiry*, *30*(4), 301-337.
- Epley, N., Keysar, B., Van Boven, L., & Gilovich, T. (2004). Perspective Taking as
 Egocentric Anchoring and Adjustment. *Journal of Personality and Social Psychology; Journal of Personality and Social Psychology, 87*(3), 327-339.
 doi:10.1037/0022-3514.87.3.327
- Epley, N., Morewedge, C. K., & Keysar, B. (2004). Perspective taking in children and adults:
 Equivalent egocentrism but differential correction. *Journal of Experimental Social Psychology*, 40(6), 760-768. doi:10.1016/j.jesp.2004.02.002
- Fiedler, K., Semin, G. R., & Finkenauer, C. (1993). The Battle of Words Between Gender
 Groups A Language-Based Approach to Intergroup Processes. *Human* Communication Research, 19(3), 409-441.
- Flavel, J. H., Botkin, P. T., Fry, C. L., Wright, J. W., & Jarvis, P. E. (1968). The development
 of role-taking and communication skills in children. New York: Wiley.
- Fussell, S. R., & Krauss, R. M. (1991). Accuracy and bias in estimates of others' knowledge.
 European Journal of Social Psychology, 21(5), 445-454.
 doi:10.1002/ejsp.2420210507
 - Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464-1480. doi:10.1037/0022-3514.74.6.1464
 - Hackel, L. M., Looser, C. E., & Van Bavel, J. J. (2014). Group membership alters the threshold for mind perception: The role of social identity, collective identification, and intergroup threat. *Journal of Experimental Social Psychology*, 52, 15-23.
- Haslam, N. (2006). Dehumanization: an integrative review. *Personality and Social Psychology Review*, 10(3), 252-264.

705

706 707

- Haslam, N., & Levy, S. R. (2006). Essentialist beliefs about homosexuality: structure and implications for prejudice. *Pers Soc Psychol Bull*, *32*(4), 471-485.
- Heyes, C. M. (1998). Theory of mind in nonhuman primates. *Behav Brain Sci*, *21*(1), 101-114; discussion 115-148.
- Holmes, J. (1986). Functions of you know in women's and men's speech. *Language in society*, *15*(01), 1-21.
- Howard, J. W., & Rothbart, M. (1980). Social categorization and memory for in-group and out-group behavior. *Journal of Personality and Social Psychology*, *38*(2), 301-310. doi:10.1037/0022-3514.38.2.301
- Ito, T. A., & Urland, G. R. (2003). Race and gender on the brain: Electrocortical measures of attention to the race and gender of multiply categorizable individuals. *Journal of Personality and Social Psychology*, 85(4), 616-626. doi:10.1037/0022-3514.85.4.616
- Jenkins, J. M., Turrell, S. L., Kogushi, Y., Lollis, S., & Ross, H. S. (2003). A Longitudinal Investigation of the Dynamics of Mental State Talk in Families. *Child Development*, 74(3), 905-920. doi:10.1111/1467-8624.00575
- Keysar, B., Barr, D. J., Balin, J. A., & Brauner, J. S. (2000). Taking perspective in conversation: the role of mutual knowledge in comprehension. *Psychological Science*, 11(1), 32-38.

- Keysar, B., Ginzel, L. E., & Bazerman, M. H. (1995). States of affairs and states of mind:
 The effect of knowledge of beliefs. *Organizational Behavior and Human Decision*Processes, 64(3), 283-293.
- 732 Keysar, B., Lin, S., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, 733 89(1), 25-41.
- Leyens, J. P., Rodriguez-Perez, A., Rodriguez-Torres, R., Gaunt, R., Paladino, M. P., Vaes,
 J., & Demoulin, S. (2001). Psychological essentialism and the differential attribution
 of uniquely human emotions to ingroups and outgroups. *European Journal of Social Psychology*, 31(4), 395-411.
 - Lin, S., Keysar, B., & Epley, N. (2010). Reflexively mindblind: Using theory of mind to interpret behavior requires effortful attention. *Journal of Experimental Social Psychology*, 46(3), 551-556. doi:10.1016/j.jesp.2009.12.019

739 740

747 748

- Maass, A., Salvi, D., Arcuri, L., & Semin, G. R. (1989). Language use in intergroup contexts: the linguistic intergroup bias. *Journal of personality and social psychology*, *57*(6), 981.
- McClung, J. S., Jentzsch, I., & Reicher, S. D. (2013). Group Membership Affects
 Spontaneous Mental Representation: Failure to Represent the Out-Group in a Joint
 Action Task. *PLoS ONE*, 8(11), e79178. doi:10.1371/journal.pone.0079178
 - Meins, E., Fernyhough, C., Johnson, F., & Lidstone, J. (2006). Mind-mindedness in children: Individual differences in internal-state talk in middle childhood. *British Journal of Developmental Psychology*, 24(1), 181-196.
- Meins, E., Fernyhough, C., Wainwright, R., Das Gupta, M., Fradley, E., & Tuckey, M. (2002). Maternal mind–mindedness and attachment security as predictors of theory of mind understanding. *Child Development*, *73*(6), 1715-1726.
- Nickerson, R. (1999). How we know—and sometimes misjudge—what others know:
 Imputing one's own knowledge to others. *Psychological Bulletin*, *125*(6), 737-759.
 doi:10.1037/0033-2909.125.6.737
- Nickerson, R., Baddeley, A., & Freeman, B. (1987). Are people's estimates of what other people know influenced by what they themselves know? *Acta Psychologica*, *64*(3), 245-259. doi:10.1016/0001-6918(87)90010-2
- Piaget, J., & Inhelder, B. (1966). *La psychologie de l'enfant*: Presses universitaires de France Paris.
- Potter, J., Hepburn , A., & Tileagă, C. (2011). INEQUALITY IN ACTION. *International Journal of Education and Psychology in the Community*, 19(2), 43-60.
- Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind?
 Behavioral and Brain Sciences, 1(04), 515-526.
 doi:doi:10.1017/S0140525X00076512
- Rice, K., & Redcay, E. (2014). Spontaneous mentalizing captures variability in the cortical thickness of social brain regions. *Social cognitive and affective neuroscience, 10*(3), 327-334.
- Royzman, E. B., Cassidy, K. W., & Baron, J. (2003). 'I Know, you know': Epistemic Egocentrism in children and adults. *Review of General Psychology*, 7(1), 38-65.
- Schwanenflugel, P. J., Henderson, R. L., & Fabricius, W. V. (1998). Developing organization
 of mental verbs and theory of mind in middle childhood: Evidence from extensions.
 Developmental Psychology, 34(3), 512-524. doi:10.1037/0012-1649.34.3.512
- Schwanenflugel, P. J., Martin, M., & Takahashi, T. (1999). The organization of verbs of
 knowing: Evidence for cultural commonality and variation in theory of mind. *Memory* & Cognition, 27(5), 813-825. doi:10.3758/bf03198534

- Semin, G. R., & Fiedler, K. (1988). The cognitive functions of linguistic categories in
 describing persons: Social cognition and language. *Journal of Personality and Social Psychology*, 54(4), 558.
- Semin, G. R., & Fiedler, K. (1991). The linguistic category model, its bases, applications and range. *European review of social psychology, 2*(1), 1-30.
- Shatz, M., Wellman, H. M., & Silber, S. (1983). The acquisition of mental verbs: A
 systematic investigation of the first reference to mental state. *Cognition*, 14(3), 301-321. doi:10.1016/0010-0277(83)90008-2
- Stott, C., & Reicher, S. (1998). How Conflict Escalates: The Inter-Group Dynamics of
 Collective Football Crowd `Violence'. *Sociology*, 32(2), 353-377.
 doi:10.1177/0038038598032002007
- Tajfel, H., Billig, M. G., Bundy, R. P., & Flament, C. (1971). Social Categorization and Intergroup Behavior. *European Journal of Social Psychology*, *1*(2), 149-177.
- 790 Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences*, 792 28(5), 675-+.
 - Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, 72(3), 655-684.

794

795 796

797

- Wimmer, H., & Perner, J. (1983). Beliefs About Beliefs Representation and Constraining Function of Wrong Beliefs in Young Childrens Understanding of Deception. *Cognition*, 13(1), 103-128.
- Wu, S., & Keysar, B. (2007). The effect of culture on perspective taking. *Psychological Science*, *18*(7), 600-606.