Spontaneous categorization along competence in partner and leader evaluations

Alexander Bor (alexander.bor@ps.au.dk)

in press Evolution and Human Behavior. DOI:10.1016/j.evolhumbehav.2017.03.006

1 Introduction

Maintaining reciprocal relationships is challenging, as there is uncertainty both about our partner's intentions and their competences. When investing in a partner, we are making two bets: that they are motivated to reciprocate *and* that they are able to do so. There is compelling evidence that humans have evolved cognitive mechanisms to inform the first bet. People routinely detect, avoid or even punish cheaters and free-riders in reciprocal relations (Cosmides & Tooby, 1992; Delton, Cosmides, Guemo, Robertson, & Tooby, 2012). We also take motivations into account when making decisions about who to help (Petersen, 2015). The evidence is less clear when it comes to the second bet regarding abilities and competences.

Research in psychology gives the impression that people are unskilled at distinguishing between competent and incompetent social partners. In a laboratory experiment, Delton and Robertson (2012) described targets who either contributed to the group's food pool or not and found no evidence that this information led to categorization. In a similar experiment, van Leeuwen, Park & Penton-Voak (2012) presented targets who varied both in their morality and competence, finding that respondents spontaneously categorized the targets along morality but not competence. Focusing on motivational systems related to these two dimensions, Petersen and colleagues found that incompetence does not trigger strong emotions, whereas a lack of motivation (intention) to cooperate does (Petersen, Sznycer, Cosmides, & Tooby, 2012; Petersen & Aarøe, 2013). In sharp contrast to these findings, however, research in political science on the evaluation of political leaders suggests that competence is a key character trait shaping evaluations (Funk, 1996, 1997; Kinder, Peters, Abelson, & Fiske, 1980; Todorov, Mandisodza, Goren, & Hall, 2005; Tyler & Degoey, 1996).

In short, there is a puzzle regarding the role of competence in social evaluations. First, there are discrepancies between theoretical expectations and the empirical evidence in psychology; and second, there are discrepancies between the findings of political science and psychological research. The present paper therefore seeks to answer the following research questions: Do people spontaneously categorize others along competence in social interactions? And under what circumstances, if any, are people particularly likely to categorize by competence?

1.1 The theoretical relevance of competence in leader and partner evaluations

Leadership and followership are designed to solve coordination and collective action problems in many species (King, Johnson, & Van Vugt, 2009). Humans in particular evolved rather sophisticated leader–follower relations, which are not limited to social coordination or information sharing but include intricate reciprocal service-for-prestige exchanges (Price & Van Vugt, 2014). Importantly, ancestral societies were egalitarian and leadership was distributed, which allowed the most competent person to emerge and lead the group to solve the problem at hand (Van Vugt, Hogan, & Kaiser, 2008). However, the evolutionary theory of leadership depends on the ability of followers to decide who to follow. Accordingly, there is a sizable literature arguing that followership psychology must be sensitive to variations in the competence of potential leaders (Price & Van Vugt, 2014; Van Vugt & Grabo, 2015; Van Vugt & Kurzban, 2007). In fact, there is some empirical evidence that competent leaders receive higher evaluations from citizens (Funk, 1997; Popkin, 1991).

There is no theoretical reason, however, to believe that the relevance of competence is limited to leader evaluations. One clear example of this relates to food-sharing, which constitutes an essential form of cooperation (Gurven, 2005). The logic of food-sharing is as follows: If the diet of a group consists of items of high variability, such as large game, then on any given day only some hunters will provide food for themselves and their relatives. To avoid large fluctuations in calorie intake, group members may decide to pool their resources (Kaplan & Hill, 1985). The benefits of sharing outweigh the costs for most individuals, both due to the diminishing returns of consumption and because starvation can lead to death. Sacrificing the second portion of foraged meat is worthwhile if it increases the likelihood that someone will save us from starvation in return.

Despite the simplicity of the underlying evolutionary logic, maintaining a social exchange relationship on the grounds of reciprocity is cognitively very demanding. First, there is the issue of monitoring cheaters; those who do not intend on reciprocating. Second, and more importantly for the present case, it also requires cognitive mechanisms that "estimate the costs and benefits of various actions, entities, or states of affairs to oneself ... [and] to others" (Cosmides & Tooby, 1992, p. 177). There are at least three factors that render estimating potential benefits challenging: 1) the need to account for potential future exchanges, 2) the need to account for multiple currencies and 3) the need to account for supply and demand in an intricate biological market of cooperative relationship.

The first challenge is that calculations must extend beyond the current exchange, especially because, ancestrally, most interactions were delayed due to the absence of a universal medium of exchange, such as money. In other words, humans require an idea of what benefit another person is likely to grant them in the future. Staying with the example of food sharing, helping a highly competent hunter could be extremely beneficial, even if it means incurring relatively large costs, as his future actions could potentially yield us large benefits. In environments in which future gains from a social exchange partner are largely a function of their competences, paying attention to competence benefits the individual.

The second challenge is to make estimations involving multiple currencies. Food could be honoured with information, protection, sexual access or a large number of other benefits (Gurven & Jaeggi, 2015). Given the existence of individual differences in competences and social roles, the algorithm estimating expected benefits is both important and complicated. In particular, this places a premium on gauging the unique and specific competences of others (see also Tooby & Cosmides, 1996). If one is not a great hunter, for example, it becomes particularly important to maintain a cooperative relationship with a competent hunter to secure a regular food supply and reciprocate via some other medium. Consistent with this, there is anthropological evidence that more competent hunters enjoy higher status and get more respect in the group (Von Rueden, Gurven, & Kaplan, 2008). In fact, the competence–status relationship goes beyond hunter–gatherer societies and is also observed in group stereotypes in modern societies (Cuddy, Fiske, & Glick, 2007). Gauging competence across multiple domains could thus increase chances of survival even for individuals with poor skills in some essential domain.

Finally, anthropological evidence suggests that competences in valuable domains can function as insurance. Studying two hunter–gatherer tribes (Yora and Shiwiar) in Peru, Sugiyama and Chacon (2000) found that if the best hunter gets ill or injured, several members of the group are willing to provide care – including food – for a prolonged period. This is remarkable inasmuch as individuals in such a scenario incur a very large cost even though the risk of the partner's death makes potential benefits highly uncertain. In an analysis of the underlying adaptive logic, this phenomenon is referred to as the Banker's Paradox: "When an ancestral hunter–gatherer is in most dire need of assistance, she becomes a bad 'credit risk' and, for this reason, is less attractive as a potential recipient of assistance" (Tooby & Cosmides, 1996, p. 131). Solving this paradox requires cognitive adaptations for identifying irreplaceable individuals. An individual becomes irreplaceable if they are able to provide benefits that other group members cannot or are unwilling to provide. Being irreplaceable need not be based on objective criteria. The best hunter is probably irreplaceable for several people in the group, but even a mediocre hunter may be irreplaceable from someone's perspective if they are the only person who shares food with them. Individuals can form deep friendships in which they become irreplaceable for each other (Sugiyama & Sugiyama, 2003). Needless to say, the resulting niche specialization requires an accurate representation of competence. This is the third and final challenge for cognitive estimations; they must be sensitive to the skills demanded and supplied in a biological market of cooperative relationships (Barclay, 2013). Individuals who could identify which skills were sought and by whom were more likely to choose, attract and maintain good partners.

These arguments suggest that the intricate social relationships we can observe in both leadership-followership and food-sharing predict the existence of sophisticated cognitive mechanisms for gauging the competence of social partners. To examine the existence of these mechanisms, four experiments using memory confusion protocol were conducted. The memory confusion protocol is an established experimental tool designed to tap into how people are categorized in social perception (Pietraszewski, 2016; Taylor, Fiske, Etcoff, & Ruderman, 1978). In these experiments, participants first receive information about the actions of eight targets. Following a brief distractor, they are asked to recall who performed

each action. As information about the targets is manipulated in such a way that four of them appear as competent and four as incompetent, errors made in the surprise recall phase reveal whether participants are more likely to confuse people with the same traits (i.e., same category). Evidence for such categorization effects implies that people make spontaneous distinctions relying on competence.

In the experiments, competence was operationalized as foraging skills. Delton and colleagues (Delton, Cosmides, Guemo, Robertson, & Tooby, 2012; Delton & Robertson, 2012) have conducted valuable research on social categorization related to foraging. Their experiments rely on the memory confusion protocol and describe a fictitious scenario in which the targets engage in social foraging after finding themselves on a deserted island. These studies highlight the importance of intentions and making an effort, but they provide no clear evidence that foraging competence is used to spontaneously categorize partners.[1] In the present experiments, this research is extended to demonstrate that reliable cues of competence are indeed categorized when properly operationalized. The experiments also rule out key alternative explanations for the observed findings.

2 Study 1: Variation in contribution to the common pool

Study 1 is a pilot replicating Delton and Robertson's (2012) Study 4. Studying social foraging, the authors convincingly argue that partners need to be judged by psychological factors that shape long-term behaviour, such as the costs they incur to gather food. This claim is not tested here further. However, the authors also propose that, due to the large effect of luck in foraging, contributions to the common pool on any given day are an unreliable indicator of long-term behaviour. Accordingly, they demonstrate that information regarding the amount of food a partner gathers and contributes to the common pool is not used to categorize targets and does not affect subsequent trait evaluations.

Importantly for this study, the amount of food gathered on a given day is a function of both luck and competence; it is not the most reliable cue of competence. Therefore, Delton and Robertson's (2012) null result may suggest both a general negligence to categorize partners according to competence or a tendency to neglect unreliable cues. In order to tease apart these explanations, it is valuable to start with a replication of the original finding to ensure its robustness. This also allows for a systematic transition to another experimental platform. The original study used laboratory experiments, whereas the

[1] Delton, Cosmides, Guemo, Robertson and Tooby (2012) find categorization between targets who contributed to a common food pool and targets who did not because they lost the food they found. This provides limited evidence for categorization along competence, as all of the targets were described as finding food (a cue of competence), and all targets were described as losing something (a cue of incompetence). The real difference between the groups, whether their incompetence affected the group or not, was due to undetermined reasons.

experiments presented in this paper are based on online experiments. Initially replicating the original findings on an online platform implies that any differences in subsequent studies are not a function of some difference between the lab and online contexts.

2.1 Methods[2]

2.1.1 Subjects

Using Amazon's Mechanical Turk (MTurk), 154 people living in the United States were recruited to participate in the experiment (M_{age} = 35.4 years, 56 females). A power analysis suggests that a sample this large has over 99% power to detect a medium effect size (d > 0.5) with a two-sided one sample t-test at 5% significance level. The experiment was administered in Qualtrics and took on average 8.5 minutes to complete. Participants were reimbursed with $0.55.

2.1.2 Procedure

Following the introduction of a scenario in which a handful of people emergency land on a deserted island and need to find food for the entire group, four targets are presented with diagnostic sentences describing a contribution of a large amount of food (e.g., "After someone else dropped a *great many juicy* oranges into a cave, he ventured into the dark alone to get them and bring them to camp"), whereas the other four targets are presented with sentences describing a collection of a minor amount of food (e.g., "He noticed a *few peaches* in the grove at the top of a perilous cliff and climbed up to gather them to bring back for the group"). The presentation of the sentences was randomized within a balanced order. The targets were represented with photographs of the same eight young white men used by Delton and colleagues (Delton, Cosmides, Guemo, Robertson, & Tooby, 2012; Delton & Robertson, 2012). Then, as a distractor, participants were asked to list as many countries as they could in 60 seconds. Finally, in the surprise recall phase, each action was presented and participants need to pick which of the simultaneously displayed eight targets they believed performed the given action.

Following the standards in the literature, correct answers from the surprise recall are neglected, whereas incorrect answers are identified either as a within-category error or a between-category error. A within-category error refers to an instance when an (in)competent action is attributed to one of the other three targets who executed an (in)competent action. In other words, the two confused targets belong to the same category. In a between-category error, the two confused targets are in different categories; a competent action is attributed to an incompetent target or *vice versa*. A categorization score is calculated as the difference between these two types of errors after correcting for different base-rates by multiplying the number of between-category errors with 0.75.

---

[2] Detailed experimental reports adhering to open science principles as defined by Gerber et al. (2014) are available in the supplementary materials for all four studies, which are uploaded to https://osf.io/h9jbu/ .

Descriptive statistics of the error rates for each study are reported in the supplementary materials. A mean categorization score significantly above zero signals that participants spontaneously categorize targets along this information. To aid comparison between the experiments, categorization scores are also translated into a measure of effect size (Pearson's *r*) ranging between -1 and 1. All test statistics reported in the paper are a result of two-tailed hypothesis tests.

This is a slightly revised version of the original experiment adapted to the constrains of an online platform following the guidelines set out by Petersen (2012). Notably, each target was only presented once, whereas each target appeared five times in the original design and performed actions diagnostic of their performance on two occasions. The authors' original 32 diagnostic sentences were balanced and pseudo-randomized into four batteries. Each participant saw only one battery of eight sentences. The results are reported for the pooled estimates. Although the restriction to provide only one action per target may weaken categorization effects as the categories are less rehearsed, it is a necessary sacrifice to take into consideration the shorter attention spans of participants in online experiments compared to laboratory subjects. Similarly, as a secondary set of dependent variables, the original authors report differences in average trait evaluations between the two groups. These additional measures are not part of this study.

2.2 Results

Study 1 provides no evidence for the categorization of targets regarding the amount of food contributed to the common pool. The categorization score is not statistically significant from 0 (M = -0.05, r = -0.02, p = 0.77). Reassuringly, the original estimate obtained by Delton and Robertson (2012) (M = -0.15, SD = 2.44, r = -0.06, p = 0.34) falls in the 95% confidence interval of this study (95% CI: [-0.22, 0.12]). The finding that the level of contribution to the common pool is not used to categorize partners thus appears robust. Moreover, the implementation of the experimental design to a new platform was successful, as the estimates are very similar despite the methodological changes between the two studies.

This result reinforces that the contribution to the common pool *per se* is not a reliable cue for long-term cooperative behaviour. If there is high *random* variation in foraging success, people may have many reasons not to contribute on any given day, and punishing someone for being unlucky is not an adaptive strategy. However, this also implies that contributions to the common pool may be a poor indicator of competence itself. In order to establish if people are able to utilize competence-relevant information, Study 2 employs a slightly modified design to distinguish more clearly between competent and incompetent targets.

3 Study 2: Variation in foraging competence

In Study 1, the variation in the contributions was due to luck. All targets showed considerable effort and skill in obtaining food, such as by "venturing into a dark cave" or "climbing a perilous cliff" – some of them just happened to find more food than others. Importantly however, the targets displayed little variation in actual foraging skills. To address more directly the role of competence, all sixteen sentences describing incompetent

targets were modified, so the reason for little contribution to the common pool was clearly a lack of skills (e.g., "In an attempt to collect many cups worth of honey, he clumsily moved through a dense cloud of angry bees, startled them, was attacked by the bees and ran away without any honey"). The sentences were pre-tested to ensure that incompetent targets are rated on average as less competent than the competent targets. The eight sentences providing the best contrast between competent and incompetent targets were selected and used in the subsequent experiment (on a scale from 1 to 7, $M_{competent} = 5.875$, $M_{incompetent} = 2.573$, $p < 0.001$).

3.1 Method

MTurk was used to recruit 151 people to participate in Study 2 ($M_{age} = 36.2$ years, 74 females). They spent 8.25 minutes on average on the task and were reimbursed with $0.55. The participants in Study 1 were not allowed to enrol. The procedure of the experiment was identical to Study 1, but the incompetent sentences were modified and only one battery of sentences was used.

3.2 Results

Participants categorized targets according to their foraging skills. Categorization scores are significantly above zero ($M = 0.63$, $t_{150} = 3.58$, $r = 0.28$, $p < 0.001$). This provides considerable evidence for the prediction that competence-relevant information is spontaneously used to categorize social partners. It is difficult to compare effect sizes across different experiments. That said, this categorization score is somewhat smaller than Delton and Robertson (2012) found in a laboratory setting (comparing cheaters and reciprocators) and is roughly similar to other memory-confusion protocol experiments executed online (comparing deserving and undeserving targets) (Petersen, 2012).

4 Study 3: Is competence categorization facilitated by leadership evaluation?

Study 2 demonstrated how potential foraging partners are spontaneously categorized according to their competence. In an attempt at investigating under what circumstances people are particularly likely to rely on this mechanism, Study 3 experimentally prompted leadership-evaluation psychology in half of the sample. This provides a test if – in line with theoretical expectations – competence categorization is in fact present or even facilitated in evaluating and selecting leaders. It also allows us to judge if the proposed importance of competence in leader evaluations has deep-rooted psychological foundations.

4.1 Method

MTurk was used to recruit 741 new participants from the U.S. in Study 3 ($M_{age} = 37.6$ years, 354 females). They spent 8 minutes on average on the task and were reimbursed with $0.50. All participants were randomly assigned to either the control or treatment group. Those in the former read a slightly simplified version of the introduction from Studies 1 and 2 and were prompted to "try to gain an impression of each individual". Subjects in the treatment group, however, were told in the introductions that the survivors on the island agreed to elect a leader the "next evening" and were instructed to "try to gain an

impression of who [they] would vote for to become the leader of the group". This framing is expected to activate psychological mechanisms relevant to leadership evaluation. To reinforce the framing, the treatment group was presented with the pictures of the eight targets and asked to select who they would vote for. This was administered following the distractor task but before the surprise recall. This procedure builds on previous works that demonstrated how leadership and partner evaluative psychologies can even be activated with simple prompts in fictitious scenarios (Antonakis & Dalgas, 2009; Laustsen & Petersen, 2015) and combines it with previous works on priming in the memory confusion protocol (Pietraszewski, 2016). Otherwise, the procedure for both groups was identical to the previous experiments. The sentences describing competent and incompetent behaviour were taken from Study 2.

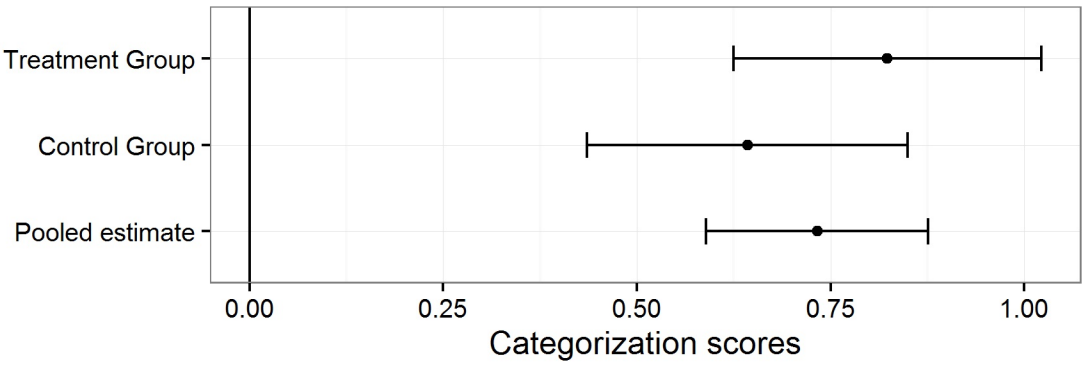

Figure 1. Categorization scores in Study 3.

4.2 Results

Figure 1 shows the categorization scores in the pooled sample and in both experimental groups separately. Positive categorization scores provide evidence that categorization along competence occurred both in the control (M = 0.64, t_{371} = 6.1, r = 0.30, p < 0.001) and treatment groups (M = 0.82, t_{368} = 8.15, r = 0.39, p < 0.001). Thus, Study 2 was successfully replicated on an independent sample. However, there is no evidence that a leader framing significantly increased the level of categorization (t_{738} = 1.24, r = 0.05, p = 0.22); this, despite the fact that there is good reason to believe that participants in the treatment group were manipulated effectively. An analysis of the actual leader choice of participants in the treatment group shows that 91% of the participants opted for a competent target as their preferred leader (χ^2 = 239, p < 0.001). Although potential leaders are categorized along their competence, this categorization is not stronger than in regular interactions.

5 Study 4: Variation on competence and likability

Study 3 showed that competence categorization is present to the same extent in the evaluation of general, social partners in the specific case of leadership evaluation. However, the evidence for competence categorization also being present in the evaluation of peers with equal social status is weak. Perhaps participants in the control condition implicitly assumed that leadership evaluation is relevant to the situation. As a further limitation, Study 3 only provided evidence of changes in the absolute levels of categorization. Target

traits were only manipulated on one dimension, competence, so little could be said about whether leader framing increases (relative) categorization according to competence as opposed to other traits. Study 4 provides remedies to both of these issues. Firstly in the control group, instead of a not specifying the social relationship between the target and the participant, it relies on a framing, which experimentally activates partner evaluation psychology. Secondly, it incorporates a second trait dimension – likability – orthogonal to competence, in order to test whether there are significant differences in the relative categorization between partner and leader evaluation (Kurzban, Tooby, & Cosmides, 2001; Petersen, 2012). Likability is a trait defining how easy it is to get along with someone. Likable targets were described as friendly, good-natured and understanding (Funk, 1997). Finally, as a secondary dependent variable, leader/partner preferences are measured to investigate differences between the two groups beyond implicit categorization.

5.1 Method

In Study 4, 402 people living in the United States were recruited on MTurk (M_{age} = 38.5 years, 233 females). None of the participants from the previous three studies or the pre-tests were allowed to enrol. The procedure was identical to Study 3 with two exceptions: First, a second set of sentences was added, describing four likable and four unlikable actions. A likable target is warm, easy-going or agreeable (e.g., "In the afternoon, he entertained a group of peers with his good sense of humour."), whereas an unlikable target is cold, annoying or pessimistic (e.g., "He refused to join a small game at the camp, saying that it was 'just silly'"). The sentences were pre-tested to ensure a significant difference between the manipulations (on a scale from 1 to 7, $M_{likable}$ = 5.61, $M_{unlikable}$ = 2.71, $p <$ 0.001). The likability sentences were pseudo-randomly paired with original (competence) sentences to ensure that the two dimensions are uncorrelated and counterbalanced. Two targets thus became both competent and likable, two targets were neither, and the remaining four were either competent or likable. Each target was presented once with the order of the sentences balanced and the target pictures randomly allocated. To calculate the categorization scores, first within- and between-category errors for the two trait dimensions are aggregated independently. Next, base-rates are corrected for and categorization scores are calculated as in the previous three studies.[3]

The second modification was that both groups received specific instructions framing their relationships with their partners. Whereas the leader framing group received the same prompts as in Study 3, the partner framing group was instructed to "Try to gain an impression of who [they] would prefer to spend time with from the group". Both groups were prompted to select a partner/leader after the distractor but before the surprise recall. Their choices constitute a secondary dependent variable tapping into partner and leader choice preferences.

---

[3] This correction method for a two-dimensional memory confusion paradigm has been used before by van Leeuwen, Park and Penton-Voak (2012).

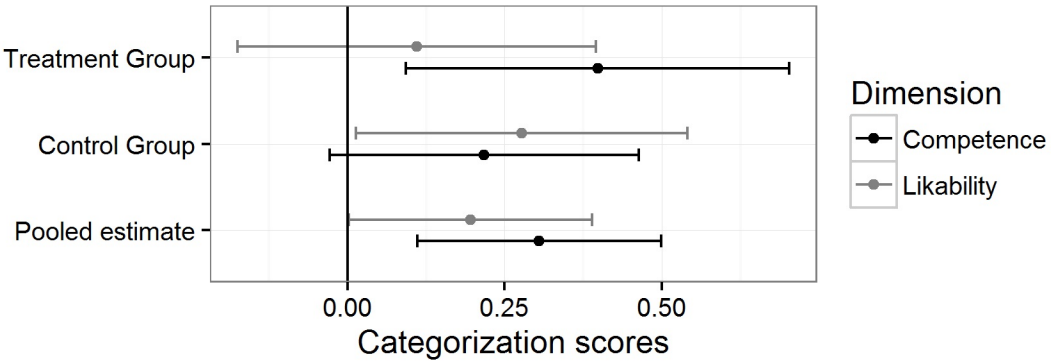

Figure 2. Categorization scores in Study 4.

5.2 Results

5.2.1 Categorization along competence and likability

First, considering competence categorization, the results are consistent with the previous studies. Targets are categorized on competence in the pooled sample and in both treatment groups, even in the presence of a second trait dimension, albeit to a lesser extent (pooled sample: M = 0.30, t_{401} = 3.09, r = 0.15, p < 0.01, partner framing group: M = 0.22, t_{206} = 1.74, r = 0.12, p = 0.08[4], leader framing group: M = 0.40, t_{194} = 2.57, r = 0.18, p < 0.05). However, there is no evidence of differential regulation of competence categorization (t_{378} = 0.91, r = 0.05, p = 0.37). Targets are categorized along competence equally strongly in both partner evaluation and leader evaluation. Secondly, there is also categorization along likability, although it is weaker than for competence (pooled sample: M = 0.20, t_{401} = 2, r = 0.1, p < 0.05, partner framing group: M = 0.28, t_{206} = 2.07, r = 0.14, p < 0.05, leader framing group: M = 0.11, t_{194} = 0.76, r = 0.05, p = 0.44). Priming participants' peer-evaluation or leader evaluation psychologies has no effect on likability categorization either (pooled sample: t_{395} = -0.85, r = 0.04, p = 0.40). The difference between the strength of the categorization by the two dimensions is not statistically significant in either group (paired t-test, pooled sample: t_{401} = 0.75, p = 0.46, partner framing group: t_{206} = 0.32, p = 0.75, leader framing group: t_{194} = 1.26, p = 0.21). All categorization scores are plotted in Figure 2. Study 4 thus demonstrates that competence categorization appears to be robust across partner and leader evaluations, even if multiple trait dimensions are present.

5.2.2 Partner and leader choice

Concerning partner and leader preferences, results show that in both groups a significant majority prefers a competent target as a partner (76%, χ^2 = 57.4, p < 0.001) and as a leader (86%, χ^2 = 101.9, p < 0.001). The leader framing causes a significant increase in

[4] This categorization score is significant at conventional levels if a one-tailed test is applied. The higher p-values in Study 4 (compared to Study 3) are partly the result of a loss of power due to a smaller sample size.

preferences for competent targets (χ^2 = 5.7, p < 0.05). Interestingly, similar tendencies are also present in the preferences for likable targets. Most of the participants (79%) chose likable targets in the partner framing group, and this tendency was significantly more pronounced (88%) in the leader framing group (χ^2 = 4.6, p < 0.05). In fact, the difference in the distribution of the two trait variables is statistically insignificant (χ^2 = 0.70, p = 0.87). An alternative way of interpreting the data is to look at what share of the participants pick a target high on both traits as a preferred partner/leader. It appears as though there is little trade-off between the two traits, and the majority of respondents pick a target both competent and likable (65% and 77% in the partner framing and leader framing groups, respectively – this difference is also statistically significant χ^2 = 9.4, p < 0.05). Importantly, these results are not driven by a single action description being more memorable than all others: the targets introduced as both competent and likable were selected substantially more often (44% and 27%) than chance (12.5%) would indicate. Thus, Study 4 also shows that people are rather competent in picking a competent and likable partner, which is even further improved if they face a leader choice dilemma.

6 Discussion

This research sought to establish if social partners are spontaneously categorized along cues revealing competence. A series of four experiments demonstrated that potential partners are spontaneously categorized as competent or incompetent in social interactions. Importantly, they show that categorization is present both in peer-to-peer and leader–follower interactions. Moreover, people's skills in gauging competence are not limited to implicit categorization, as they appeared to make ecologically rational partner and leader choice decisions. These results demonstrate the presence of an impressive, evolved cognitive mechanism in humans.

It is beyond the scope of the current paper to offer an explanation as to why previous studies showed no evidence of competence categorization effects. However, an interesting possibility worth noting is that previous experiments usually pitted competence against intentions. We know cognitive mechanisms regulating social exchange need to be extremely context-sensitive (Cosmides & Tooby, 1992). Perhaps competence categorization is inhibited at recall in environments in which there is a significant variation in effort (Pietraszewski, 2016). This could help to avoid situations in which someone is excluded from a reciprocal relationship despite being a valuable group member. The Ache men in Paraguay, for example, are very uncomfortable with discussing hunting competences, possibly in order not to interfere with their effort-based food-pooling schemes (Cosmides and Tooby, 1992, p. 214).

Finally, although no evidence was found that leader framings prompt higher levels of categorization, the findings demonstrate that categorization is also present in this domain and that this contributes to the ability to choose a competent leader. This result provides some indirect evidence for evolutionary leadership theories emphasizing the domain-specific nature of ancestral leadership. It may also imply some careful optimism about the ability of citizens to judge the competence of political candidates. More research is needed, however, to investigate how partner choice and leader choice psychologies differ. One

particularly promising direction could be to investigate additional forms of competence. This paper focused on foraging-related, task-specific competence (i.e., foraging skills), which is a particularly important set of competences over human evolutionary history. At the same time, other competences could be more specifically important for potential leaders (Price & Van Vugt, 2014); competences related to group defence or diplomacy, for example. It will be an important direction for future research to test whether categorization along leadership-specific competences show a stronger tendency to be up-regulated in leader versus peer categorization.

Acknowledgements

I am grateful to Michael Bang Petersen, Kim Mannemar Sønderskov, Florian van Leeuwen, Troels Bøggild, the editor and three anonymous reviewers for their insightful comments on earlier drafts of this paper.

Data Availability

The data, code and research materials associated with this research are available at https://osf.io/h9jbu/

Bibliography

Antonakis, J., & Dalgas, O. (2009). Predicting elections: child's play! *Science, 323*, 1183. http://doi.org/10.1126/science.1167748

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

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. In J. Barkow, L. Cosmides, & J. Tooby (Eds.), *The adapted mind: evolutionary psychology and the generation of culture* (pp. 163–228). New York, NY: Oxford University Press.

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2007). The BIAS map: behaviors from intergroup affect and stereotypes. *Journal of Personality and Social Psychology*, *92,* 631–648. doi:10.1037/0022-3514.92.4.631

Delton, A. W., Cosmides, L., Guemo, M., Robertson, T. E., & Tooby, J. (2012). The psychosemantics of free riding: dissecting the architecture of a moral concept. *Journal of Personality and Social Psychology, 102,* 1252–1270. doi:10.1037/a0027026

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: partner selection by underlying valuation. *Evolution and Human Behavior, 33,* 715–725. doi:10.1016/j.evolhumbehav.2012.05.007

Funk, C. L. (1996). The impact of scandal on candidate evaluations: an experimental test of the role of candidate traits. *Political Behavior, 18,* 1–24. doi:10.1007/BF01498658

Funk, C. L. (1997). Implications of political expertise in candidate trait evaluations. *Political Research Quarterly*, *50,* 675–697. doi:10.1177/106591299705000309

Gerber, A., Arceneaux, K., Boudreau, C., Dowling, C., Hillygus, S., Palfrey, T., Biggers, D. R., Hendry, D. J. (2014). Reporting guidelines for experimental research: a report from the experimental research section standards committee. *Journal of Experimental Political Science 1*, 81–98. doi:10.1017/xps.2014.11

Gurven, M. (2005). To give and to give not: the behavioral ecology of human food transfers. *Behavioral and Brain Sciences, 27,* 543–583. doi:10.1017/S0140525X04000123

Gurven, M., & Jaeggi, A. V. (2015). Food sharing. In R. Scott, & S. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences* (pp. 1–12). Hoboken, NJ: John Wiley & Sons. doi:10.1002/9781118900772.etrds0133

Kaplan, H., & Hill, K. (1985). Food sharing among Ache foragers: tests of explanatory hypotheses. *Current Anthropology*, *26,* 223–246. doi:10.1086/203251

Kinder, D. R., Peters, M. D., Abelson, R. P., & Fiske, S. T. (1980). Presidential prototypes. *Political Behavior 2,* 315–337. doi:10.1007/BF00990172

King, A. J., Johnson, D. D. P., & Van Vugt, M. (2009). The origins and evolution of leadership. *Current Biology, 19,* R911–R916. doi:10.1016/j.cub.2009.07.027

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences of the United States of America, 98,* 15387–92. doi:10.1073/pnas.251541498

Laustsen, L., & Petersen, M. B. (2015). Does a competent leader make a good friend? Conflict, ideology and the psychologies of friendship and followership. *Evolution and Human Behavior, 36*, 286–293. doi:10.1016/j.evolhumbehav.2015.01.001

Van Leeuwen, F., Park, J. H., & Penton-Voak, I. S. (2012). Another fundamental social category? Spontaneous categorization of people who uphold or violate moral norms. *Journal of Experimental Social Psychology*, *48,* 1385–1388. doi:10.1016/j.jesp.2012.06.004

Petersen, M. B. (2012). Social welfare as small-scale help: evolutionary psychology and the deservingness heuristic. *American Journal of Political Science, 56,* 1–16. doi:10.1111/j.1540-5907.2011.00545.x

Petersen, M. B. (2015). Evolutionary political psychology: the origins and structure of heuristics and biases in politics. *Advances in Political Psychology*, *36,* 45–78. doi:10.1111/pops.12237

Petersen, M. B., & Aarøe, L. (2013). Politics in the mind's eye: imagination as a link between social and political cognition. *American Political Science Review, 107,* 275–293. doi:10.1017/S0003055413000026

Petersen, M. B., Sznycer, D., Cosmides, L. & Tooby, J. (2012). Who deserves help? Evolutionary psychology, social emotions, and public opinion about welfare. *Political Psychology, 33,* 395–418. doi:10.1111/j.1467-9221.2012.00883.x

Pietraszewski, D. (2016). Priming race: does the mind inhibit categorization by race at encoding or recall? *Social Psychological and Personality Science*, *7,* 85–91. doi:10.1177/1948550615602934

Popkin, S. L. (1991). *The reasoning voter: communication and persuasion in presidential campaigns.* Chicago, IL: University of Chicago Press.

Price, M. E., & Van Vugt, M. (2014). The evolution of leader–follower reciprocity: the theory of service-for-prestige. *Frontiers in Human Neuroscience*, *8,* 1–17. doi:10.3389/fnhum.2014.00363

Von Rueden, C., Gurven, M., & Kaplan, H. (2008). The multiple dimensions of male social status in an Amazonian society. *Evolution and Human Behavior, 29,* 402–415. doi:10.1016/j.evolhumbehav.2008.05.001

Sugiyama, L. S., & Chacon, R. (2000). Effects of illness and injury on foraging among the Yora and Shiwiar: pathology risk as adaptive problem. In L. Cronk, N. Chagnon, & W.

Irons (Eds.), *Human behavior and adaptation: an anthropological perspective* (pp. 371–395). New York, NY: Aldine de Gruyter.

Sugiyama, L. S., & Sugiyama, M. S. (2003). Social roles, prestige, and health risk: social niche specialization as a risk-buffering strategy. *Human Nature, 14,* 165–190.

Taylor, S. E., Fiske, S. T., Etcoff, N. L., & Ruderman, A. J. (1978). Categorical and contextual bases of person memory and stereotyping. *Journal of Personality and Social Psychology, 36,* 778–793. doi:10.1037/0022-3514.36.7.778

Todorov, A., Mandisodza, A. N., Goren, A., & Hall, C. C. (2005). Inferences of competence from faces predict election outcomes. *Science*, *308,* 1623–1626. doi:10.1126/science.1110589

Tooby, J., & Cosmides, L. (1996). Friendship and the banker's paradox: other pathways to the evolution of adaptations for altruism. *Proceedings of the British Academy, 88,* 119–143.

Tyler, T. R., & Degoey, P. (1996). Trust in organizational authorities: the influence of motive attributions on willingness to accept decisions. In R. M. Kramer, & T. R. Tyler (Eds.), *Trust in organizations: frontiers of theory and research.* (pp. 331-357). Thousand Oaks, CA: Sage.

Van Vugt, M., & Grabo, A. E. (2015). The many faces of leadership: an evolutionary-psychology approach. *Current Directions in Psychological Science, 24,* 484–489. doi:10.1177/0963721415601971

Van Vugt, M., Hogan, R., & Kaiser, R. B. (2008). Leadership, followership, and evolution: some lessons from the past. *The American Psychologist, 63,* 182–196. doi:10.1037/0003-066X.63.3.182

Van Vugt, M., & Kurzban, R. (2007). Cognitive and social adaptations for leadership and followership: evolutionary game theory and group dynamics. In J. Forgas, M. Haselton, & W. Von Hippel (Eds.), *The evolution of the social mind: evolutionary psychology and social cognition*. (pp. 229-244). New York, NY: Psychology Press.