Prosociality, social tolerance and partner choice facilitate mutually beneficial cooperation in common marmosets, *Callithrix jacchus*

J. S. Martin [a, b, c, d, *], S. E. Koski [e], T. Bugnyar [b], A. V. Jaeggi [a, d], J. J. M. Massen [b, f]

[a] *Human Ecology Group, Institute of Evolutionary Medicine, University of Zurich, Zurich, Switzerland*
[b] *Department of Cognitive Biology, University of Vienna, Vienna, Austria*
[c] *Department of Anthropology, Miami University, Oxford, OH, U.S.A.*
[d] *Department of Anthropology, Emory University, Atlanta, GA, U.S.A.*
[e] *Faculty of Social Sciences, Anthropology, University of Helsinki, Helsinki, Finland*
[f] *Animal Ecology Group, Department of Biology, Utrecht University, Utrecht, The Netherlands*

Prosocial motivation has been suggested to facilitate the initiation and maintenance of cooperative interactions, as well as the evolution of social systems reliant on helping behaviour and social coordination. Previous comparative research on the link between prosociality and cooperation has been limited, however, by the absence of directly comparable measures of these traits among the same individuals. In the present study, we therefore examined intraspecific variation in prosociality and cooperative behaviour within a captive colony of group-living, cooperatively breeding common marmosets to provide a direct experimental test of these hypothesized benefits. We measured prosociality using a group service food-provisioning paradigm, and we assessed mutually beneficial dyadic cooperation with the loose string coordinated pulling paradigm. In addition, we also investigated the effects of individual social tolerance and partner choice, which have previously been identified as key factors promoting prosociality and cooperation among primates. As predicted, successful cooperation in the loose string paradigm was positively associated with prosociality, as well as with social tolerance and partner choice. These effects were independent of age, sex, personality, food motivation and learning across experimental sessions. Our results therefore suggest that prosocial motivation, social tolerance and partner choice can each facilitate mutually beneficial cooperation and social coordination in marmosets, supporting the hypothesized role of these mechanisms in the evolution of cooperative behaviour among primates.

© 2020 The Author(s). Published by Elsevier Ltd on behalf of The Association for the Study of Animal Behaviour. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

Human cooperation is exceptional among animals for its high degree of expression, complexity and plasticity (Fehr & Fischbacher, 2003; Melis & Semmann, 2010). Nevertheless, highly cooperative social interactions are also found throughout the natural world and have been experimentally demonstrated in diverse taxa ranging from mycorrhizal fungi (Kiers et al., 2011) and joint hunting fish (Bshary, Hohner, Ait-el-Djoudi, & Fricke, 2006), to ravens, *Corvus corax* (Massen, Ritter, & Bugnyar, 2015), Asian elephants, *Elephas maximus* (Plotnik, Lair, Suphachoksahakun, & de Waal, 2011), and chimpanzees, *Pan trogolodytes* (Melis, 2006). While the emergence and maintenance of cooperation has presented many explanatory challenges for social evolutionary theory

(Okasha, 2006), considerable progress has been made in identifying common evolutionary and ecological drivers of cooperative behaviour (Bourke, 2011; Rubenstein & Abbot, 2017; West, Griffin, & Gardner, 2007a), defined here as any behaviour that provides short- or long-term benefits to social partners (Bshary & Bergmüller, 2008). However, the proximate mechanisms underlying cooperative behaviour appear to be much more heterogeneous across species, as reflected by ongoing debates about whether cognitive demands can constrain the evolution of cooperation within particular lineages (Clutton-Brock, 2009; McNally, Brown, & Jackson, 2012; Ohtsuki, Hauert, Liebermann, & Nowak, 2006), or otherwise promote the evolution of distinct forms of social intelligence (González-Forero & Gardner, 2018; McNally et al., 2012; Moll & Tomasello, 2007). In addition to the specialized cognitive skills that may facilitate complex forms of sociality, such as inequity aversion (Brosnan & de Waal, 2003) or calculated reciprocity

* Corresponding author.
 E-mail address: jordan.martin@iem.uzh.ch (J. S. Martin).

(Dufour, Pele, Neumann, Thierry, & Call, 2009), cooperation also fundamentally requires attentional and motivational mechanisms (Akçay, Van Cleve, Feldman, & Roughgarden, 2009; Jaeggi, Burkart, & van Schaik, 2010), which can enhance and expand upon the behavioural capacities of pre-existing cognitive abilities (Finlay, Hinz, & Darlington, 2011; Morand-Ferron, Cole, & Quinn, 2016). For example, differential levels of circulating oxytocin and patterns of oxytocin receptor expression, which are strongly implicated in social motivation (McCall & Singer, 2012), have been linked to variation in social attention, recognition, learning and memory across a variety of mammals (Johnson & Young, 2017; Ziegler & Crockford, 2017). Tolerance towards social partners, partner preferences and other-regarding motivations are also key factors promoting social learning and coordination (Akçay et al., 2009; Asakawa-Haas, Schiestl, Bugnyar, & Massen, 2016; Crane & Ferrari, 2015; Schuster & Perelberg, 2004).

While the importance of social motivations for cooperative behaviour has been thoroughly demonstrated in human research (Fehr & Gintis, 2007; Kurzban, Burton-Chellew, & West, 2015; Tomasello, Carpenter, Call, Behne, & Moll, 2005), their role in explaining nonhuman animal (herein 'animal') cooperation remains controversial and less well established. This disconnect is due in part to the common assumption that psychological mechanisms do not constrain the expression of adaptive animal behaviour, which has been dubbed the 'behavioural gambit' (Fawcett, Hamblin, & Giraldeau, 2013). Despite the utility of this assumption for behavioural ecological research, studies have consistently shown that psychological processes such as motivation, attention and learning play a key role in explaining otherwise unexpected patterns of decision making and behavioural plasticity in both the laboratory and the wild (e.g. Clark & Dukas, 2003; Kedar, Rodriguez-Gironés, Yedvab, Winkler, & Lotem, 2000; McNamara & Houston, 2009). Understanding the structure and effects of these proximate mechanisms is therefore crucial for explaining the diversity of behavioural processes within and across animal species, which may otherwise be unaccounted for by evolutionary ecological theory alone.

One of the most well-studied social motivations in human and nonhuman primates is prosociality, which is defined as a motivation for benefitting or helping social partners, also known as an 'other-regarding' preference. Formally, prosociality is modelled as a parameter controlling the degree to which an individual weighs the observable payoffs of their social partners in their decision-making processes (Akçay et al., 2009). Prosocially motivated individuals are those who positively weigh the payoffs of others in their decisions and thus, all else being equal, are more likely to exhibit actions that increase the observed benefits accrued to social partners. Given that such motivational parameters cannot be directly observed, prosociality must instead be inferred from empirical variation in individuals' expected probability of acting to help or benefit others, as observed under standardized experimental conditions (Burkart et al., 2014; Jaeggi et al., 2010; Massen, Haley, & Bugnyar, 2020). This experimental approach to investigating prosociality is also employed in the present study.

Importantly, because prosociality is hypothesized to specifically regulate the weighing of social partner payoffs, it is expected that prosociality will increase the probability of exhibiting any cooperative behaviour that an animal understands as providing observable benefits to its social partners (Burkart, Hrdy, & van Schaik, 2009). Formal models suggest that generalized social motivations such as prosociality are most likely to evolve in socioecological contexts that generate mutually beneficial outcomes, as well as those that select for complementary action and role specialization, such as in contexts of group hunting, resource defence and cooperative offspring care (Akçay et al., 2009). Consistent with these

models, prosociality has been hypothesized to be a key proximate mechanism promoting the initiation and maintenance of cooperation across primate societies (Burkart et al., 2014; Callaghan & Corbit, 2018; Silk, 2007), as well as a central target of selection within social systems reliant on costly forms of helping behaviour and social coordination, such as the societies characteristic of Pleistocene hominins (Isler & van Schaik, 2012; Martin, Ringen, Duda, & Jaeggi, 2020) and other cooperatively breeding animals (e.g. Horn, Scheer, Bugnyar, & Massen, 2016; Massen et al., 2020).

Support for the generalized role of prosociality in regulating cooperative behaviour has been found across callitrichid monkeys, who often perform exceptionally well in cognitive tasks requiring social attention and coordination among group members, as compared to primates of similar brain and group size (Burkart et al., 2009). The so-called cooperative breeding hypothesis suggests that these findings can be explained by selection for prosociality to facilitate the helping behaviours characteristic of cooperatively breeding species such as callitrichids. Given that prosociality is expected to increase the probability of cooperative behaviour in general, it is also expected that prosociality will, all else being equal, positively associate with performance in any experimental task where individuals can generate food rewards or other benefits for their social partners (Burkart et al., 2014; Burkart et al., 2009; Hrdy, 2009; but see Burkart & van Schaik, 2016; Thornton & McAuliffe, 2015; Thornton et al., 2016). This hypothesis therefore predicts that individual prosocial motivation can explain both intra- and interspecific variation in outcomes such as social coordination and problem solving across a variety of cooperative domains beyond food provisioning and offspring care.

While many studies have investigated prosociality in animals, particularly primates (Cronin, 2012), its potentially generalized effects on problem solving and the capacity to achieve social coordination in cooperative contexts remain unclear. For example, although chimpanzees and ravens show a suite of cooperative behaviours in both behavioural experiments (e.g. Massen, Lambert, Schiestl, & Bugnyar, 2015, Melis, 2006) and their daily lives (Heinrich, 2009; Muller & Mitani, 2005; Samuni, Preis, Deschner, Crockford, & Wittig, 2018), they do not consistently exhibit evidence of prosocial motivation in experimental tasks (Lambert, Massen, Seed, Bugnyar, & Slocombe, 2017; Massen, Lambert et al., 2015, Melis, Engelmann, & Warneken, 2018; Tennie, Jensen, & Call, 2016). Cooperatively breeding primates (Burkart et al., 2014) and corvids (pinyon jays, *Gymnorhinus cyanocephalus*: Duque, Leichner, Ahmann, & Stevens, 2018; azure-winged magpies, *Cyanopica cyana*: Horn et al., 2016; Massen et al., 2020) have also been found to exhibit relatively high prosocial motivation, while cooperatively breeding meerkats, *Suricata suricatta*, have not (Amici, Mimó, von Borell, & Bueno-Guerra, 2017). Currently, however, interpretation of such findings is inhibited by a sole reliance on interspecific comparisons, particularly in the absence of directly comparable measures of prosocial motivation and cooperative behaviour among the same individuals. To our knowledge, no prior animal experiments have utilized intraspecific variation in these traits to directly examine whether prosociality predicts individuals' tendency to achieve social coordination and problem solving in a cooperative task.

In addition to prosociality, individual differences in social tolerance and partner choice have previously been found to explain variability in cooperative behaviour both within and among a variety of primate taxa (e.g. Hare, Melis, Woods, Hastings, & Wrangham, 2007; Jaeggi & Gurven, 2013; Kaigaishi, Nakamichi & Yamadam 2019; Melis, Hare, & Tomasello, 2006; Molesti & Majolo, 2016; Sabbatini, De Bortoli Vizioli, Visalberghi, & Schino, 2012). Partner choice effects have also been shown to reflect key hormonal mechanisms regulating social behaviour and

performance in cognitive tasks, such as in longtailed macaques, *Macaca fascicularis*, where cooperation with closely bonded individuals is accompanied by a reduction in cortisol levels (Stocker, Loretto, Sterck, Bugnyar, & Massen, 2020). Social tolerance refers to an individual's tendency to remain in proximity to a conspecific during potentially competitive situations, such as when an individual co-feeds on a limited resource near to a group member without attempting to displace them, while partner choice broadly describes the tendency of particular individuals to more frequently engage in cooperative acts together, as compared to other possible pairings within their social group. Importantly, some authors have hypothesized that prosociality may be a by-product of selection for increased social tolerance (Hare, Wobber, & Wrangham, 2012), in contrast to the predictions of the cooperative breeding hypothesis (Burkart et al., 2014). This suggests that the effects of prosociality on cooperation may be better explained by individual differences in social tolerance per se. The expression of helping behaviour is also influenced by partner preferences and relationship quality (Finkenwirth & Burkart, 2018; Martin & Olson, 2015), further suggesting that the apparent effects of prosociality on cooperation may be a by-product of partner choice in addition to social tolerance. Directly testing these predictions has remained difficult, however, due to the necessity of disentangling individual variation in these dimensions within a comparable experimental framework.

In the present study, we therefore experimentally compare prosocial motivation, social tolerance, partner choice and cooperative behaviour in common marmosets. Common marmosets are a particularly valuable model system for this investigation because they engage in extensive cooperative breeding (Digby & Barreto, 1993) and have also been found to exhibit relatively high prosociality in previous experimental studies (Burkart et al., 2014; Burkart, Fehr, Efferson, & van Schaik, 2007; Burkart & van Schaik, 2013). Moreover, marmosets have been found to coordinate their behaviour in joint action tasks (Miss & Burkart, 2018), providing the opportunity to examine how prosociality influences success in an experimental cooperative task requiring partner coordination.

To assess prosocial motivation, we tested individuals in a group service food-provisioning paradigm (Fig. 1a), which relied on a see-saw mechanism adaptation (Horn et al., 2016) of a paradigm previously tested in marmosets and other primates (Burkart et al., 2014; Burkart & van Schaik, 2013). In this paradigm, individuals could step on a platform to provision a group member on an adjacent platform without receiving a reward themselves (0/1 payoff). We further employed a cooperative pulling task known as the loose string paradigm (Hirata, 2003; Fig. 1b) to assess social tolerance, partner choice and dyadic cooperation, using an in-group testing procedure (Massen, Ritter et al., 2015). In contrast to the group service paradigm, the loose string paradigm required active coordination within a dyad, as two subjects needed to simultaneously pull on the ends of a tethered string to retrieve food rewards from a sliding platform. Given that success resulted in rewards for both partners (1/1 payoff), we interpreted this paradigm as measuring mutually beneficial cooperation. We predicted that individuals with higher prosocial motivation in the group service paradigm would exhibit higher rates of dyadic cooperation in the loose string paradigm, consistent with the hypothesized benefits of prosociality for achieving social coordination and cooperative problem solving in contexts where individuals can produce benefits for their social partners (Burkart et al., 2009). In addition, we predicted that individual social tolerance and partner choice would also influence dyadic cooperation, and we further tested whether prosociality had an independent effect on mutually beneficial cooperation after accounting for these factors. Finally, to examine the generalizability of these potential effects across individuals and social partners, we also controlled for personality differences in sociability and arousal, age, sex and food motivation, all of which have been found to affect primates' task motivation and performance in social behavioural experiments (e.g. Altschul, Wallace, Sonnweber, Tomonaga, & Weiss, 2017; Morton, Lee, & Buchanan-Smith, 2013; Wergård, Westlund, Spångberg, Fredlund, & Forkman, 2016).
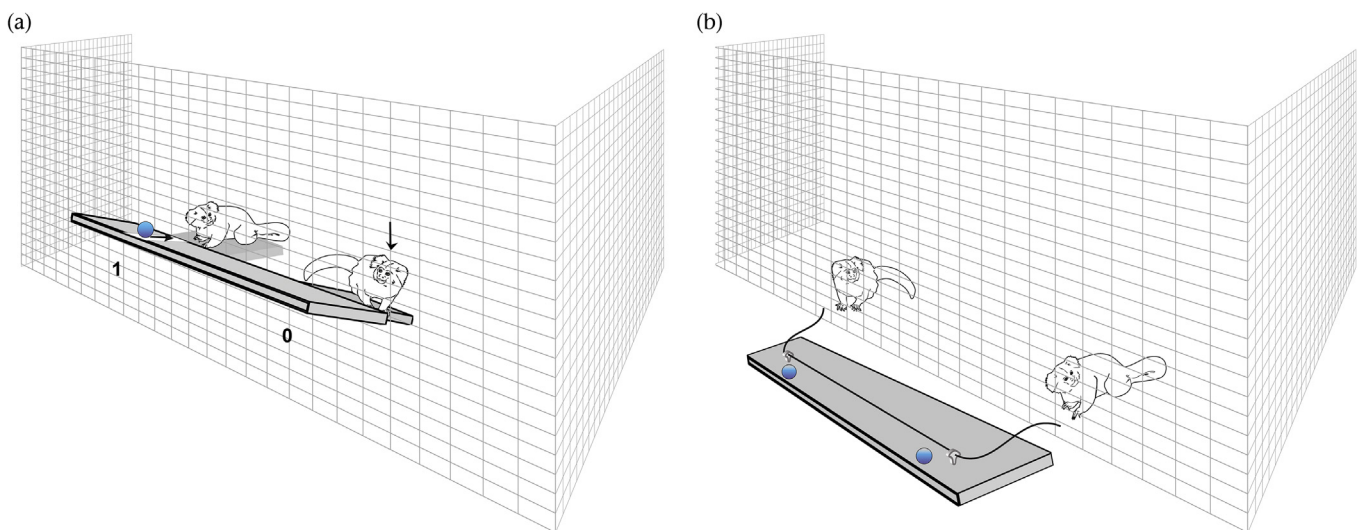

Figure 1. Experimental paradigms. (a) The group service paradigm. A see-saw mechanism was fixed to the enclosure of a social group with a provisioning platform (Position 0) extended into the enclosure, which could be stepped on to weigh the platform down — causing a food reward to roll towards the enclosure (Position 1). An additional platform not attached to the see-saw mechanism was placed within the enclosure to facilitate retrieval of this food reward. (b) The loose string paradigm. A moving board was stationed outside the enclosure of a social group, with a tethered string placed within arm's length of two platforms fixed within the enclosure. Two equally sized food rewards were placed on the board across from each platform. Simultaneous pulling on both ends of the string resulted in the board moving forward, facilitating retrieval of the food rewards. Conversely, uncoordinated pulling on either end would result in the string becoming untethered, preventing access to the food rewards.

METHODS

Subjects and Housing

We tested 23 common marmosets in five social groups (13 males; 10 females; age range 0–12 years of age, group size range 3–6 individuals; see Appendix, Table A1) at the University of Vienna. Each group enclosure was encased with thin wire mesh, contained opaque barriers to prevent visual contact between adjacent groups and facilitated free movement between indoor and outdoor housing areas (indoor: $250 \times 250 \times 250$ cm; outdoor: $250 \times 250 \times 250$ cm), which both contained ample supply of enrichment objects and bedding material. We maintained the laboratory temperature between $24\,°C$ and $26\,°C$, with $40–60\%$ humidity and individual heating lamps for each group, and we employed lighting lamps on a 12:12 h light:dark cycle in addition to natural sunlight. All groups had ad libitum access to water, and monkey pellets were provided daily in addition to nutritionally balanced meals in the morning and afternoon, along with weekly enrichment activities that provided mealworms and marmoset gum.

Ethical Note

This research was approved by the Animal Ethics and Experimentation Board, Faculty of Life Sciences, University of Vienna (No. 2018-013). All housing conditions accorded with Austrian legislation as well as the Callitrichidae husbandry guidelines of the European Association of Zoos and Aquaria. We conducted our experiments within group enclosures without separating any animals from their social groups so as to prevent unnecessary stress during testing. Subjects also received regular meals during the experimental period to prevent an undesirable increase in competition for food rewards and any subsequent social stress.

Experimental Paradigms

The order of paradigm presentation was counterbalanced across groups to control for order effects. As depicted in Fig. 1, all experimental sessions were conducted in the subjects' home enclosures with the entire social group present to enhance the ecological validity of the observed responses. As a result, individuals could spontaneously pair and engage in partner choice with any group members during testing, rather than being placed into specified dyads.

Group service paradigm

Our apparatus used a see-saw mechanism, such that stepping on a provisioning platform (Position 0) would weigh the see-saw down, causing food placed on the outside to roll down towards the fence and become accessible at an adjacent receiving platform (Position 1; see Fig. 1a). Crucially, a subject who landed at Position 0 would need to leave the provisioning platform if it wanted to access the food at Position 1 for itself, which would in turn cause the see-saw mechanism to revert to its original position with the food out of reach. Consequently, an individual was only able to provide food for someone else.

A multistage habituation and training procedure (Horn et al., 2016) was implemented prior to testing to ensure that subjects understood the basic mechanics of the apparatus. This was followed by two experimental phases composed of the group service task ('test' condition), during which individuals could provision their groupmates with a food reward without receiving a reward themselves, as well as two control conditions designed to account for potential effects of stimulus enhancement and food motivation

on provisioning behaviour. In the 'empty' control condition, the experimenter approached the apparatus and pretended to place food down on the Position 1 receiving tray; in the 'blocked' control condition, food was placed on the Position 1 tray, but access to the food was blocked by a transparent boundary at the receiving platform. Subjects stepping on the Position 0 platform could therefore cause the food to move closer to the enclosure but not provision their group members. Trial numbers per experimental session were determined by group size (i.e. $5 \times n$) to ensure adequate opportunity for participation across subjects. Trials were a maximum of 2 min in duration. Following previous studies (Burkart et al., 2010; Burkart & van Schaik, 2013; Horn et al., 2016), we first conducted three training sessions allowing the marmosets to learn about the full contingencies of both test and control set-ups beyond the basic see-saw mechanism trained in previous phases. This was followed by two experimental sessions of each condition, which were used for assessing prosociality. The blocked control was conducted after initial testing and empty control sessions, consistent with prior research (Burkart et al., 2014; Horn et al., 2016), while the test and empty control sessions were repeated sequentially for counterbalancing. Note that previous research has shown that no order effect occurs when marmosets are retested after the blocked control using the original group service paradigm (Burkart & van Schaik, 2016), nor for azure-winged magpies using the modified paradigm employed in our study (Horn et al., 2016). Please see Appendix for more detailed testing procedures.

Consistent with prior research (Burkart et al., 2007; Burkart et al., 2014; Burkart & van Schaik, 2013; Horn et al., 2016), we considered higher rates of stepping on the provisioning platform during the test compared to the control conditions as providing necessary evidence for interpreting provisioning behaviour as prosocially motivated. Thus, for subjects who clearly differentiated when they could and could not provision groupmates, we considered higher rates of stepping on the provisioning platform to be indicative of higher prosocial motivation. Note that this is a conservative criterion intended to strengthen our inferences about the motivational basis of observed provisioning behaviour, as all individuals participated in extensive training with the apparatus and thus exhibited a basic comprehension of the paradigm prior to test sessions. Therefore, while many individuals did not step on the provisioning platform frequently across any experimental sessions, it is likely that they nevertheless understood the task and simply were not motivated to provide food (or they may have been more motivated to retrieve rather than provision food). As described further below, some subjects also used the provisioning platform more during test sessions but exhibited very little prosocial motivation overall. In contrast, very few subjects used the provisioning platform frequently in both the test and control trials, suggestive of a stimulus enhancement effect (see Appendix, Fig. A3 and Table A3). All subjects were provided standardized daily meals outside of the testing period, and we therefore also used each individual's average number of retrieved food rewards across test sessions as an approximate measure of individual food motivation.

Loose string paradigm

Social tolerance assessment. Each social group was habituated to the apparatus (Fig. 1b) and underwent an initial social tolerance task used in previous research on cooperation in the loose string paradigm (Massen, Ritter et al., 2015). Two untethered strings were placed adjacent to one another on the apparatus across 18 sessions of 10 trials each, such that subjects could retrieve a food reward by pulling a string without the help of a partner. By observing how often subjects retrieved the food in proximity to each of their groupmates at the adjacent string, we were able to calculate a previously validated measure (Massen, Ritter, et al., 2015) of

average individual social tolerance (see Appendix). In addition, this phase facilitated training the subjects on the basic string-pulling mechanism, as acquisition of a food reward required a subject to pull in a single string from the apparatus towards itself. All subjects succeeded in accessing food in at least 9/18 training sessions, suggesting a basic comprehension of the string-pulling component of the task.

Dyadic task

Following this procedure, each group received 40 test sessions of 10 trials using the tethered string, with a maximum 1 min duration per trial. The rewards could not be accessed during these trials without simultaneous pulling by both partners on their respective ends of the string, thus requiring dyadic coordination in the timing of their pulling. We considered trials in which subjects were able to retrieve these rewards as instances of successful cooperation. In contrast to prior studies utilizing a single clump of food rewards (e.g. Melis, 2006), we distributed food evenly on the moving platform so that both partners could retrieve equal rewards. This allowed us to more directly disentangle the effects of individual prosociality and social tolerance on task performance. Moreover, given that prosociality is formally predicted to influence decision making irrespective of personal payoffs (Akçay et al., 2009), this design allowed us to investigate the unique effect of other-regarding preferences while holding the personal benefits of action constant across individuals. Furthermore, while previous research motivated our use of three training sessions across the group service conditions (Burkart et al., 2010; Burkart & van Schaik, 2013; Horn et al., 2016), no prior studies had yet been done on the loose string paradigm with common marmosets. We therefore did not designate a fixed number of additional training sessions with the full dyadic loose string task. Rather, we provided all groups with a large number of experimental trials (400 loose string trials) to facilitate ample opportunity for subjects in all groups to learn the temporal coordination necessary for the dyadic task (see Appendix, Fig. A5). As described below, we controlled for general learning across subjects in all analyses to effectively distinguish between the independent effects of task comprehension and the factors of interest on performance in the dyadic loose string task. Please see Appendix for more detailed testing procedures. Note that two older adult subjects did not consistently participate in either the social tolerance task or the dyadic loose string paradigm during the entire study period and were therefore excluded from data analysis due to insufficient evidence of task comprehension (see Appendix, Table A1).

Personality Scores

Personality scores from two factor dimensions describing individual differences in sociability (+allogrooming, +contact sitting, +social proximity) and arousal (+activity level, +gnawing, +scent marking) were available from previous research (Martin et al., 2019) on a subset of our sample (19/23 subjects). These scores were estimated using the Exploratory Graph Analysis + Generalized Network Modeling (EGA + GNM) statistical framework, from focal observational data collected within 3—7 months of the present study. See Appendix for further details on the structure and estimation of these personality dimensions. Note that we here refer to personality as temporally consistent among-individual variation in behaviour, following the behavioural ecological tradition in animal personality research (Dingemanse & Dochtermann, 2013). The sociability and arousal dimensions are therefore statistical constructs quantifying stable patterns of interindividual differences observed across multiple months of behavioural data. We investigated these traits in particular as

activity level is expected to predict neophilia among common marmosets (Koski et al., 2017), and previous work has shown sociability to reduce participation in experimental tasks among captive primates (Morton et al., 2013). In addition, many species exhibit increased personality similarity within both friendships and pair bonds (e.g. Gabriel & Black, 2012; Massen & Koski, 2014; Youyou, Stillwell, Schwartz, & Kosinki, 2017). This phenomenon, often referred to as homophily ('love of the same'), has been hypothesized to benefit dyadic cooperation through an enhanced capacity to coordinate and synchronize behaviour with phenotypically similar partners (Laubu, Dechaume-Moncharmont, Motreuil, & Schweitzer, 2016; Massen & Koski, 2014).

Statistical Analysis

Bayesian generalized linear mixed-effects models (GLMMs) were fitted for all analyses using the 'brms' package (Bürkner, 2017) for the R statistical environment (R Core Team, 2018). Weakly regularizing priors − $\beta \sim Normal(0, 2)$ for fixed effects, $\sigma \sim Half - Cauchy(0, 2)$ for random effects, and $R \sim lkj(2)$ for random effect correlations − were placed on all model parameters to penalize extreme estimates and reduce our risk of inferential error (McElreath, 2016). We modelled responses in the group service paradigm by investigating subject stepping rates at the Position 0 provisioning platform. To assess whether subjects exhibited prosociality in the experiment, we estimated a binomial model with experimental condition (test, control, blocked) as a fixed effect with accompanying random slopes across subjects. We further included a fixed effect for time of day to control for variation in performance across morning and afternoon sessions, as well as random intercepts to account for any unobserved heterogeneity across social groups and observations.

Cooperation in the loose string paradigm was investigated using both group level and dyad level success rates per experimental session. We assessed our central hypothesis by comparing rates of successful cooperation between dyads with and without subjects exhibiting prosociality in the group service task. The group level analysis therefore considered the total proportion of successful trials between prosocial and nonprosocial dyads within a group for each experimental session, while the dyad level analysis modelled the proportion of successful trials per experimental session for each possible prosocial and nonprosocial dyad within a social group. The total proportion of prosocial dyads within each group was controlled for in the group level analysis to ensure that the relative difference in success between prosocial and nonprosocial dyads was not confounded with the baseline probability of a prosocial dyad participating in the task. To examine the effects of partner choice across experimental sessions, we modelled dyadic cooperation in the prior session (number of successful trials/10 trials) as a fixed effect predictor of success in the subsequent session, capturing how well prior cooperation predicted future cooperation. To account for the effect of individual social tolerance on dyadic cooperation, we included an additive effect (i.e. sum of partner values) of individual social tolerance for each dyad. Additional exploratory fixed effects were also estimated for sex combination within a dyad, the additive effects of partners' age, sociability and arousal, as well as for partner similarity in age and personality − calculated on a 0−1 scale as $1/(1 + |value_{subject\ 1} - value_{subject\ 2}|)$. We also included the additive effect of food motivation within each dyad to ensure that the primary effects of interest were not confounded by experimental motivation across tasks, as well as time of day to control for variation in performance across morning and afternoon sessions. We included random intercepts across dyads and social groups to account for the multilevel structure of our repeated measures data, as well as so-called multimembership

random intercepts to appropriately handle repeated observations of individuals nested within multiple dyads (Browne, Goldstein, & Rasbash, 2001). Additional random intercepts were included to capture any unobserved heterogeneity across observations and days of experimentation independent of our fixed effects. A multiple imputation procedure (van Ginkel, Linting, Rippe, & van der Voort, 2019) was used to account for unreliable identification of two twin juvenile subjects during the loose string task. Additionally, given support for random missingness in a Bayesian imputation model, we used a simpler mean imputation procedure to account for missing personality scores in four subjects. Please see Appendix for further details on our statistical models and procedures.

Rather than relying on null hypothesis tests and discrete designations of statistical significance, we provide multiple measures to summarize and draw inferences from our posterior model estimates (McElreath, 2016; McShane, Gal, Gelman, Robert, & Tackett, 2019). In particular, to interpret the strength and uncertainty of estimated effects, we used the posterior median slope (i.e. the log odds for a 0−1 or 1 SD change; $\tilde{\beta}$), the median absolute deviation (MAD) as a robust measure of statistical uncertainty around the median, the 90% Bayesian credible interval (90% CI) and the probability of observing a positive or negative effect, i.e. the proportion of the posterior greater or smaller than 0 in the direction of the median (p_+ or p_-). Note that in contrast to classical P values, which consider the probability of observing data under a null hypothesis $p(\text{data}|H_0)$, the reported p_+ and p_- directly estimate the probability in support of hypothesized positive or negative effects given the observed data $p(H_1|\text{data})$. Larger values of p_+ or p_- therefore indicate greater support for positive or negative effects, respectively. In addition, we calculated posterior median and MAD estimates of Cohen's d for our fixed effects, which provides a standardized mean difference effect size with values of 0.2, 0.5 and 0.8 traditionally interpreted as small, medium and large effects, respectively. Finally, to avoid overfitting and enhance model generalizability, we used the fully Bayesian Watanabe−Akaike information criterion (WAIC; Gelman, Hwang, & Veharti, 2014) to select a final model of dyadic cooperation. In particular, we first estimated a model with all parameters included and compared it to a reduced model without fixed effect parameters exhibiting highly uncertain effects (p_+ or $p_- < 0.90$). As with other information criteria, $\Delta\text{WAIC}_{\text{full-reduced model}} = -2$ provides minimal support for the full model. Note that this procedure did not meaningfully change inferences about the primary effects of interest (see Appendix, Table A4).

RESULTS

Group Service Paradigm

Prosociality is evidenced by higher rates of stepping on the Position 0 platform during test trials, when groupmates could be provisioned, as well as relatively lower rates of stepping on the platform during the empty and blocked controls that assessed potential stimulus enhancement effects. Rates of stepping on the Position 0 provisioning platform were moderate to high across all conditions (average rate: 88% test trials, 37% empty control, 52% blocked control; see Appendix, Fig. A2). On average, the marmosets used the Position 0 platform more during test trials than during empty control trials, in which no rewards were placed on the see-saw ($\tilde{\beta} = 0.67$ [MAD = 0.29], 90% CI [0.20, 1.15], $p_+ = 0.99$, $\tilde{d} = 0.37$ [MAD = 0.16]), but there was a high degree of uncertainty in the difference between the test and blocked control sessions, in which rewards were placed on the see-saw yet access to these rewards was blocked ($\tilde{\beta} = 0.34$ [0.36], 90% CI [−0.24, 0.94], $p_+ = 0.84$,

$\tilde{d} = 0.19$ [0.20]). However, the stepping rates of 7/23 subjects provided clear evidence of differentiating between conditions with and without opportunities to provision group members (i.e. more use of the Position 0 platform in test trials than both controls; see Appendix, Fig. A3 and Table A3). Of these seven individuals, three nulliparous adult females, one subadult male and one subadult female exhibited high levels of prosociality (see Fig. 2). These 5/23 subjects engaged in appreciably higher rates of provisioning than the other two subjects who evidenced clear understanding of the task but had low test session provisioning rates overall ($\tilde{\beta} = 1.74$ [0.66], 90% CI [0.49, 2.85], $p_+ = 0.98$, $\tilde{d} = 0.96$ [0.36]; see Appendix, Fig. A4). We therefore considered these five subjects in particular to exhibit strong evidence of prosocial motivation.

Overall, these prosocial subjects successfully provisioned their groupmates between 24% and 70% of test trials. As expected, there was a resultant negative association between food received during testing and prosociality ($r_{\text{biserial}} = -0.35$). Given that these subjects differentiated between conditions in which their groupmates could and could not receive rewards, this further suggests that these individuals were not merely motivated by the expectation of reward at the apparatus after being provisioned themselves, but were instead motivated to provision their group members. Given that the loose string paradigm is intrinsically a dyadic task, we therefore classified all dyads containing at least one of these prosocial subjects as being prosocial dyads ($N = 14/37$ prosocial dyads, $N = 23/37$ nonprosocial dyads; 3−15 total dyads per group).

Loose String Paradigm

Successful cooperation in the loose string paradigm increased with subsequent experience across experimental trials at the group level ($\tilde{\beta} = 0.56$ [0.08], 90% CI [0.42, 0.70], $p_+ = 1.00$, $\tilde{d} = 0.31$ [0.05]; see Fig. 3), as well as for all dyadic pairs on average within those groups ($\tilde{\beta} = 0.38$ [0.05], 90% CI [0.30, 0.47], $p_+ = 1.00$, $\tilde{d} = 0.21$ [0.03]; see Fig. 4a, Appendix, Fig. A5). This indicates that the task required sustained learning among partners beyond the initial single string social tolerance task and further demonstrates that, on average, all dyads became better at coordinating their pulling in the task over time. Independently of these learning effects, and consistent with our central hypothesis, dyads containing prosocial subjects also contributed to an appreciably higher proportion of successful cooperation trials per group session ($\tilde{\beta} = 2.08$ [0.16], 90% CI [1.81, 2.34], $p_+ = 1.00$, $\tilde{d} = 1.15$ [0.09]; see Fig. 3). Indeed, prosocial dyads averaged 5.58 successful trials/session, but nonprosocial dyads averaged only 2.57 trials/session.

When comparing all possible dyads within each group, dyads containing a prosocial subject continued to exhibit a higher average probability of successful cooperation than those without a prosocial subject ($\tilde{\beta} = 0.62$ [0.28], 90% CI [0.14, 1.10], $p_+ = 0.98$, $\tilde{d} = 0.34$ [0.16]; see Fig. 4a, Appendix, Fig. A6). This suggests that the prosociality effect is not due to specific prosocial dyads within each group driving the overall difference in dyad types across group sessions. Indeed, this dyadic prosociality effect remained after controlling for partner choice (as indicated by the association between dyadic success in the previous session and success in the current session; $\tilde{\beta} = 0.57$ [0.31], 90% CI [0.06, 1.08], $p_+ = 0.97$, $\tilde{d} = 0.32$ [0.17]; see Fig. 4b, Appendix, Fig. A6), as well as for social tolerance ($\tilde{\beta} = 0.79$ [0.18], 90% CI [0.49, 1.10], $p_+ = 1.00$, $\tilde{d} = 0.44$ [0.10]; see Fig. 4d) and food motivation ($\tilde{\beta} = 0.28$ [0.12], 90% CI [0.07, 0.49], $p_+ = 0.98$, $\tilde{d} = 0.15$ [0.07]; see Fig. 4e). This demonstrates that while the pertinent factors of partner choice and individual social tolerance also influenced cooperation, prosociality nevertheless had an independent positive effect on average cooperative success across social partners. In addition, individuals of similar age also achieved higher rates of success ($\tilde{\beta} = 0.46$ [0.26],
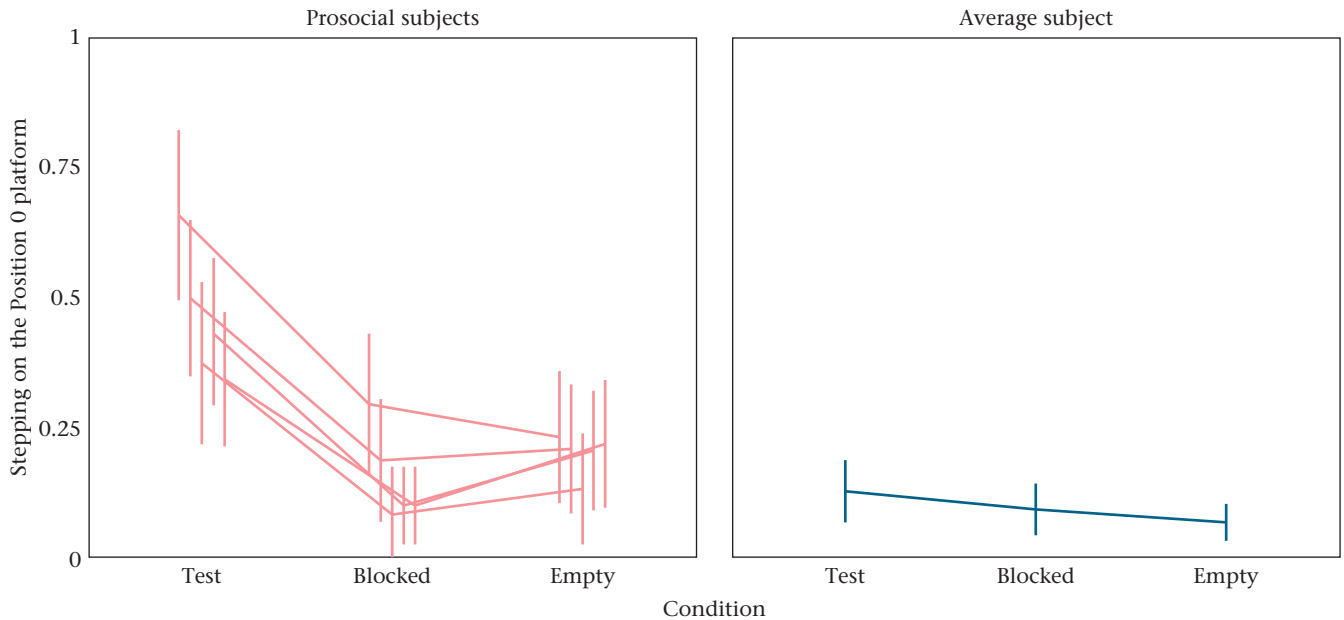
Figure 2. Group service task. The expected probability (Y axis) of stepping on the Position 0 provisioning platform across experimental conditions. Posterior medians ± SD are reported. Estimates are shown for the average subject as well as the five prosocial subjects who exhibited moderate to high rates of provisioning behaviour (see Fig. A3 for plots of all subjects and Table A3 for estimated posterior probabilities and between-condition random slope effects).
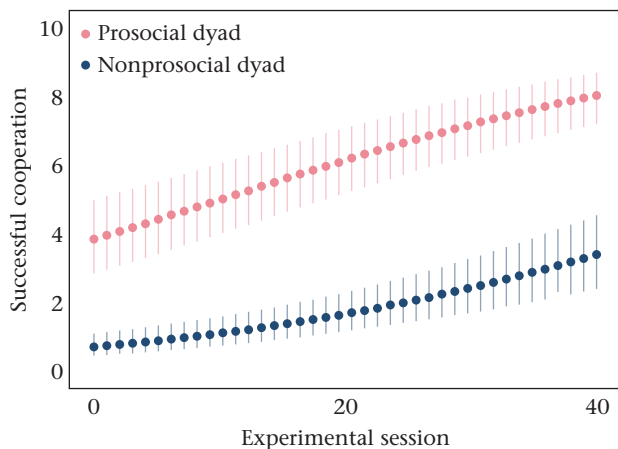

Figure 3. Cooperation across loose string experimental sessions. The total expected count of successful cooperation out of 10 trials/session (Y axis) for all dyads with ('Prosocial') and without ('Nonprosocial') a prosocial partner across group experimental sessions. Posterior medians are shown with 90% credible intervals. Estimates account for expected learning effects across sessions and are conditional on the average proportion of prosocial dyads within a group, marginalized over random effects. See Fig. A6 for raw counts across groups.

90% CI [0.03, 0.89], $p_+ = 0.96$, $\tilde{d} = 0.25$ [0.14]; see Fig. 4c). Our final model excluded parameters for personality, total dyadic age, sex combination and time of day, as model comparison suggested that these highly uncertain effects did not enhance model quality ($\Delta WAIC_{full - reduced\ model} = 0.00$ [SD = 1.30]).

DISCUSSION

We built on previous research by directly testing whether prosociality, social tolerance and partner choice facilitate individuals' capacity to achieve social coordination and problem solving in a mutually beneficial cooperative task. While suggestive, previous attempts to address the role of prosociality in animal cooperation have been limited by the employment of distinct behavioural measures and a reliance on interspecific comparisons, preventing an unambiguous assessment of the benefits of prosocial motivation for cooperation among conspecifics. We found that intraspecific variation in prosociality had a positive association with coordinated cooperation across social partners in general, independently of social tolerance and partner choice. These findings are consistent with the cooperative breeding hypothesis, which predicts that the prosocial motivation underlying costly helping behaviour also increases the probability of achieving coordination and problem solving across cooperative contexts more generally (Burkart & van Schaik, 2010), in addition to other mechanisms such as partner choice and social tolerance (Burkart et al., 2014). It is important to note that we did not examine performance in a task for which cooperative behaviour resulted in clear individual costs. We therefore cannot confidently infer that prosocial motivation as measured in the group service paradigm, where individuals experience negligible costs for food provisioning, would also facilitate coordination and problem solving in a behavioural experiment with a costlier altruistic payoff. Nevertheless, by using a mutually beneficial cooperation task with a large number of repeated trials, we were able to control for the role of individual benefits in motivating behaviour, showing that dyads with prosocial animals still performed better than nonprosocial dyads in securing food rewards for both social partners, independently of learning or other nonsocial motivational processes.

Despite the observed benefits of prosocial motivation for securing food rewards in the loose string task, we found that most of our subjects were not clearly prosocial (see Appendix, Fig. A3), with only 5/23 individuals evidencing a strong preference for provisioning their group members. Interestingly, none of these prosocial subjects were socially dominant within their groups, either being nulliparous adult females or subadult juveniles. While previous research has found higher prosociality among male compared to female helpers (Burkart, 2015), sex differences in marmoset provisioning behaviour have also been somewhat inconsistent across studies (Finkenwirth & Burkart, 2018). Such

apparent sex effects may reflect more fundamental variation in factors such as relationship quality and stability within groups (Finkenwirth & Burkart, 2017). These components of social integration are expected to predict further social tolerance from dominant individuals, who might otherwise expel subordinates from the group, and may thus function to regulate subordinates' motivation for costly helping behaviour. In this regard, our findings are consistent with the pay-to-stay hypothesis of alloparental care, which predicts that nonbreeders offset their costs to breeders through helping (Erb & Porter, 2017), although this interpretation remains speculative in light of our small sample size. While our study suggests generalized benefits of prosociality for coordination and cooperative problem solving, it therefore also supports previous work emphasizing that social roles and relationships are central determinants of marmosets' prosocial motivation and the expression of helping behaviour (Finkenwirth & Burkart, 2018). Whether the observed benefits of prosociality extend to such forms of cooperation among less familiar individuals thus remains an important question for future investigation, as does the potential role of reciprocity in facilitating food provisioning. In addition, it is likely that the uneven distribution of prosociality across our sample partially reflects the heterogeneous structure of the social groups, which included both prototypical family groups and groups containing multiple unrelated adults (see Appendix, Fig. A4). As is typical of research in captivity, our study thus provides desirable experimental control while also being limited by the reduced evolutionary and ecological relevance of the experimental context (Massen, Behrens, Martin, Stocker, & Brosnan, 2019).

Individual variation in social motivations such as prosociality is consistent with the more general observation that cooperative behaviour tends to be facultative and highly context dependent (Bourke, 2014; Gurven & Winking, 2008). Previous work has, for example, demonstrated the importance of factors such as relatedness (Green, Freckelton, & Hatchwell, 2016; Lukas & Clutton-Brock, 2012), local ecology (Shen, Emlen, Koenig, & Rubenstein, 2017) and social group dynamics (Smith et al., 2016) in the occurrence of cooperation across a variety of taxa. Whenever such cooperative behaviours are proximately influenced by prosocial motivation, we expect that individual levels of prosocial motivation may also be highly contingent. Similarly, while helping in the absence of overt signs of need suggests proactive prosociality (Burkart et al., 2009, 2014; Jaeggi et al., 2010), these motivations may nevertheless be conditional on the provisioner's and the receiver's state (Thornton & McAuliffe, 2015), including the constraints imposed by their social roles. The degree to which human cooperation relies on generalized, proactively prosocial motivations also remains unclear, at least within the context of economic games (Burton-Chellew & West, 2013). Caution should therefore be taken in drawing inferences about levels of species-typical prosociality and cooperation (Decety, Bartal, Uzefovsky, & Knafo-Noam, 2016; Thornton & McAuliffe, 2015), and appropriate consideration should be given to the influence of relevant social and ecological factors on individual motivation and the expression of cooperative behaviour (Massen et al., 2019).

It is also important to emphasize that the immediate benefits of a motivational trait on performance in experimental tasks need not translate into long-term fitness benefits observed in the wild
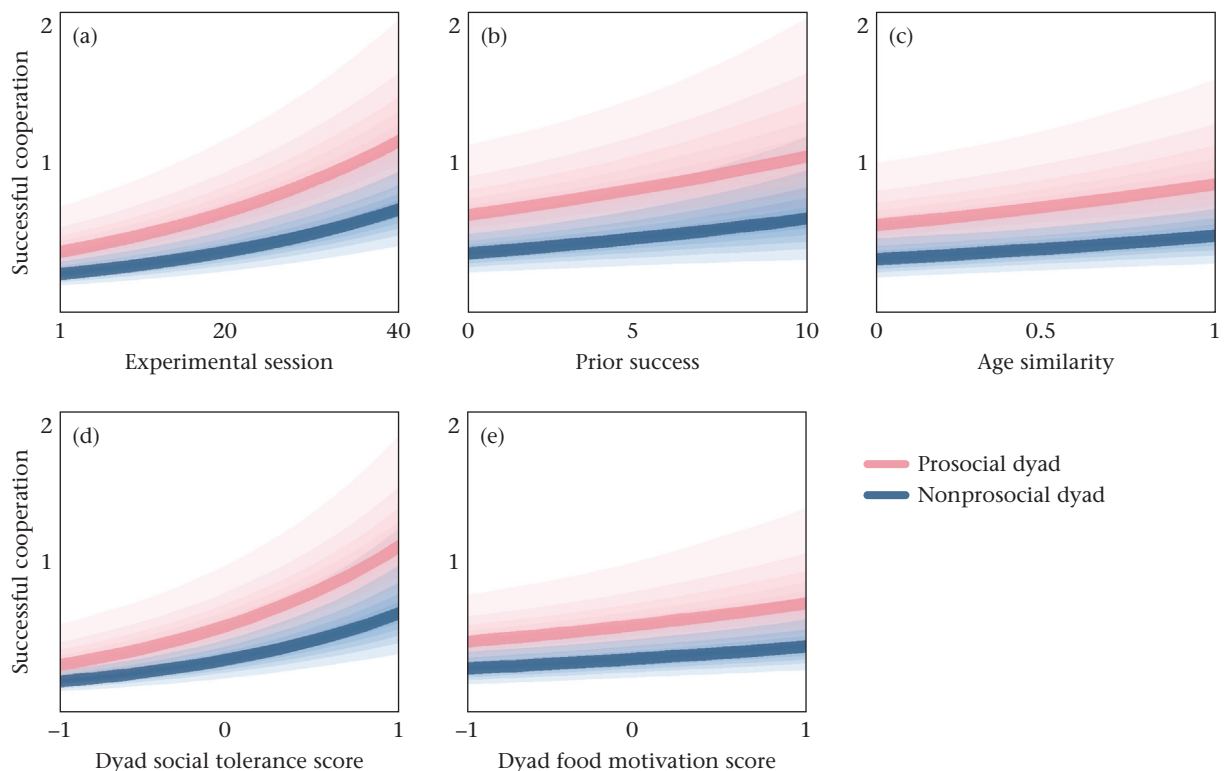

Figure 4. Effects on cooperation for a single dyad. The expected count of successful cooperation out of 10 trials/session (Y axis) for a typical dyad is shown as a function of (a) learning across experimental sessions, (b) success in the previous session (indicative of partner choice), (c) similarity in age, (d) total dyadic social tolerance score and (e) total dyadic food motivation score. Predictions are shown for dyads containing a prosocial subject (pink; N = 14) and those without one (blue; N = 23). Median model estimates are shown by the thick curve, with shaded posterior quantiles representing credible intervals from 10% to 90%. Estimates are conditional on average covariate values and are marginalized over random effects. Note that predicted counts of cooperation for a single dyad are relatively low because multiple dyads tended to use the paradigm per experimental session. Thus, while the absolute effect of a prosocial partner on success with any random group member was relatively small (Cohen's $d = 0.34$, independently of multiple correlated factors such as social tolerance; see Table A5), the aggregate effect across dyads at the group level was quite large ($d = 1.15$; see Fig. 3).

(Scott-Phillips, Dickins, & West, 2011; West, Griffin, & Gardner, 2007b). Our study demonstrates that prosociality predicted the capacity of marmosets to achieve mutually beneficial cooperation in a coordination task, suggesting that prosocial motivation may be a target of selection whenever the immediate benefits of such prosocially motivated behaviours translate into ultimate fitness benefits. The application of evolutionary ecological theory is needed, however, to generate sensible predictions about the conditions under which such long-term benefits would in fact be realized (West et al., 2007b). For example, the potential benefits of prosociality for coordination and problem solving may nevertheless be offset by greater long-term costs of defection or exploitation from nonprosocial partners, leading to selection against generalized prosociality. As discussed above, asymmetric payoffs between social partners — such as in the dominant—subordinate relations typical of cooperative breeders (Phillips, 2018) — may also select for the state-dependent expression of prosociality, contingent on social roles.

Given that prosociality is most likely to evolve in contexts that generate mutually beneficial or synergistic fitness benefits (Akçay et al., 2009), it would be particularly valuable to investigate the benefits of prosociality in the context of behaviours such as cooperative hunting, which is most likely to occur when hunters have a low probability of successfully capturing large prey alone (Packer & Ruttan, 1988). Alloparenting is also commonly observed in cooperatively hunting species, such as coyotes, *Canis latrans*, and grey wolves, *Canis lupus* (Smith, Lacey, & Hayes, 2017), which thus present ideal model systems for testing the conditions under which the hypothesized benefits of prosociality for social coordination and problem solving may also translate into lifetime fitness benefits. Wild marmosets also coordinate their behaviour in larger groups for cooperative territorial defence (Lazaro-Perea, 2001). Resource defence benefits are predicted to select for cooperative breeding in saturated environments (Lin, Chan, Rubenstein, Liu, & Shen, 2019; Shen et al., 2017) and may have enhanced prosociality among early human populations reliant on dense and predictable resources (Marean, 2016). Examining whether differential levels of group competition influence marmoset prosociality is therefore another exciting target for future investigation, as well as further understanding the influence of individual differences in prosociality (Carter, English, & Clutton-Brock, 2014; Schachner, Newton, Thompson, & Goodman-Wilson, 2018) on within-group cooperation and between-group competition more generally (Gavrilets, 2015; Majolo & Maréchal, 2017).

In addition to prosociality, we also observed positive effects of partner choice and individual social tolerance (see Fig. 4b and d) on performance in the loose string paradigm, supporting previous work identifying these factors as key mechanisms for primate cooperation (e.g. Hare et al., 2007; Jaeggi & Gurven, 2013; Melis, 2006; Melis et al., 2006; Molesti & Majolo, 2016; Sabbatini et al., 2012). The observed positive effect of age similarity on cooperation (Fig. 4c), irrespective of the total age of a dyad, may additionally reflect the fact that individuals of similar age tend to have stronger social bonds (e.g. de Waal & Luttrell, 1986; Silk, Alberts, & Altmann, 2006) and may be better able to coordinate their behaviour due to the benefits of phenotypic similarity for cooperation (Gabriel & Black, 2012; Laubu et al., 2016; Massen & Koski, 2014). The lack of support for personality similarity effects in our final model is surprising in this regard, as previous work has found positive associations between relationship quality and personality similarity among humans and nonhuman primates (Morton, Weiss, Buchanan-Smith, & Lee, 2015; Youyou, Stillwell, Schwartz, & Kosinski, 2017). The implications of this finding should be interpreted with caution, however, as previous evidence suggests high within-individual variability and group level consistency in

marmoset behaviour (Koski & Burkart, 2015). The importance of personality similarity for marmosets may therefore be more apparent during tasks requiring group cooperation. The small number of social groups in the present study and the dyadic nature of our paradigm, as well as the absence of other pertinent personality trait measures, prevent us from clearly addressing this question.

While we conducted the loose string experiment within group enclosures to enhance ecological validity (Burkart et al., 2014; Cronin, Jacobson, Bonnie, & Hopper, 2017; Massen, Ritter et al., 2015), this limited our capacity to differentiate partner choice per se (see Fig. 4b) from differential access to the experimental paradigm. In addition, although the loose string paradigm necessarily requires temporal coordination in pulling, we were unable to effectively distinguish behavioural indicators of coordination among partners, as various uncontrolled stimuli in the home enclosure also influenced subjects' attention, visual orientation and vocalizations during testing. Despite these limitations, given that rates of success steadily increased for the average dyad across experimental sessions (Figs. 3, 4a), it is clear that subjects improved the coordination of their pulling over time (Asakawa-Haas et al., 2016). In addition, our results suggest that competition for access to the paradigm was unlikely to be an important determinant of task success, as individual social tolerance was a clear factor promoting dyadic cooperation. It is also unlikely that partner choice for cooperation was confounded by task motivation, as we controlled for individual differences in food motivation observed during the group service task. Nevertheless, the use of forced dyad designs, as well as paradigms designed for more than two participants, would be a useful tool for fully eliminating such confounds in future studies, as well for better isolating potentially distinct individual, dyadic and group level factors relevant for marmosets' performance in these tasks.

Several alternative explanations for the observed effect of prosociality on cooperation are also possible, but unlikely. First, it is possible that our prosocial subjects performed better in the loose string task not because prosociality enhanced their cooperation with social partners, but rather because both measures reflect underlying differences in test comprehension among our subjects. We think this explanation is improbable, however, due to the extensive training provided to each social group prior to testing, as well as the large number of trials during testing. Indeed, all participating subjects evidenced sufficient comprehension to manipulate the loose string and group service apparatuses and procure food for themselves. Clear evidence of stimulus enhancement effects was only apparent for a few subjects in the group service paradigm (see Appendix, Fig. A3), with most adult subjects simply being motivated to receive food rather than provision their groupmates. General learning effects were also observed and accounted for across all dyads during the loose string paradigm, following training with the string-pulling mechanic during the social tolerance assessment. It is therefore unlikely that comprehension can account for the observed differences in dyadic cooperation between prosocial and nonprosocial subjects. However, given that the cooperative breeding hypothesis predicts that prosociality per se enhances performance in such tasks, it would be valuable for future research to further disentangle any benefits of generalized cognitive or problem-solving ability for performance from the specific effects of social motivation.

Alternatively, selfish rather than prosocial motivation may explain our results, as the variable reinforcement provided by motivational trials may have increased provisioning behaviour for some subjects during test sessions of the group service paradigm. If these individuals also had higher food motivation than their conspecifics during the loose string task, provisioning behaviour would

be expected to predict cooperative success. While plausible, this explanation is poorly supported by our data, as prosocial subjects in the group service paradigm tended to receive fewer rewards than the groupmates they provisioned. All subjects also received multiple food rewards at both the provisioning and receiving platforms prior to our testing sessions, making it unlikely that provisioning in particular can be explained by selfish motivation or reinforcement learning. Given that the prosociality effect on cooperation remains after controlling for general food motivation, selfish motivation is also an unlikely explanation for performance in the loose string paradigm.

In conclusion, our study demonstrates the value of examining individual differences, which often remain a large but relatively unexplained source of variation in experimental tasks (e.g. Barrett, McElreath, & Perry, 2017; Watson et al., 2018), to better understand the proximate determinants of cooperative behaviour. By utilizing repeated experimental measures and multilevel statistical models, we detected heterogeneous individual motivation and behaviour among our subjects, which were otherwise obscured by the aggregate pattern in our sample. In addition, by using experimental controls to more confidently identify subjects who both understood the task and exhibited high prosociality, we were able to nevertheless test our hypothesis in the absence of evidence for prosociality across all subjects. In so doing, we found direct support for the positive, independent effects of prosocial motivation, social tolerance and partner choice on social coordination and problem solving in a mutually beneficial task among group members, consistent with the broader claim that these traits are potential targets of selection for cooperative behaviour in primates. These benefits, which were independent of age, sex, personality, food motivation and learning, may therefore help to explain why prosociality, social tolerance and partner choice are associated with various cooperative behaviours across primates (Burkart et al., 2014; Hare et al., 2007; Jaeggi & Gurven, 2013; Kaigaishi, Nakamichi, & Yamada, 2019; Melis, 2006; Melis et al., 2006; Molesti & Majolo, 2016; Sabbatini et al., 2012), as well as their putative role in the evolution of uniquely human forms of social cognition and cooperation (Barclay, 2016; Burkart et al., 2009; Hrdy, 2009; Tomasello et al., 2005).

Data Availability

The data set and R code supporting this article are available as supplementary material.

Author Contributions

J.S.M., S.E.K., J.J.M.M. and T.B. collaboratively designed this experiment, J.S.M. conducted the experiment under the supervision of J.J.M.M. and S.E.K., J.S.M. conducted the statistical analyses under the supervision of A.V.J., J.S.M., and J.J.M.M. composed the manuscript, and all authors contributed to significant revision of the manuscript.

Acknowledgments

Supplementary Material

Supplementary material associated with this article is available, in the online version, at https://doi.org/10.1016/j.anbehav.2020.12.016.

References

Akçay, E., Van Cleve, J., Feldman, M. W., & Roughgarden, J. (2009). A theory for the evolution of other-regard integrating proximate and ultimate perspectives. *Proceedings of the National Academy of Sciences of the United States of America, 106,* 19061—19066.

Altschul, D. M., Wallace, E. K., Sonnweber, R., Tomonaga, M., & Weiss, A. (2017). Chimpanzee intellect: Personality, performance and motivation with touchscreen tasks. *Royal Society Open Science, 4,* Article 170169.

Amici, F., Mimó, M. C., von Borell, C., & Bueno-Guerra, N. (2017). Meerkats (*Suricata suricatta*) fail to prosocially donate food in an experimental set-up. *Animal Cognition, 20*(6), 1059—1066.

Asakawa-Haas, K., Schiestl, M., Bugnyar, T., & Massen, J. J. M. (2016). Partner choice in raven (*Corvus corax*) cooperation. *PLoS One, 11*(6), Article e0156962.

Barclay, P. (2016). Biological markets and the effects of partner choice on cooperation and friendship. *Current Opinion in Psychology, 7,* 33—38.

Barrett, B. J., McElreath, R. L., & Perry, S. E. (2017). Pay-off-biased social learning underlies the diffusion of novel extractive foraging traditions in a wild primate. *Proceedings of the Royal Society B: Biological Sciences, 284*(1856), Article 20170358.

Bourke, A. F. (2011). *Principles of social evolution.* Oxford, U.K.: Oxford University Press.

Bourke, A. F. (2014). Hamilton's rule and the causes of social evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences, 369*(1642), Article 20130362.

Brosnan, S. F., & de Waal, F. B. M. (2003). Monkeys reject unequal pay. *Nature, 425*(6955), 297—299.

Browne, W. J., Goldstein, H., & Rasbash, J. (2001). Multiple membership multiple classification (MMMC) models. *Statistical Modelling, 1*(2), 103—124.

Bshary, R., & Bergmüller, R. (2008). Distinguishing four fundamental approaches to the evolution of helping. *Journal of Evolutionary Biology, 21,* 405—420.

Bshary, R., Hohner, A., Ait-el-Djoudi, K., & Fricke, H. (2006). Interspecific communicative and coordinated hunting between groupers and giant moray eels in the red sea. *PLoS Biology, 4*(12), Article e431.

Burkart, J. M. (2015). Opposite effects of male and female helpers on social tolerance and proactive prosociality in callitrichid family groups. *Scientific Reports, 5,* Article 9622.

Burkart, J. M., Allon, O., Amici, F., Fichtel, C., Finkenwirth, C., Heschl, A., et al. (2014). The evolutionary origin of human hyper-cooperation. *Nature Communications, 5,* Article 4747.

Burkart, J. M., Fehr, E., Efferson, C., & van Schaik, C. P. (2007). Other-regarding preferences in a non-human primate: Common marmosets provision food altruistically. *Proceedings of the National Academy of Sciences of the United States of America, 104*(50), 19762—19766.

Burkart, J. M., Hrdy, S. B., & van Schaik, C. P. (2009). Cooperative breeding and human cognitive evolution. *Evolutionary Anthropology, 18*(5), 175—186.

Burkart, J. M., & van Schaik, C. P. (2010). Cognitive consequences of cooperative breeding in primates? *Animal Cognition, 13*(1), 1—19.

Burkart, J. M., & van Schaik, C. P. (2013). Group service in macaques (*Macaca fuscata*), capuchins (*Cebus apella*) and marmosets (*Callithrix jacchus*): A comparative approach to identifying proactive prosocial motivations. *Journal of Comparative Psychology, 127*(2), 212—225.

Burkart, J. M., & van Schaik, C. P. (2016). Revisiting the consequences of cooperative breeding. *Journal of Zoology, 299,* 77—83.

Bürkner, P. C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software, 80*(1), 1—28.

Burton-Chellew, M. N., & West, S. A. (2013). Prosocial preferences do not explain human cooperation in public-goods games. *Proceedings of the National Academy of Sciences of the United States of America, 110,* 216—221.

Callaghan, T., & Corbit, J. (2018). Early prosocial development across cultures. *Current Opinion in Psychology, 20,* 102—106.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., et al. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software, 76*(1), 1—32.

Carter, A. J., English, S., & Clutton-Brock, T. H. (2014). Cooperative personalities and social niche specialization in female meerkats. *Journal of Evolutionary Biology, 27*(5), 815—825.

Clark, C. W., & Dukas, R. (2003). The behavioral ecology of a cognitive constraint: Limited attention. *Behavioral Ecology, 14,* 151—156.

Clutton-Brock, T. (2009). Cooperation between non-kin in animal societies. *Nature, 462*(7269), 51–57.

Collins, L. M., Schafer, J. L., & Kam, C. M. (2001). A comparison of inclusive and restrictive strategies in modern missing data procedures. *Psychological Methods, 6*(3), 330–351.

Crane, A. L., & Ferrari, M. C. (2015). Minnows trust conspecifics more than themselves when faced with conflicting information about predation risk. *Animal Behaviour, 100*, 184–190.

Cronin, K. A. (2012). Prosocial behaviour in animals: The influence of social relationships, communication and rewards. *Animal Behaviour, 84*(5), 1085–1093.

Cronin, K. A., Jacobson, S. L., Bonnie, K. E., & Hopper, L. M. (2017). Studying primate cognition in a social setting to improve validity and welfare: A literature review highlighting successful approaches. *PeerJ, 5*, Article e3649.

Decety, J., Bartal, I. B. A., Uzefovsky, F., & Knafo-Noam, A. (2016). Empathy as a driver of prosocial behaviour: Highly conserved neurobehavioural mechanisms across species. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1686), Article 20150077.

Digby, L. J., & Barreto, C. E. (1993). Social organization in a wild population of *Callithrix jacchus. Folia Primatologica, 61*(3), 123–134.

Dingemanse, N. J., & Dochtermann, N. A. (2013). Quantifying individual variation in behaviour: Mixed-effect modelling approaches. *Journal of Animal Ecology, 82*(1), 39–54.

Dufour, V., Pele, M., Neumann, M., Thierry, B., & Call, J. (2009). Calculated reciprocity after all: Computation behind token transfers in orang-utans. *Biology Letters, 5*(2), 172–175.

Duque, J. F., Leichner, W., Ahmann, H., & Stevens, J. R. (2018). Mesotocin influences pinyon jay prosociality. *Biology Letters, 14*, Article 20180105.

Erb, W. M., & Porter, L. M. (2017). Mother's little helpers: What we know (and don't know) about cooperative infant care in callitrichines. *Evolutionary Anthropology, 26*(1), 25–37.

Fawcett, T. W., Hamblin, S., & Giraldeau, L.-A. (2013). Exposing the behavioral gambit: The evolution of learning and decision rules. *Behavioral Ecology, 24*(1), 2–11.

Fehr, E., & Fischbacher, U. (2003). The nature of human altruism. *Nature, 425*(6960), 785–791.

Fehr, E., & Gintis, H. (2007). Human motivation and social cooperation: Experimental and analytical foundations. *Annual Review of Sociology, 33*, 43–64.

Finkenwirth, C., & Burkart, J. M. (2017). Long-term-stability of relationship structure in family groups of common marmosets, and its link to proactive prosociality. *Physiology & Behavior, 173*, 79–86.

Finkenwirth, C., & Burkart, J. M. (2018). Why help? Relationship quality, not strategic grooming predicts infant-care in group-living marmosets. *Physiology & Behavior, 193*, 108–116.

Finlay, B. L., Hinz, F., & Darlington, R. B. (2011). Mapping behavioural evolution onto brain evolution: The strategic roles of conserved organization in individuals and species. *Philosophical Transactions of the Royal Society B: Biological Sciences, 366*(1574), 2111–2123.

Gabriel, P. O., & Black, J. M. (2012). Behavioural syndromes, partner compatibility and reproductive performance in Steller's jays. *Ethology, 118*(1), 76–86.

Gavrilets, S. (2015). Collective action problem in heterogeneous groups. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1683), Article 20150016.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, S. B., Vehtari, A., & Rubin, D. B. (2014a). *Bayesian data analysis* (3rd ed.). Boca Raton, FL: CRC Press.

Gelman, A., Hwang, J., & Vehtari, A. (2014b). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24*(6), 997–1016.

van Ginkel, J. R., Linting, M., Rippe, R. C., & van der Voort, A. (2019). Rebutting existing misconceptions about multiple imputation as a method for handling missing data. *Journal of Personality Assessment, 102*, 1–12.

González-Forero, M., & Gardner, A. (2018). Inference of ecological and social drivers of human brain-size evolution. *Nature, 557*(7706), 554–557.

Green, J. P., Freckleton, R. P., & Hatchwell, B. J. (2016). Variation in helper effort among cooperatively breeding bird species is consistent with Hamilton's rule. *Nature Communications, 7*, Article 12663.

Gurven, M., & Winking, J. (2008). Collective action in action: Prosocial behavior in and out of the laboratory. *American Anthropologist, 110*(2), 179–190.

Hare, B., Melis, A. P., Woods, V., Hastings, S., & Wrangham, R. (2007). Tolerance allows bonobos to outperform chimpanzees on a cooperative task. *Current Biology, 17*(7), 619–623.

Hare, B., Wobber, V., & Wrangham, R. (2012). The self-domestication hypothesis: Evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour, 83*, 573–585.

Harrison, X. A. (2014). Using observation-level random effects to model overdispersion in count data in ecology and evolution. *PeerJ, 2*, Article e616.

Heinrich, B. (2009). *The mind of the raven*. New York, NY: Harper Collins.

Heip, C. H. R., Herman, P. M. J., & Soetaert, K. (1998). Indices of diversity and evenness. *Oceanis, 24*(4), 61–87.

Hirata, S. (2003). Cooperation in chimpanzees. *Hattatsu, 95*, 103–111.

Horn, L., Scheer, C., Bugnyar, T., & Massen, J. J. M. (2016). Proactive prosociality in a cooperatively breeding corvid, the azure-winged magpie (*Cyanopica cyana*). *Biology Letters, 12*(10), Article 20160649.

Hrdy, S. B. (2009). *Mothers and others: The evolutionary origins of mutual understanding*. Cambridge, MA: Harvard University Press.

Isler, K., & van Schaik, C. P. (2012). How our ancestors broke through the gray ceiling: Comparative evidence for cooperative breeding in early homo. *Current Anthropology, 53*(Suppl. 6), S453–S465.

Jaeggi, A. V., Burkart, J. M., & van Schaik, C. P. (2010). On the psychology of cooperation in humans and other primates: Combining the natural history and experimental evidence of prosociality. *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1553), 2723–2735.

Jaeggi, A. V., & Gurven, M. (2013). Natural cooperators: Food sharing in humans and other primates. *Evolutionary Anthropology, 22*, 186–195.

Johnson, Z. V., & Young, L. J. (2017). Oxytocin and vasopressin neural networks: Implications for social behavioral diversity and translational neuroscience. *Neuroscience & Biobehavioral Reviews, 76*, 87–98.

Kaigaishi, Y., Nakamichi, M., & Yamada, K. (2019). High but not low tolerance populations of Japanese macaques solve a novel cooperative task. *Primates, 60*(5), 421–430.

Kedar, H., Rodriguez-Gironés, M. A., Yedvab, S., Winkler, D. W., & Lotem, A. (2000). Experimental evidence for offspring learning in parent–offspring communication. *Proceedings of the Royal Society B: Biological Sciences, 267*(1454), 1723–1727.

Kiers, E. T., Duhamel, M., Beesetty, Y., Mensah, J. A., Franken, O., Verbruggen, E., et al. (2011). Reciprocal rewards stabilize cooperation in the mycorrhizal symbiosis. *Science, 333*(6044), 880–882.

Koski, S. E., Buchanan-Smith, H. M., Ash, H., Burkart, J. M., Bugnyar, T., & Weiss, A. (2017). Common marmoset (*Callithrix jacchus*) personality. *Journal of Comparative Psychology, 131*, 326–336.

Koski, S. E., & Burkart, J. M. (2015). Common marmosets show social plasticity and group-level similarity in personality. *Scientific Reports, 5*, Article 8878.

Kurzban, R., Burton-Chellew, M. N., & West, S. A. (2015). The evolution of altruism in humans. *Annual Review of Psychology, 66*, 575–599.

Lambert, M. L., Massen, J. J. M., Seed, A. M., Bugnyar, T., & Slocombe, K. E. (2017). An 'unkindness' of ravens? Measuring prosocial preferences in *Corvus corax. Animal Behaviour, 123*, 383–393.

Laubu, C., Dechaume-Moncharmont, F. X., Motreuil, S., & Schweitzer, C. (2016). Mismatched partners that achieve postpairing behavioral similarity improve their reproductive success. *Science Advances, 2*(3), Article e1501013.

Lazaro-Perea, C. (2001). Intergroup interactions in wild common marmosets, *Callithrix jacchus*: Territorial defence and assessment of neighbours. *Animal Behaviour, 62*(1), 11–21.

Lin, Y. H., Chan, S. F., Rubenstein, D. R., Liu, M., & Shen, S. F. (2019). Resolving the paradox of environmental quality and sociality: The ecological causes and consequences of cooperative breeding in two lineages of birds. *American Naturalist, 194*(2), 207–216.

Lukas, D., & Clutton-Brock, T. (2012). Cooperative breeding and monogamy in mammalian societies. *Proceedings of the Royal Society B: Biological Sciences, 279*(1736), 2151–2156.

Majolo, B., & Maréchal, L. (2017). Between-group competition elicits within-group cooperation in children. *Scientific Reports, 7*, Article 43277.

Marean, C. W. (2016). The transition to foraging for dense and predictable resources and its impact on the evolution of modern humans. *Philosophical Transactions of the Royal Society B: Biological Sciences, 371*(1698), Article 20150239.

Martin, J. S., Massen, J. J. M., Šlipogor, V., Bugnyar, T., Jaeggi, A. V., & Koski, S. E. (2019). The EGA+ GNM framework: An integrative approach to modelling behavioural syndromes. *Methods in Ecology and Evolution, 10*, 245–257.

Martin, A., & Olson, K. R. (2015). Beyond good and evil: What motivations underlie children's prosocial behavior? *Perspectives on Psychological Science, 10*(2), 159–175.

Martin, J. S., Ringen, E. J., Duda, P., & Jaeggi, A. V. (2020). Harsh environments promote alloparental care across human societies. *Proceedings of the Royal Society B: Biological Sciences, 287*, Article 20200758.

Massen, J. J. M., Behrens, F., Martin, J. S., Stocker, M., & Brosnan, S. F. (2019). A comparative approach to affect and cooperation. *Neuroscience & Biobehavioral Reviews, 107*, 370–387.

Massen, J. J. M., Haley, S. M., & Bugnyar, T. (2020). Azure-winged magpies' decisions to share food are contingent on the presence or absence of food for the recipient. *Scientific Reports, 10*, Article 16147.

Massen, J. J. M., & Koski, S. E. (2014). Chimps of a feather sit together: Chimpanzee friendships are based on homophily in personality. *Evolution and Human Behavior, 35*, 1–8.

Massen, J. J. M., Lambert, M., Schiestl, M., & Bugnyar, T. (2015a). Subadult ravens generally don't transfer valuable tokens to conspecifics when there is nothing to gain for themselves. *Frontiers in Comparative Psychology, 6*, Article 885.

Massen, J. J. M., Ritter, C., & Bugnyar, T. (2015b). Tolerance and reward equity predict cooperation in ravens (*Corvus corax*). *Scientific Reports, 5*, Article 15021.

McCall, C., & Singer, T. (2012). The animal and human neuroendocrinology of social cognition, motivation and behavior. *Nature Neuroscience, 15*(5), Article 681.

McElreath, R. (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Boca Raton, FL: CRC Press.

McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46.

McNally, L., Brown, S. P., & Jackson, A. L. (2012). Cooperation and the evolution of intelligence. *Proceedings of the Royal Society B: Biological Sciences, 279*(1740), 3027–3034.

McNamara, J. M., & Houston, A. I. (2009). Integrating function and mechanism. *Trends in Ecology & Evolution, 24*(12), 670–675.

McShane, B. B., Gal, D., Gelman, A., Robert, C., & Tackett, J. L. (2019). Abandon statistical significance. *American Statistician, 73*, 235–245.

Melis, A. P. (2006). Chimpanzees recruit the best collaborators. *Science, 311*(5765), 1297–1300.

Melis, A. P., Engelmann, J. M., & Warneken, F. (2018). Correspondence: Chimpanzee helping is real, not a byproduct. *Nature Communications, 9*, Article 615.

Melis, A. P., Hare, B., & Tomasello, M. (2006). Engineering cooperation in chimpanzees: Tolerance constraints on cooperation. *Animal Behaviour, 72*(2), 275–286.

Melis, A. P., & Semmann, D. (2010). How is human cooperation different? *Philosophical Transactions of the Royal Society B: Biological Sciences, 365*(1553), 2663–2674.

Miss, F. M., & Burkart, J. M. (2018). Corepresentation during joint action in marmoset monkeys (*Callithrix jacchus*). *Psychological Science, 29*(6), 984–995.

Molesti, S., & Majolo, B. (2016). Cooperation in wild Barbary macaques: Factors affecting free partner choice. *Animal Cognition, 19*, 133–146.

Moll, H., & Tomasello, M. (2007). Cooperation and human cognition: The Vygotskian intelligence hypothesis. *Philosophical Transactions of the Royal Society B: Biological Sciences, 362*(1480), 639–648.

Morand-Ferron, J., Cole, E. F., & Quinn, J. L. (2016). Studying the evolutionary ecology of cognition in the wild: A review of practical and conceptual challenges. *Biology Reviews of the Cambridge Philosophical Society, 91*(2), 367–389.

Morton, F. B., Lee, P. C., & Buchanan-Smith, H. M. (2013). Taking personality selection bias seriously in animal cognition research: A case study in capuchin monkeys (*Sapajusapella*). *Animal Cognition, 16*, 677–684.

Morton, F. B., Weiss, A., Buchanan-Smith, H. M., & Lee, P. C. (2015). Capuchin monkeys with similar personalities have higher-quality relationships independent of age, sex, kinship and rank. *Animal Behaviour, 105*, 163–171.

Muller, M. N., & Mitani, J. C. (2005). Conflict and cooperation in wild chimpanzees. *Advances in the Study of Behavior, 35*, 275–331.

Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature, 441*(7092), 502–505.

Okasha, S. (2006). *Evolution and the levels of selection*. Oxford, U.K.: Oxford University Press.

Packer, C., & Ruttan, L. (1988). The evolution of cooperative hunting. *American Naturalist, 132*, 159–198.

Phillips, T. (2018). The concepts of asymmetric and symmetric power can help resolve the puzzle of altruistic and cooperative behaviour. *Biological Reviews, 93*, 457–468.

Plotnik, J. M., Lair, R., Suphachoksahakun, W., & de Waal, F. B. M. (2011). Elephants know when they need a helping trunk in a cooperative task. *Proceedings of the National Academy of Sciences of the United States of America, 108*(12), 5116–5121.

R Core Team. (2018). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Rubenstein, D. R., & Abbot, P. (Eds.). (2017). *Comparative social evolution*. Cambridge, U.K.: Cambridge University Press.

Sabbatini, G., De Bortoli Vizioli, A., Visalberghi, E., & Schino, G. (2012). Food transfers in capuchin monkeys: An experiment on partner choice. *Biology Letters, 8*(5), 757–759.

Samuni, L., Preis, A., Deschner, T., Crockford, C., & Wittig, R. M. (2018). Reward of labor coordination and hunting success in wild chimpanzees. *Communications Biology, 1*(1), 1–9.

Schachner, A. C., Newton, E. K., Thompson, R. A., & Goodman-Wilson, M. (2018). Becoming prosocial: The consistency of individual differences in early prosocial behavior. *Early Childhood Research Quarterly, 43*(2), 42–51.

Schuster, R., & Perelberg, A. (2004). Why cooperate? An economic perspective is not enough. *Behavioural Processes, 66*(3), 261–277.

Scott-Phillips, T. C., Dickins, T. E., & West, S. A. (2011). Evolutionary theory and the ultimate–proximate distinction in the human behavioral sciences. *Perspectives on Psychological Science, 6*, 38–47.

Shen, S. F., Emlen, S. T., Koenig, W. D., & Rubenstein, D. R. (2017). The ecology of cooperative breeding behaviour. *Ecology Letters, 20*(6), 708–720.

Silk, J. B. (2007). Empathy, sympathy, and prosocial preferences in primates. In R. I. M. Dunbar, & L. Barrett (Eds.), *The Oxford handbook of evolutionary psychology* (pp. 115–126). Oxford, U.K.: Oxford University Press.

Silk, J. B., Alberts, S. C., & Altmann, J. (2006). Social relationships among adult female baboons (*Papio cynocephalus*) II. Variation in the quality and stability of social bonds. *Behavioral Ecology and Sociobiology, 61*, 197–204.

Smith, D., Dyble, M., Thompson, J., Major, K., Page, A. E., Chaudhary, N., et al. (2016). Camp stability predicts patterns of hunter–gatherer cooperation. *Royal Society of Open Science, 3*(7), Article 160131.

Smith, J. E., Lacey, E. A., & Hayes, L. D. (2017). Sociality in non-primate mammals. In D. R. Rubenstein, & P. Abbot (Eds.), *Comparative social evolution* (pp. 284–319). Cambridge, U.K.: Cambridge University Press.

Stocker, M., Loretto, M. C., Sterck, E. H. M., Bugnyar, T., & Massen, J. J. M. (2020). Cooperation with closely bonded individuals reduces cortisol levels in long-tailed macaques. *Royal Society Open Science, 7*, Article 191056.

Tabachnick, B. G., & Fidell, L. S. (2014). *Using multivariate statistics* (6th ed.). Essex, U.K.: Pearson.

Tennie, C., Jensen, K., & Call, J. (2016). The nature of prosociality in chimpanzees. *Nature Communications, 7*, Article 13915.

Thornton, A., & McAuliffe, K. (2015). Cognitive consequences of cooperative breeding? A critical appraisal. *Journal of Zoology, 295*(1), 12–22.

Thornton, A., McAuliffe, K., Dall, S. R. X., Fernandez-Duque, E., Garber, P. A., & Young, A. J. (2016). Fundamental problems with the cooperative breeding hypothesis. A reply to Burkart & van Schaik. *Journal of Zoology, 299*(2), 84–88.

Tomasello, M., Carpenter, M., Call, J., Behne, T., & Moll, H. (2005). Understanding and sharing intentions: The origins of cultural cognition. *Behavioral and Brain Sciences, 28*, 675–691.

de Waal, F. B., & Luttrell, L. M. (1986). The similarity principle underlying social bonding among female rhesus monkeys. *Folia Primatologica, 46*, 215–234.

Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's statement on *p*-values: Context, process, and purpose. *American Statistician, 70*(2), 129–133.

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*, 3571–3594.

Watson, S. K., Vale, G. L., Hopper, L. M., Dean, L. G., Kendal, R. L., Price, E. E., et al. (2018). Chimpanzees demonstrate individual differences in social information use. *Animal Cognition, 21*, 639–650.

Wergård, E. M., Westlund, K., Spångberg, M., Fredlund, H., & Forkman, B. (2016). Training success in group-housed long-tailed macaques (*Macaca fascicularis*) is better explained by personality than by social rank. *Applied Animal Behaviour Science, 177*, 52–58.

West, S. A., Griffin, A. S., & Gardner, A. (2007a). Evolutionary explanations for cooperation. *Current Biology, 17*(16), R661–R672.

West, S. A., Griffin, A. S., & Gardner, A. (2007b). Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection. *Journal of Evolutionary Biology, 20*, 415–432.

Youyou, W., Stillwell, D., Schwartz, H. A., & Kosinski, M. (2017). Birds of a feather do flock together: Behavior-based personality-assessment method reveals personality similarity among couples and friends. *Psychological Science, 28*, 276–284.

Ziegler, T. E., & Crockford, C. (2017). Neuroendocrine control in social relationships in non-human primates: Field based evidence. *Hormones and Behavior, 91*, 107–121.

Appendix

Detailed Testing Procedures

Subjects and housing

At the time of observation, the laboratory housed seven social groups composed of two to six individuals including the juvenile and subadult offspring of dominant breeding pairs. Data were only collected for groups containing more than three individuals and for adults and juveniles who had been weened by the start of the experiments. This resulted in a total sample size of 23 individuals (mean ± SD age = 5.52 ± 4.86 years; 13 males, 10 females) in five social groups.

Personality scores

We used the regression method (Tabachnick & Fidell, 2014) to generate individual sociability and arousal scores from median posterior estimates of the Bayesian multiresponse model described in Martin et al.'s (2019) research on marmoset personality, which included 19/23 subjects from the present study. In particular, we calculated factor scores (**F**) for our subjects by

$$B = G^{-1}S$$

$$F = XB$$

where G^{-1} is the inverse among-individual random intercept correlation matrix, **S** is the matrix of factor loadings (i.e. the correlation between the latent factors and each observed behaviour) and **X** is the individual random intercept matrix for the observed behaviours. See Fig. A1 for a graphical overview of the sociability and arousal behavioural syndromes, including model parameters and specific behavioural indicators. These personality dimensions were derived using the Exploratory Graph Analysis + Generalized Network Modeling (EGA + GNM) framework (Martin et al., 2019),

which was applied to monthly counts and durations of individual behaviours, generated from repeated focal animal sampling across 6-weeks of spring (April–May) and summer (May–July) observational periods. These data were collected within 3–7 months of the present study. Given that individual differences were found to exhibit moderate to high temporal consistency across observational periods, we considered personality scores from this previous study to accurately reflect personality in the present study.

Experimental procedures

Group service paradigm. Our experimental procedure consisted of five phases based upon previous work by Horn and colleagues (Horn et al., 2016). Small pieces of blueberry and grape were provided as high-quality food rewards throughout the study period.

Phase 0: Habituation to the apparatus. During an initial habituation phase, the group service apparatus (see Fig. 1A) was installed in each group's enclosure for a 2-week habituation procedure prior to training. Rewards were placed ad libitum on the apparatus to enhance subjects' approach motivation.

Phase I: Initial training and habituation to the procedure. Training began with the see-saw mechanism locked in a downward position, so that food placed on the board automatically slid to the wire mesh. On alternating days, food was provided either in Position 0 (the provisioning platform) or in Position 1 (the receiving platform). For 5 × group size trials, the marmosets' attention was called ('Monkeys!') and one piece of food was placed on the board. The next trial started after a subject obtained the food or after a maximum of 2 min. If a subject took the food, we then placed the next piece of food on the board as soon as no subject was sitting on the platform where food was provided in this session. If no subject took the food, we called the marmosets' attention again, lifted the same piece of food and placed it back on the board. A session ended after all trials or when none of the subjects stepped on the platform for three consecutive trials. We proceeded to the next phase after each marmoset had taken at least 10 pieces of food in a minimum of five sessions.

Phase II: Food distribution assessment. As in Phase I, the see-saw mechanism was locked in a downward position so that food placed on the board automatically slid to the wire mesh. For 5 × group size trials per group, we called the marmosets' attention and placed one piece of food in Position 1. The next piece of food was placed after a subject retrieved the food or after a maximum of 2 min. The session ended after all trials were completed or when none of the subjects stepped on the apparatus for three consecutive trials. In the latter case, the session was aborted and redone the following day. Two sessions were conducted for each group.

To measure the evenness of food distribution for each group, we calculated Pielou's J' (Heip, Herman, & Soetaert, 1998) using the Shannon diversity index H', where J' is given by

$$J' = \frac{H'}{H'_{\max}}$$

$$H' = -\sum_{i=1}^{J} p_i \times \ln(p_i)$$

$$H'_{\max} = \ln(J')$$

For the proportion p_i of food retrieved by the ith individual ($i = 1, 2, ..., J$) over $2 \times (5 \times n)$ trials per group. H' quantifies the uncertainty of predicting the identity of the individual who retrieved the food on a randomly selected trial. $H'_{\max}$ quantifies the maximum possible state of uncertainty in a group, which is the case

where all individuals are equally likely to retrieve the food, i.e. $p_i = \frac{J}{2 \times (5 \times \text{group size})}$. $H'_{\max}$ therefore corrects H' for differences in group size.

See Table A2 for the J' of each group. Overall, there was a strong tendency towards equitable food sharing ($J' \approx 0.9$), consistent with previous findings for common marmosets using the original group service paradigm (Burkart & van Schaik, 2013). Notably, Aurora group had a highly unequitable distribution ($J' = 0.13$), with a single male receiving nearly all the food rewards. This idiosyncratic outcome likely reflects the unusually steep dominance hierarchy within this small social group.

Phase III: Further training. In this phase, the marmosets learned how to use the see-saw mechanism and move food towards the wire mesh by stepping on Position 0 (i.e. the provisioning platform). Food was always placed in Position 0. To facilitate learning, the see-saw mechanism was first loosened only partially and food was placed close to the wire mesh. When each marmoset obtained food from the apparatus at least once, the mechanism was then loosened further. The see-saw mechanism was gradually loosened in three steps and the food was subsequently placed further away from the wire mesh. In the final step, the see-saw mechanism was completely released and the food was placed at the other end of the board. Per trial, we called the marmosets' attention and placed one piece of food on the board. The next trial started after a subject obtained the food or after a maximum of 2 min had passed. If a subject took the food, we then placed the next piece of food on the board as soon as no subject was sitting on the platform. If no subject took the food, we then called the marmosets' attention again, lifted the same piece of food and placed it back on the board. A session ended after all 5 × group size trials or when none of the subjects stepped on the platform for three consecutive trials. Sessions continued until each marmoset took at least 10 pieces of food in a minimum of five sessions with the see-saw mechanism completely released. Two subjects did not frequently participate in the final training sessions (Smart, Ginevra; see Table A1) but attained at least 10 pieces of food across sessions. We therefore considered these individuals to have met sufficient training criteria for the subsequent testing phase.

Phase IV: Group service training and test. In this phase, the see-saw apparatus' mechanism was completely released. We conducted five test sessions and five empty control sessions on alternating days. During each trial of a test session, we called the marmosets' attention and placed a food reward on Position 1 (the receiving platform). The next trial started after a subject either retrieved the reward through provisioning by a subject stepping on Position 0 or after a maximum of 2 min. Additionally, we implemented motivation trials with food in Position 0 at the beginning of each session and after every fifth regular trial, which ensured that a lack of provisioning did not reflect a lack of food motivation. Each session therefore consisted of 5 × group size regular trials and 1 + group size motivation trials. For each trial, we recorded which subjects(s) stepped on the Position 0 platform (i.e. moved the see-saw mechanism), which subject(s) stepped on the Position 1 platform and which subject obtained the food reward. Stepping on the platform was only coded when a subject stopped moving across a platform. Following previous studies (Burkart & van Schaik, 2013; Horn et al., 2016), the first three sessions were considered as training sessions and the final two sessions were used as test sessions for data analysis. Due to experimental error, a single trial was skipped (24/25 completed) for Sprichtel group during their fourth group service session.

The empty control sessions were identical to the test sessions but no food was placed on the board. Instead, the experimenter approached the apparatus, called the marmosets' attention and pretended to leave a food reward in Position 1. Control sessions also

comprised motivation trials with food in Position 0 and contained the same total number of trials as test sessions.

Phase V: Blocked control. The see-saw apparatus' mechanism was completely released as during the test and empty control sessions, but access at Position 1 was blocked with a transparent plastic barrier, so that the food was still visible but could not be obtained. This allowed us to assess whether stepping on the provisioning platform (Position 0) reflected some stimulus enhancement effect of the food reward moving closer to the enclosure rather than prosocial motivation. During each trial, we recorded which subject(s) stepped on the Position 0 platform. There were five blocked control sessions and five empty control sessions on alternating days. The procedure was otherwise the same as in the test and control sessions of Phase IV, with each trial lasting 2 min. New trials did not begin until there were no subjects on the Position 0 platform. As in Phase IV, the first three sessions were used for training and only the final two sessions were used for comparison across conditions.

A research assistant blind to our hypotheses independently coded whether stepping occurred in each trial for 20% of sessions across experimental conditions. Using a two-way, mixed effects, absolute agreement, single rater intraclass correlation coefficient (McGraw & Wong, 1996), our measurements were found to be highly reliable (ICC (3,1) = 0.99).

Loose string paradigm.

Phase 0: Habituation to the apparatus. Our experimental procedure began with a brief habituation phase in which the loose string apparatus was placed in front of the subjects' enclosures for 15−30 min periods intermittently for 2 weeks. Food rewards were placed on the moving platform ad libitum and slowly moved towards the subjects to habituate them to the mechanism. Small pieces of blueberry, grape and banana were provided as high-quality food rewards throughout the study.

Phase I: Social tolerance task and training. In addition to the initial habituation phase, each group received a further 180 trial period to familiarize subjects with the string-pulling task, enhance motivation towards the apparatus and assess individual social tolerance towards group members in the presence of food rewards. This social tolerance assessment was based on previous work using the loose string paradigm in ravens (Massen, Ritter et al., 2015). This phase consisted of 18 sessions of 10 trials each in which two untethered strings tied to food rewards were placed on the surface of the moving platform apparatus adjacent to one another. Subjects who tolerated one another at the platform could independently retrieve the food reward tied to the distal end of the string by pulling their respective string into the enclosure. For each trial, we called the attention of the marmosets ('Monkeys!') and simultaneously placed two independent strings on the apparatus, with one end of each string within arm's length of two standing platforms in each enclosure. A trial ended after both strings had been pulled in or after a maximum of 2 min, and we coded which monkeys retrieved the strings adjacent to one another.

Following previous work on loose string cooperation (Massen, Ritter et al., 2015), we derived a measure of social tolerance for each individual by averaging their mean single string retrieval rate in the presence of each group member across experimental sessions. We found that these retrieval rates were highly consistent among individuals across the 18 training sessions (ICC (3,1) = 0.60), such that the total dyadic social tolerance score also effectively controlled for the influence of prior single string training on each dyad's performance during the testing phase.

Phase II: Testing. A total of 40 testing sessions were subsequently conducted for each group, with 10 trials per session. The marmosets' attention was called to the apparatus at the beginning of each session as two equally sized food rewards were placed on the left- and right-hand side of the apparatus, within view of the respective test platforms. A single string was then tethered through the apparatus, raised up and placed within arm's length of the test platforms as soon as two subjects were present on the platforms (see Fig. 1b). A trial was coded as successful if the two subjects were able to pull the moving apparatus towards their enclosure and retrieve the food rewards; conversely, a session was unsuccessful if the string was untethered and the subjects were unable to retrieve the reward, which could occur either through uncoordinated pulling or through a single subject pulling the entire string through to his or her platform. Each trial ended after either a successful or unsuccessful retrieval of the food reward or a maximum of 1 min. Sessions were aborted if three trials ended without any subjects attempting to use the apparatus. A total of one to four sessions were conducted per day for training and test sessions, contingent on the observed food motivation of the group.

A research assistant blind to our hypotheses independently coded whether successful cooperation occurred for 10% of experimental sessions per group. Our measurements were found to be highly reliable (ICC (3,1) = 0.99).

Statistical models and procedures

We estimated Bayesian generalized linear mixed-effects models (GLMMs) for all analyses using the R package 'brms' (Bürkner, 2017), which interfaces with the Stan statistical programming language (Carpenter et al., 2017). As noted in the main text, we employed a fully Bayesian approach to statistical estimation and inference (McElreath, 2016). Therefore, rather than relying upon null hypothesis tests and arbitrary designations of statistical significance, we used multiple sources of information to summarize and draw inferences from our posterior model estimates (McElreath, 2016; McShane et al., 2019; Wasserstein & Lazar, 2016). The R Code and the data set provided in the supplementary material can be used to replicate all analyses described below.

Group service paradigm. We first compared rates of stepping on the Position 0 provisioning platform across conditions during the test sessions (session 4 and 5) to assess whether the subjects understood the task and therefore stepped to provide food rather than because of a stimulus enhancement effect. To do this, we estimated a binomial GLMM predicting stepping on the Position 0 provisioning platform across subjects. Fixed effects included experimental condition (test, empty, control) and time of day to control for unbalanced sampling across morning and afternoon periods. Test was set as the reference category for the experimental condition. Random effects included subject-specific intercepts and slopes across experimental conditions, social group intercepts, and observation level intercepts to account for overdispersion (Harrison, 2014). We therefore estimated the following model conditional on our observed data for observation i of an individual's stepping rate during an experimental session.

Model 1.

$$\text{Step on Position } 0_i \sim \text{Binomial}(n_i, p_i)$$

$$\text{logit}(p_i) = A_i + B_i$$

$$A_i = \alpha + \alpha_{\text{individual}[i]} + \alpha_{\text{group}[i]} + \alpha_{\text{observation}[i]}$$

$$B_i = \beta_{\text{blocked}} + \beta_{\text{blocked,individual}[i]} + \beta_{\text{empty}} + \beta_{\text{empty,individual}[i]}$$
$$+ \beta_{\text{time}}$$

$$\begin{bmatrix} \alpha_{\text{ind}} \\ \beta_{\text{blocked}} \\ \beta_{\text{empty}} \end{bmatrix} \sim \text{MVNormal} \left(\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \mathbf{SRS} \right)$$

$$\alpha, \beta \sim \text{Normal}(0,2)$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0,2)$$

The five subjects with appreciably higher prosocial motivation were categorized as 'prosocial' subjects throughout subsequent analyses.

Loose string paradigm. As previously noted, we prioritized the enhanced ecological validity of testing in the home enclosure for

$$\mathbf{S} = \begin{pmatrix} \sigma_{\text{ind}} & 0 & 0 \\ 0 & \sigma_{\text{blocked}} & 0 \\ 0 & 0 & \sigma_{\text{empty}} \end{pmatrix}, \mathbf{R} = \begin{pmatrix} 1 & \text{cor}(\alpha_{\text{ind}}, \beta_{\text{blocked}}) & \text{cor}(\alpha_{\text{ind}}, \beta_{\text{empty}}) \\ \text{cor}(\alpha_{\text{ind}}, \beta_{\text{blocked}}) & 1 & \text{cor}(\beta_{\text{blocked}}, \beta_{\text{empty}}) \\ \text{cor}(\alpha_{\text{ind}}, \beta_{\text{empty}}) & \text{cor}(\beta_{\text{blocked}}, \beta_{\text{empty}}) & 1 \end{pmatrix}$$

$$\alpha_{\text{individual}}, \alpha_{\text{group}}, \alpha_{\text{observation}} \sim \text{Normal}(0, \sigma)$$

$$\alpha, \beta \sim \text{Normal}(0,2)$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0,2)$$

$$\mathbf{R} \sim \text{LKJcorr}(2)$$

where n is the group-specific trial number per session, p is the probability of stepping on Position 0, α are intercepts, β are fixed effects coefficients, $\mathbf{S}$ is the matrix of random effect standard deviations and $\mathbf{R}$ is the correlation matrix. This notation is used throughout the remainder of the Appendix. See Appendix, Fig. A3 and Table 3 below for the subject-specific stepping rates estimated from this model. Subjects estimated with high certainty to have stepped more frequently on the provisioning platform during the test condition compared to the empty and blocked control conditions were considered to have understood the task.

Although seven subjects exhibited an understanding of the group service paradigm (see Fig. A3 and Table A3), five of these subjects appeared to exhibit appreciably higher prosocial motivation (see Fig. 2 and Fig. A4). To formally assess this difference in motivation, we further modelled provisioning behaviour for only these seven subjects, comparing provisioning rates during test sessions between the two older subjects with low Position 0 stepping rates and the five subjects with moderate to high rates (coded as a binary variable: high (1) or low (0) rate).

Model 2.

$$\text{Provisioning}_i \sim \text{Binomial } (n_i, p_i)$$

$$\text{logit}(p_i) = A_i + B_i$$

$$A_i = \alpha + \alpha_{\text{observation}[i]}$$

$$B_i = \beta_{\text{high stepping rate}}$$

$$\alpha_{\text{observation}} \sim \text{Normal}(0, \sigma)$$

each social group, rather than conducting a standardized number of trials for all possible dyads. Dyad- and individual-specific outcomes were therefore contingent upon factors such as partner choice and access to the apparatus. In testing our primary hypothesis for the relationship between prosociality and cooperation, we began by analysing responses across groups before investigating dyad-specific outcomes. In particular, we predicted that a greater proportion of successful trials would involve a prosocial individual ('prosocial' trials (1)) compared to successful trials without a prosocial individual ('nonprosocial' trials (0)), irrespective of subject or dyad identity. Only one group was found to have two prosocial individuals (Sprichtel; see Table A1), and we therefore lumped successful trials containing one or two prosocial subjects together as 'prosocial trials' for comparison across groups. One group (Cleli) lacked any prosocial subjects and all of their successful trials were therefore coded as nonprosocial trials. Given that groups varied in the proportion of potential dyads containing a proactively prosocial individual ('proportion prosocial'), we controlled for the proportion of possible prosocial dyads in our analysis. Note that our low statistical power (5 social groups) prevented any substantive interpretation of group level covariates, including the estimated effect size for the proportion of prosocial dyads (raw success rates across groups are shown in Fig. A5). In addition, we included fixed effects for session number to account for learning effects, as well as time of day to control for any effect of unbalanced sampling across morning and afternoon periods.

Model 3.

$$\text{Successful cooperation}_i \sim \text{Binomial } (n_i, p_i)$$

$$\text{logit}(p_i) = A_i + B_i$$

$$A_i = \alpha + \alpha_{\text{group}[i]} + \alpha_{\text{date}[i]} + \alpha_{\text{observation}[i]}$$

$$B_i = \beta_{\text{prosocial trial}} + \beta_{\text{proportion prosocial}} + \beta_{\text{session}} + \beta_{\text{time}}$$

$$\alpha_{\text{group}}, \alpha_{\text{date}}, \alpha_{\text{observation}} \sim \text{Normal}(0, \sigma)$$

$$\alpha, \beta \sim \text{Normal}(0,2)$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0,2)$$

We subsequently analysed dyad level outcomes to account for nested responses within individuals and dyads, as well as to explore the potential influence of age, sex, personality, partner choice, social tolerance and food motivation on observed responses. We estimated the effect of prosociality on dyadic cooperation by comparing the rates of successful cooperation among dyads with (1) and without (0) one of the prosocial subjects. As described in the main text, we also included fixed effects for sex combination, partner preference and the effects of both dyadic similarity and total trait value for age (years), arousal and sociability. In addition, we included the total effect of social tolerance and food motivation to remove any bias in our prosociality measure, as these factors are expected to increase interest and success in experimental tasks and may therefore cause correlated performance across our tasks. Total values were simply the sum of each partner's scores, while similarity was calculated on a 0–1 scale as $1/(1+|$value subject 1 − value subject 2$|)$. To capture the influence of partner choice on cooperation, we used dyadic success in the previous experimental session as a predictor of current success (0/10 for the first experimental session, and $x/10$ for x successful trials in the previous session).

Sociability and arousal scores were missing for four subjects who were not observed in our previous personality study (Martin et al., 2019). Removal of observations including these subjects would result in a significant and undesirable loss of statistical power. Mean imputation is a common solution to this problem, but this approach can potentially bias model estimates due to the assumption that responses are missing completely at random (Collins, Schafer, & Kam, 2001; McElreath, 2016). To assess the possibility of nonrandom missingness in our data, we used Bayesian imputation to examine missingness as a function of age, sex, and prosociality. No clear effects were observed, suggesting that mean imputation provided a simpler and appropriate solution. Scores of 0 were therefore imputed for the sociability and arousal z scores of these four subjects.

During any particular trial, subjects engaged in the loose string task may have used the left or right test platform, but we were interested in modelling the random effect of subject identity on dyadic outcomes irrespective of their position. Appropriately accounting for such arbitrary dyadic structuring, and thus accurately estimating uncertainty in subject random effects, requires the use of so-called multiple membership modeling (Collins et al., 2001). In addition to our multimembership random subject effect, we also included random effects for dyad, social group, observation and day of observation to account for unbalanced sampling across days. Given a high proportion of observed zeroes, we also estimated the probability z_i of not cooperating during a test session to account for zero-inflation in the dyadic responses. We did not include additional effects to predict zero-inflation due to insufficient statistical power, and we therefore made the simplifying assumption that z_i was constant across trials. Our final fully parameterized model therefore estimated the following effects conditional on our data for observation i of dyadic cooperation during each loose string session.

Model 4.0.

$$\text{Successful dyadic cooperation}_i \sim \text{ZIbinomial}(n_i, z_i, p_i)$$

$$\text{logit}(p_i) = A_i + B_i$$

$$A_i = \alpha + \alpha_{\text{individual}_1[i]} + \alpha_{\text{individual}_2[i]} + \alpha_{\text{dyad}[i]} + \alpha_{\text{group}[i]} + \alpha_{\text{date}[i]} + \alpha_{\text{observation}[i]}$$

$$\begin{aligned} B_i = &\beta_{\text{prosocial dyad}} + \beta_{\text{age total}} + \beta_{\text{age similarity}} + \beta_{\text{Arousal total}} \\ &+ \beta_{\text{Arousal similarity}} + \beta_{\text{Sociability total}} + \beta_{\text{Sociability similarity}} \\ &+ \beta_{\text{Social tolerance total}} + \beta_{\text{Food motivation total}} \\ &+ \beta_{\text{prior success}} + \beta_{\text{session}} + \beta_{\text{time}} \end{aligned}$$

$$\alpha_{\text{individual}}, \alpha_{\text{group}}, \alpha_{\text{date}}, \alpha_{\text{observation}} \sim \text{Normal}(0, \sigma)$$

$$z_i \sim \text{Beta}(1, 1)$$

$$\alpha, \beta \sim \text{Normal}(0, 2)$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0, 2)$$

To avoid overfitting our model given the moderate power of our sample, we used the fully Bayesian Watanabe–Akaike information criterion (WAIC) (Watanabe, 2010) to compare this full model to a reduced model excluding any terms that exhibited highly uncertain effects in Model 4.0. This reduced model had the following structure.

Model 4.1.

$$\text{Successful dyadic cooperation}_i \sim \text{ZIbinomial}(n_i, z_i, p_i)$$

$$\text{logit}(p_i) = A_i + B_i$$

$$A_i = \alpha + \alpha_{\text{individual}_1[i]} + \alpha_{\text{individual}_2[i]} + \alpha_{\text{dyad}[i]} + \alpha_{\text{group}[i]} + \alpha_{\text{date}[i]} + \alpha_{\text{observation}[i]}$$

$$\begin{aligned} B_i = &\beta_{\text{prosocial dyad}} + \beta_{\text{age similarity}} + \beta_{\text{Social tolerance total}} \\ &+ \beta_{\text{Food motivation total}} + \beta_{\text{prior success}} + \beta_{\text{session}} \end{aligned}$$

$$\alpha_{\text{individual}}, \alpha_{\text{group}}, \alpha_{\text{date}}, \alpha_{\text{observation}} \sim \text{Normal}(0, \sigma)$$

$$z_i \sim \text{Beta}(1, 1)$$

$$\alpha, \beta \sim \text{Normal}(0, 2)$$

$$\sigma \sim \text{Half} - \text{Cauchy}(0, 2)$$

As reported in the main text, our model comparison provided support for the reduced Model 4.1 excluding uncertain effects (see Table A4), suggesting that this more parsimonious structure better represents our data. Results reported in the main text are therefore estimated based upon Model 4.1.

Two juvenile subjects in Cleli group (see Table A1) could not be reliably sexed during loose string testing, as our laboratory does not use invasive tagging procedures and these twins could not yet be distinguished by their facial features or genitalia. To account for uncertainty in the assignment of subject identity for these individuals, we used an additional Bayesian multiple imputation procedure. Relative to the covariates in our model, these subjects

only differed in their sex. Given that the sexes were not observed to differ in their frequency of cooperation (mean ± SD: males: 3.16 ± 2.90 trials; females: 3.60 ± 2.89 trials), we therefore assumed an equal probability of observing either the male or female juvenile across trials ($p = 0.50$) and produced five data sets with different randomly generated trial level sequences of identity assignment. We subsequently fitted Model 4.1 to each of these data sets and pooled their posteriors together to account for the influence of uncertainty in trial level assignment on our model estimates (Gelman et al., 2014a). Values reported in the main text are based on this multiple imputation model. Note that very little variance in parameter estimates was observed across the imputed data sets, suggesting that our results were highly robust to this procedure.

Table A1
Subject demographics

Social group	Subject (abbrev.)	Sex	Years of age
Aurora			
	Aurora (AUR)	F	4
	Jack (JCK)	M	6
	Mink (MNK)	M	6
Cleli			
	Blinky Bill (BLK)	M	1
	Clever (CLV)[a]	M	5
	Juvenile F (JUVF)[b]	F	0
	Juvenile M (JUVM)[b]	M	0
	Wall-E (WLE)	M	1
	Veli (VLI)[a]	F	7
Sprichtel			
	Kobold (KBL)[a]	M	7
	Nala (NLA)	F	1
	Simba (SMB)	M	1
	Smart (SMRT)[a]	M	5
	Sparrow (SPR)[a]	F	6
V			
	Ernesto (ERN)[a, c]	M	12
	Genevra (GNV)[a, c]	F	12
	Mathilda (MAT)	F	2
	Melvin (MEL)	M	2
	Vincent (VIN)	M	4
Zaphod			
	Luna (LNA)	F	3
	Nemo (NMO)	F	7
	Oli (OLI)[a]	F	6
	Zaphod (ZPH)[a]	M	7

Years of age are rounded to the nearest whole number and reported for the start of the study period. Note that the subject abbreviations are used in Supplementary data set.
[a] Current or previous breeders within a social group.
[b] These twin juveniles had been weaned by the time of testing but were not yet named by the laboratory.
[c] These subjects did not participate during any loose string sessions.

Table A2
Food distribution assessment (group service paradigm, Phase II)

Social group	Subject	p_i	J'
Aurora			0.13
	Aurora	0.03	
	Jack	0.97	
	Mink	0.00	
Cleli			0.96
	Blinky Bill	0.08	
	Clever	0.15	
	Juvenile F	0.12	
	Juvenile M	0.28	
	Wall-E	0.2	
	Veli	0.17	

Table A2 (continued)

Social group	Subject	p_i	J'
Sprichtel			0.89
	Kobold	0.04	
	Nala	0.42	
	Simba	0.12	
	Smart	0.24	
	Sparrow	0.18	
V			0.90
	Ernesto	0.16	
	Genevra	0.30	
	Mathilda	0.04	
	Melvin	0.16	
	Vincent	0.34	
Zaphod			0.91
	Luna	0.23	
	Nemo	0.23	
	Oli	0.10	
	Zaphod	0.45	
Average			0.76

Higher values of Pielou's J' indicate a more equitable distribution of food across subjects. The calculation of J' is described in the Appendix (Group service paradigm, Phase II) above. p_i refers to individual-specific probabilities of attaining the food reward across two experimental sessions. See Table A1 for further details on the subjects and social groups.

Table A3
Group service random slope estimates

Social group	Subject	$\bar{p}_{Test}$ (SD)	$p_{Blocked\_}$	$p_{Empty\_}$
Aurora				
	Aurora[b]	0.66 (0.16)	0.99	1.00
	Jack	0.09 (0.07)	0.03	0.69
	Mink	0.15 (0.10)	0.19	0.52
Cleli				
	Blinky Bill	0.09 (0.07)	0.10	0.74
	Clever	0.03 (0.04)	0.73	0.70
	Juvenile F	0.56 (0.14)	0.78	1.00
	Juvenile M	0.28 (0.13)	0.02	0.73
	Wall-E	0.07 (0.06)	0.17	0.77
	Veli	0.02 (0.03)	0.68	0.29
Sprichtel				
	Kobold	0.05 (0.05)	0.03	0.51
	Nala[b]	0.43 (0.14)	1.00	0.99
	Simba[b]	0.35 (0.13)	0.99	0.96
	Smart	0.04 (0.04)	0.22	0.33
	Sparrow	0.04 (0.04)	0.75	0.56
V				
	Ernesto[a]	0.08 (0.09)	0.97	0.96
	Genevra	0.01 (0.03)	0.67	0.50
	Mathilda[b]	0.38 (0.16)	1.00	1.00
	Melvin	0.24 (0.14)	0.61	0.88
	Vincent	0.15 (0.12)	0.77	0.95
Zaphod				
	Luna[b]	0.49 (0.15)	0.99	1.00
	Nemo	0.07 (0.07)	0.09	0.64
	Oli[a]	0.17 (0.11)	0.99	0.93
	Zaphod	0.03 (0.04)	0.59	0.69

Values are based upon posterior random slopes estimated from Model 1 described above. $\bar{p}_{Test}$ is the estimated median probability of stepping on the Position 0 platform during test condition experimental sessions. $p_{Blocked\_}$ is the posterior probability of a subject exhibiting a lower rate of stepping on Position 0 during the blocked condition than in the test condition, while $p_{Empty\_}$ is the posterior probability of a subject exhibiting a lower stepping rate in the empty condition. Probability decimals are rounded to the nearest hundredth. See Table A1 for further details on the subjects and social groups.
[a] Subjects exhibiting evidence of understanding the task.
[b] Subjects exhibiting an understanding of the task and moderate to high prosocial motivation.

Table A4
Full Model 4.0 posterior fixed effect estimates

Fixed effect	$\bar{\beta}$ [MAD]	90% CI	p_+
Experimental session	**0.39 [0.06]**	**[0.30, 0.48]**	**1.00**
Time of day	0.05 [0.11]	[−0.14, 0.24]	0.68
Prior success	**0.54 [0.31]**	**[0.04, 1.06]**	**0.96**
Prosociality	**0.69 [0.39]**	**[0.04, 1.35]**	**0.96**
Total social tolerance	**0.78 [0.22]**	**[0.40, 1.17]**	**1.00**
Total food motivation	**0.27 [0.15]**	**[0.02, 0.52]**	**0.96**
Total age	0.06 [0.21]	[−0.29, 0.42]	0.61
Age similarity	**0.48 [0.32]**	**[−0.05, 1.02]**	**0.94**
Total sociability	0.12 [0.15]	[−0.13, 0.38]	0.79
Sociability similarity	0.17 [0.58]	[−0.81, 1.13]	0.62
Total arousal	0.10 [0.15]	[−0.16, 0.35]	0.75
Arousal similarity	0.24 [0.58]	[−0.72, 1.23]	0.66

Bolded values represent parameters selected for the reduced Model 4.1, which was selected to decrease the risk of overfitting and enhance model generalization. Parameters with a posterior probability of p_+ or $p_- < 0.90$ were removed due to high statistical uncertainty for the direction of the effect.

Table A5
Correlations between prosociality and other individual level traits

	Prosociality	Social tolerance	Food motivation	Sociability	Arousal	Sex	Age
Prosociality	1						
Social tolerance	0.45	1					
Food motivation	−0.35	0.31	1				
Sociability	−0.17	−0.27	0.04	1			
Arousal	−0.02	0.35	0.36	−0.03	1		
Sex	0.64	−0.08	0.54	0.02	0.25	1	
Age	−0.66	−0.02	0.10	0.08	−0.26	−0.10	1

Correlation coefficients are reported for individual covariates used in our analysis of dyadic cooperation. Biserial and tetrachoric correlations are reported for prosociality (0 = not prosocial, 1 = prosocial) and sex (0 = male, 1 = female) because these are binary variables.
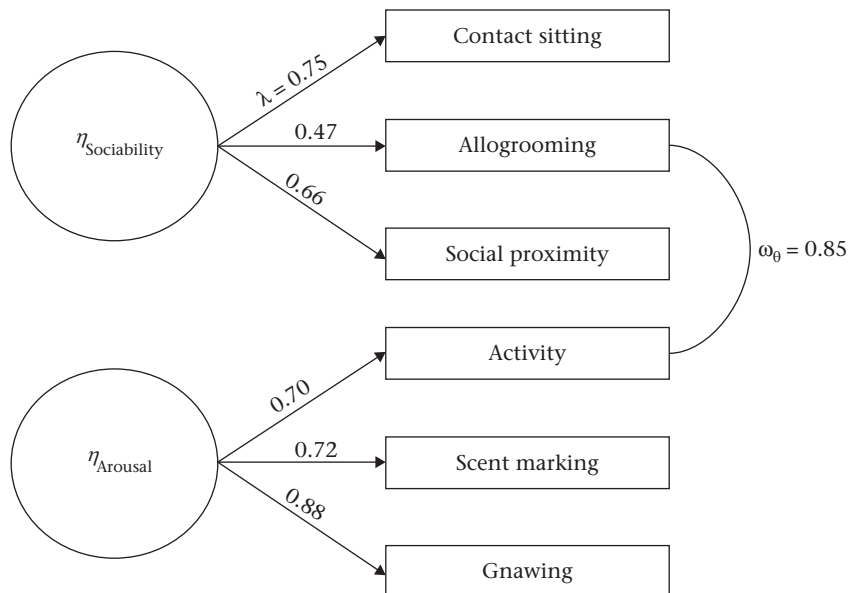

Figure A1. Marmoset behavioural syndrome. Model describing the hypothesized causal and correlational relations among behavioural traits exhibiting personality (i.e. temporally consistent among-individual variation) in marmosets. This so-called generalized network model integrates a latent factor model with a residual pairwise partial correlation network, which was derived using the EGA + GNM statistical framework (Martin et al., 2019). Circles indicate latent factors, while boxes indicate observed traits. Directional arrows encode hypothesized causal effects, and the undirected line indicates a residual pairwise partial correlation network edge. Median posterior estimates are provided for factor loadings and the residual partial correlation . See Martin et al. (2019) for further details.
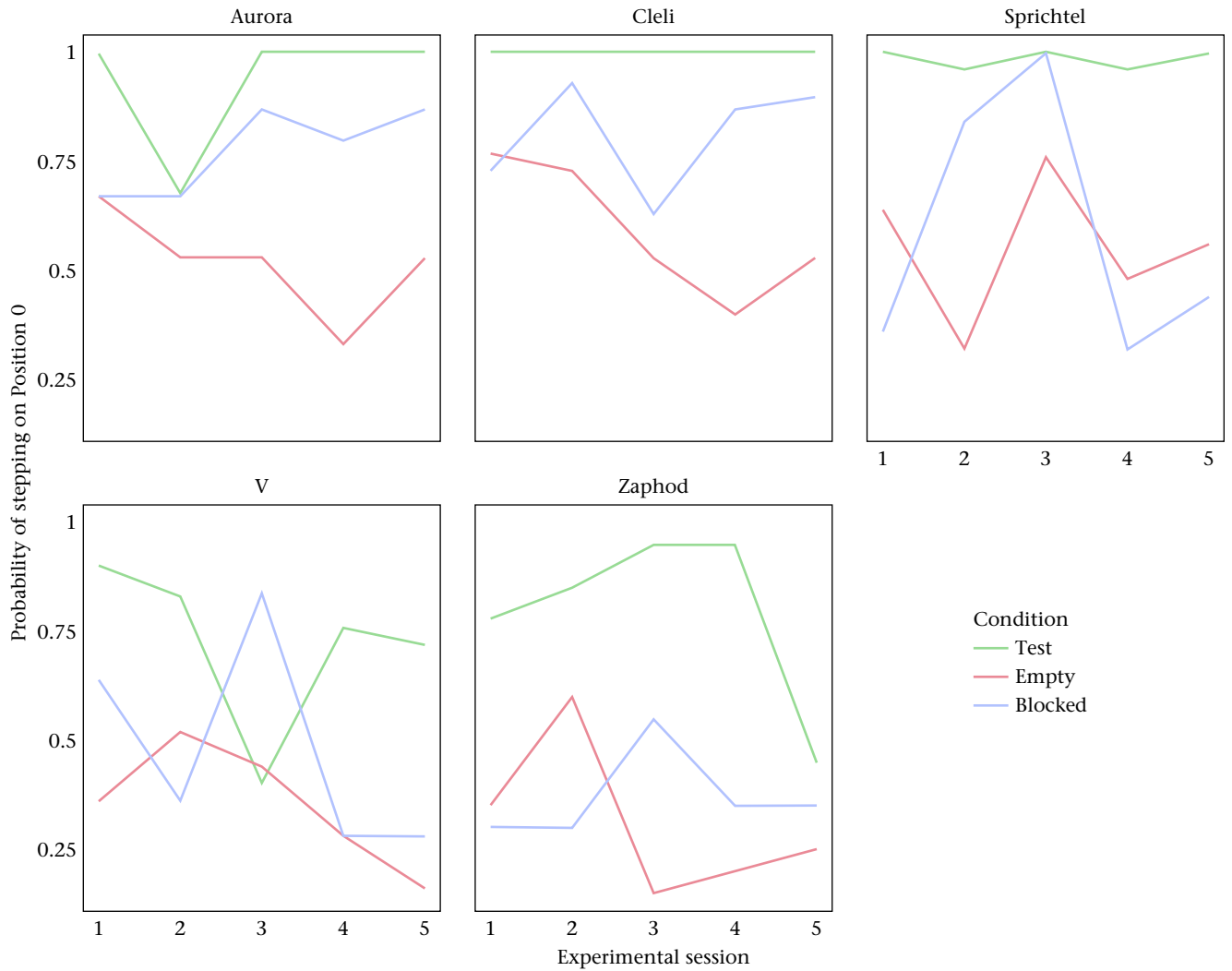
Figure A2. Proportion of trials with stepping on Position 0 across experimental conditions. Proportions are reported to account for differences in the absolute number of experimental sessions across social groups. See Fig. A3 and Table A3 for subject-specific test session results.
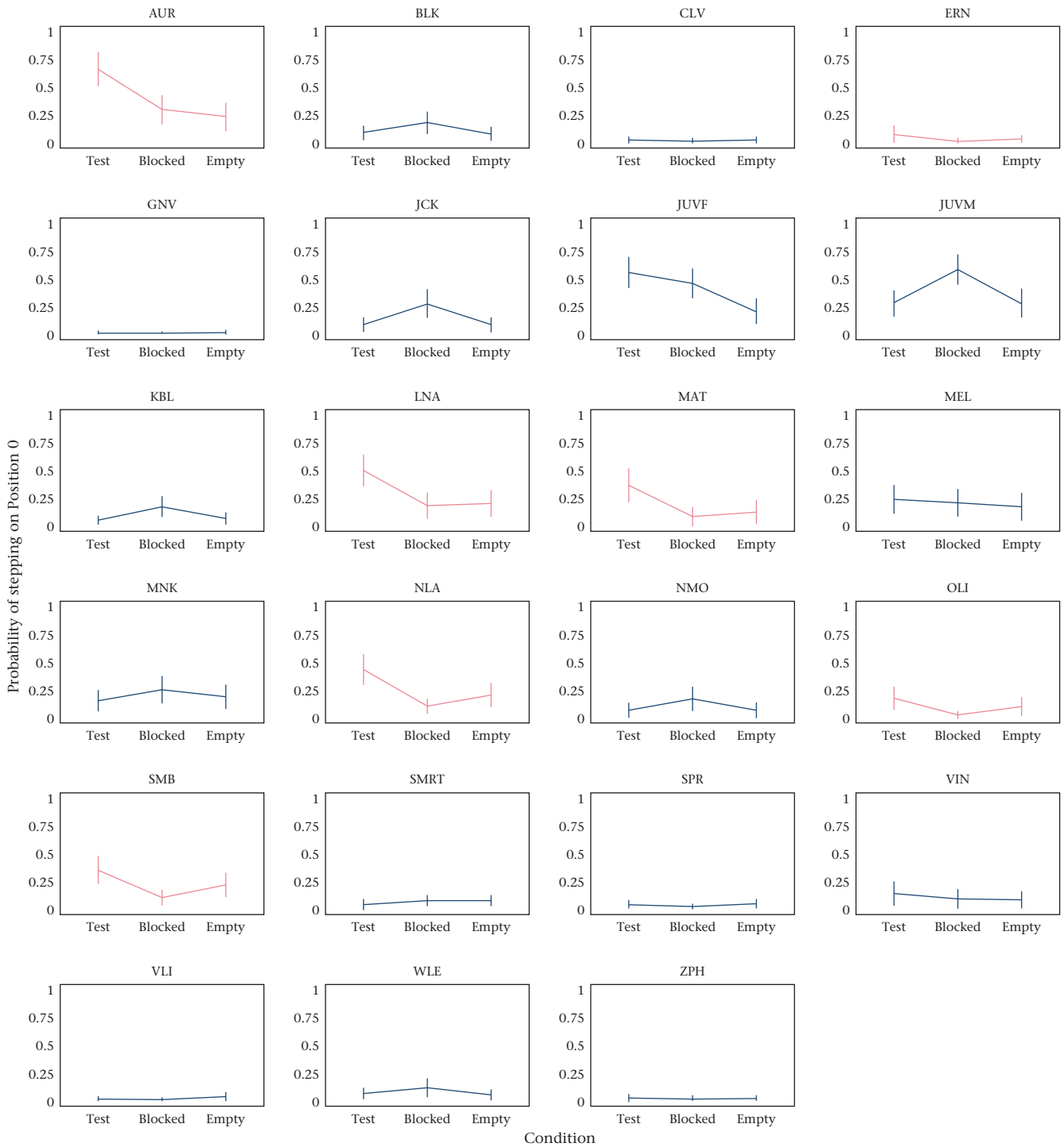
Figure A3. Subject-specific probabilities of stepping on Position 0 platform across experimental conditions. Posterior median ± SD probabilities are shown. These estimates are derived from Model 1 subject random slopes. Pink lines indicate subjects who exhibited evidence of understanding the group service task (i.e. higher test condition stepping probability). See Fig. A4 below for comparisons between the subjects with relatively low (ERN, OLI) and high (AUR, LNA, MAT, NLA, SMB) test session stepping rates. The latter subjects were classified as prosocial.
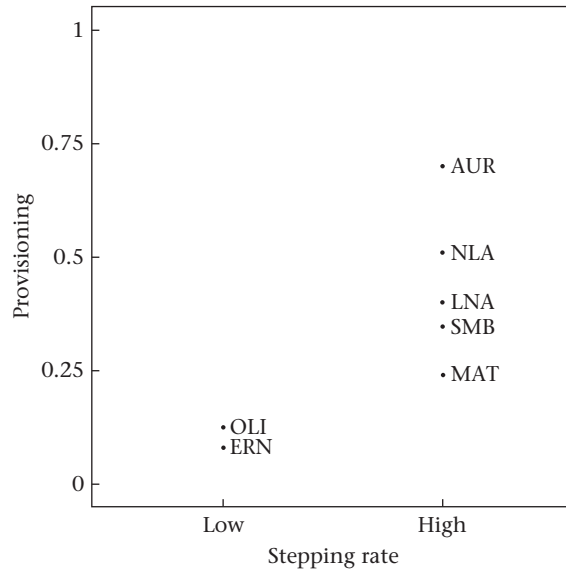
Figure A4. Proportion of provisioning behaviour among subjects passing the group service task. Subjects are categorized by whether they exhibited a high or low stepping rate at the Position 0 platform during group service test sessions (see Fig. A3 and Table A3). The *Y* axis describes the proportion of experimental trials during which the subject successfully provisioned another group member with a food reward. Proportions are reported to account for differences in the absolute number of trials across social groups. Subjects with high stepping rates were classified as prosocial. See Table A1 for subject abbreviations.
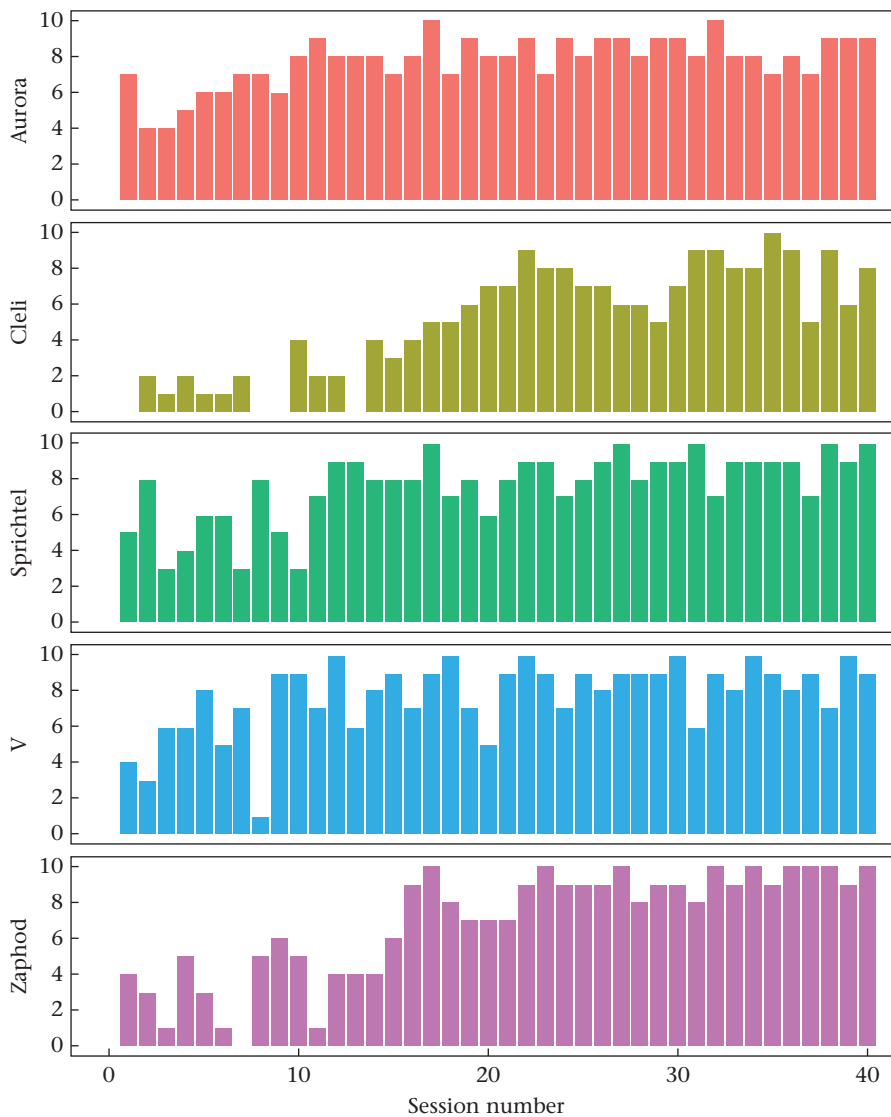

Figure A5. Loose string success across social groups. Raw counts of successful cooperation trials across experimental sessions for each social group. See Table A1 for further details on each social group.
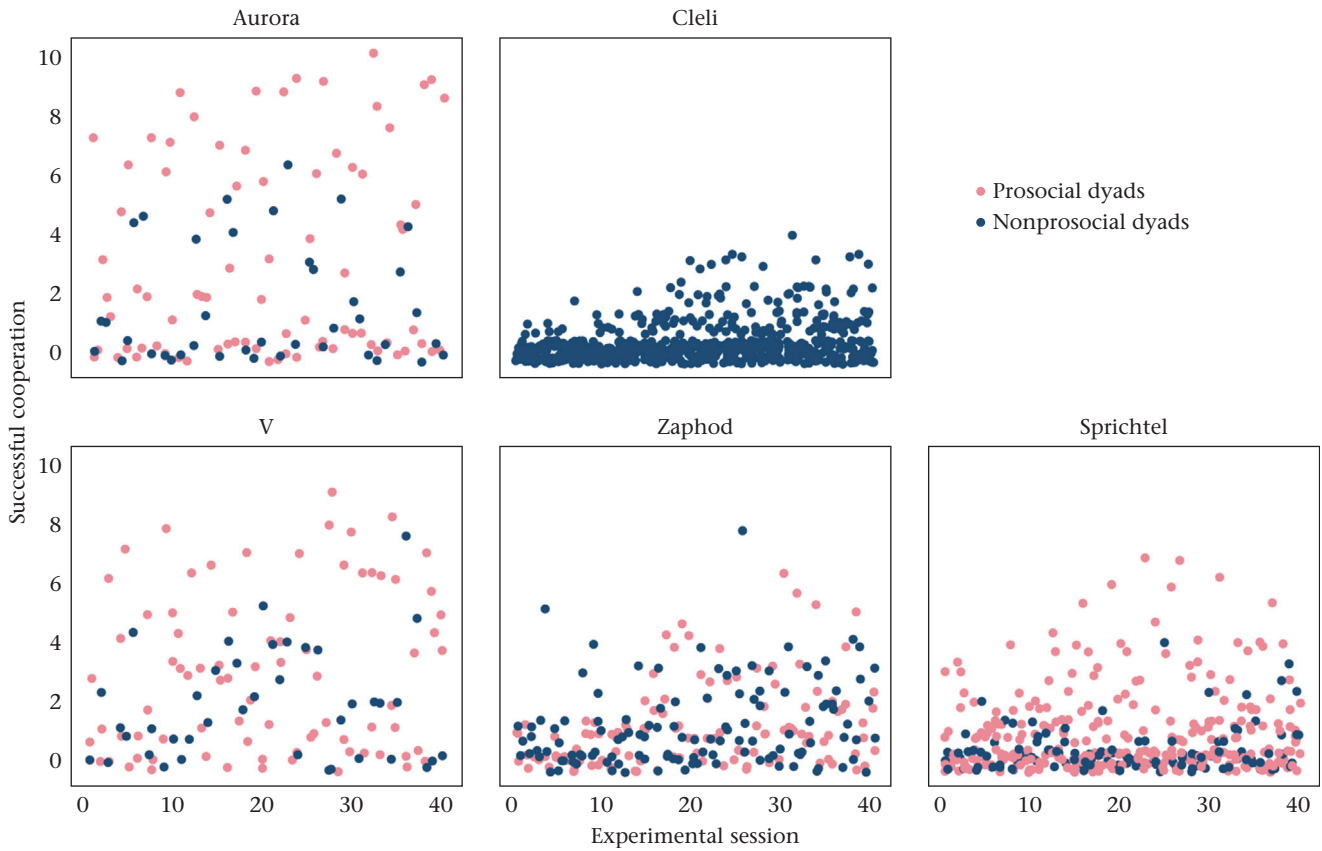
Figure A6. Cooperation as a function of prosociality. Raw data are displayed for counts of cooperative success in the loose string task for dyads containing (pink) and not containing (blue) a prosocial subject. Plots have been faceted for each of our five social groups. Only one group (Cleli) lacked a prosocial individual.
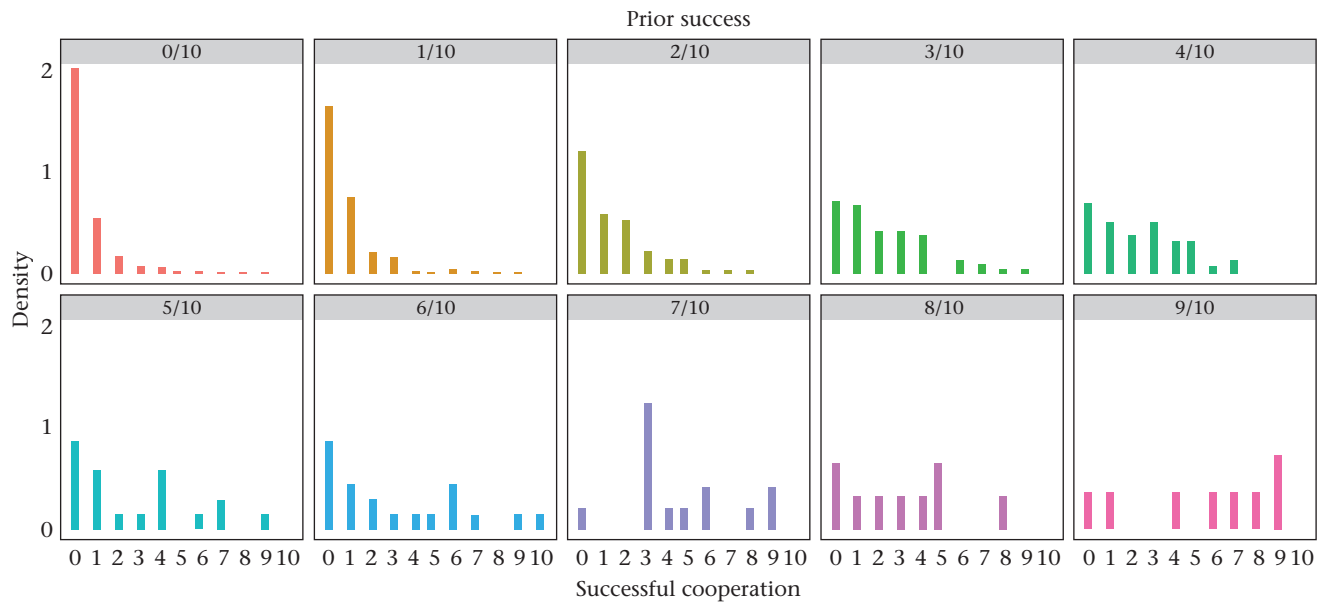

Figure A7. Dyadic cooperation as a function of success in the previous session. Density histograms are displayed for counts of dyadic success in the loose string task, faceted by the number of successful trials a dyad completed during the previous experimental session (0/10 to 9/10). Relative rather than absolute counts are displayed to facilitate more direct comparisons between low and high counts of success, given the much higher occurrence of a low success rate. Success in the first experimental session is not displayed because all dyads had zero previous success. Sessions in which a dyad successfully completed 10/10 loose string trials are also not displayed due to this only occurring in two sessions.