How taking turns communicates desired equality in social relationships

Alicia M. Chen (aliciach@mit.edu) **and Rebecca Saxe** (saxe@mit.edu)
Department of Brain and Cognitive Sciences, MIT

Abstract

When people perform generous acts for each other, they can balance out relative benefits by alternating who is generous. When and why do they do this? Here we test the explanation that sequences of generosity regulate social relationships. We find that people selectively expect reciprocal generosity in equal (vs. hierarchical) relationships, use reciprocal generosity to infer the presence of an equal relationship, and critically expect that people reciprocate generosity in order to communicate a desire for a (more) equal relationship. In a formal planning model, reciprocal generosity can emerge from the value of communicating desired equality.

Keywords: social relationships, generosity, communication, theory of mind, Bayesian modeling

Introduction

People routinely take turns exerting effort to benefit one another. What kind of social cognitive expectations do people have for why people do this?

One set of explanations for reciprocal generosity is that it emerges from people's considerations of how much each partner benefits from interacting. People may care about others' benefits (Bob and Andrew may alternate buying coffee or helping each other with their work projects because they care about each other; Powell, 2022), and/or care about their own downstream benefits (Bob and Andrew might expect that being generous to each other induces and sustains reciprocal interactions that are advantageous to themselves over the long run; Trivers, 1971; Cosmides & Tooby, 1992; Andreoni & Miller, 1993). Additional strategic benefits for the self may arise from developing a public reputation as a generous and potentially valuable partner for others (Roberts et al., 2021), and from establishing a wider norm of generosity in one's community (Bowles & Gintis, 1998).

In addition to caring about their own and their partner's benefits, people may also want the benefits to be distributed equally or fairly between the two people interacting (inequity aversion; Fehr & Schmidt, 1999; Bolton & Ockenfels, 2000). People are more generous to those who have been generous to them in the past (Harris, 1970; Ben-Ner, Putterman, Kong, & Magan, 2004), in part because they want to avoid receiving more benefits than their partner (advantageous inequity aversion; Fehr & Schmidt, 2006). Even children show advantageous inequity aversion, although this motivation develops later and more variably than the desire not to have less than other people (disadvantageous inequity aversion; Blake & McAuliffe, 2011; McAuliffe, Blake, Kim, Wrangham, & Warneken, 2013; Shaw, Choshen-Hillel, & Caruso, 2016).

Yet taking turns is not fully explained by concern for the other, strategic concern for the self, or inequity aversion. All of these goals consider the sum, or balance, of benefits derived from social interactions in the long run. Deliberate turn taking is one way, but not the only way, people can achieve these goals. So why do people alternate?

We propose that one reason people think social partners alternate generous acts is to regulate their relationships (Tatone & Csibra, 2024). Relationship regulation generally refers to the use of social actions to create, sustain, modify, or end a social relationship (Thibaut & Kelley, 1959; Molm, 2010; Rai & Fiske, 2011). We hypothesize that sequences of generous acts are expected to create, sustain or modify the equality or hierarchy of a relationship. More specifically, alternating turns being generous can be interpreted as communicating a desire for a (more) equal relationship. We test this prediction in two steps: first, by measuring the inferences humans make about turn taking for generous actions, and second, by using a model to formalize the reasoning process that promotes alternation.

In this paper, we show that in experiments presenting human observers with naturalistic vignettes, people expect alternating generosity in equal relationships (Experiment 1), and infer the presence of an equal relationship after observing alternating generosity (Experiment 2). Then, we directly ask observers why a character would act generously, either when turn taking or continuing a precedent. We find that communicating about wanting an equal relationship is identified as a primary motivation for alternating generosity (Experiment 3). We formalize this social reasoning process in a Bayesian computational model, where planners consider the above three distributive preferences — vicarious benefits, inequity aversion, and strategic self-interests — as well as the value of communicating about desired equality (based on relative weights on their and their partner's utility) to their partner. In this formal model, we find that a motivation to communicate about desired equality can specifically give rise to reciprocal generosity.

Experiments

Scenarios

We created 18 scenarios, each describing a generous action that one person can do for another in the context of a daily-life interaction (e.g. buying coffee, preparing a meal and cleaning

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

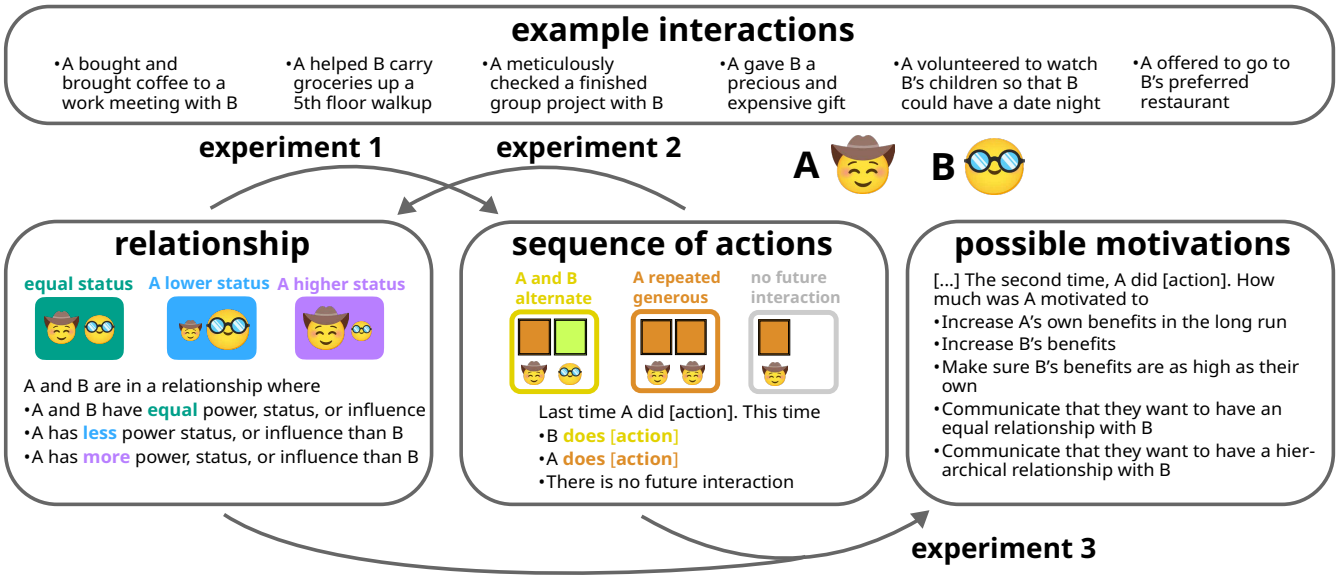

Figure 1: Schematic of experiments, and example stimuli.

after, deferring to the other's preference for cuisine, look after the other's children; Figure 1). We designed the scenarios to cover different kinds and contexts of generous actions. In each interaction, one character gains more benefits and pays less costs than the other (validated by pilot experiments measuring the perceived costs and benefits for both participants in each interaction). For all scenarios, the two characters interacting were the same gender, as indicated by names and pronouns.

Participants All participants were recruited on Prolific and pre-screened to be adult fluent English speakers from the United States. Participants gave informed consent, and all procedures were approved by the MIT Committee on the Use of Humans as Experimental Subjects. For all experiments, participants received $5 for completing the task, for an estimated pay of $15/hour. For each experiment, we excluded participants who did not pass an attention check or indicated that they did not understand the instructions. In total, 231 adults (ages 18-74, M(SD) age = 35.2(11.7) contributed judgments, in three separate waves.

Implementation and open practices All experiments were implemented using the jsPsych library (De Leeuw, 2015), and all analyses were conducted in R using the lme4, lmerTest, and emmeans packages (Bates, Mächler, Bolker, & Walker, 2014; Kuznetsova, Brockhoff, & Christensen, 2017; Lenth, Singmann, Love, Buerkner, & Herve, 2018). All stimuli, data, code, and preregistrations are available at https://osf.io/njcvf/.

Experiments

Methods

Procedure In **Experiment 1**, each participant read 18 scenarios about two characters (A and B) in an equal or hierarchical social relationship, and predicted the sequence of generous interactions. For each scenario, after A did a generous action, participants indicated the probability that the characters expected (1) A would do the same action the next time ('repeating'), (2) B would do that action the next time ('alternating'), and (3) the interaction not happening again.

In **Experiment 2**, each participant read 18 scenarios about two people engaged in a sequence of generous interactions, and inferred whether their social relationship was equal or hierarchical. For each scenario, participants indicated the probability that the characters (A and B) were in (1) a relationship where A has more power, status, or influence than B, (2) a relationship where B has more power, status, or influence than A, and (3) a relationship where A and B have equal power, status, or influence.

In **Experiment 3**, each participant read 16 scenarios[1] about two people engaged in a sequence of generous interactions, and judged the motives of the person who was generous most recently (cf. Kraft-Todd, Kleiman-Weiner, & Young, 2023). The scenarios manipulated the relative status of the person performing the generous action during the second interaction (4: 'lower' vs. 'higher' vs. 'equal' vs. 'just met'), and the sequence of the two social interactions (2: 'repeating' vs. 'alternating'). In the 'just met' condition, the two characters

[1]To ensure proper counterbalancing in Experiment 3, we omitted the two scenarios with the smallest cost plus benefit difference between the two people interacting, as measured in a separate pilot experiment.

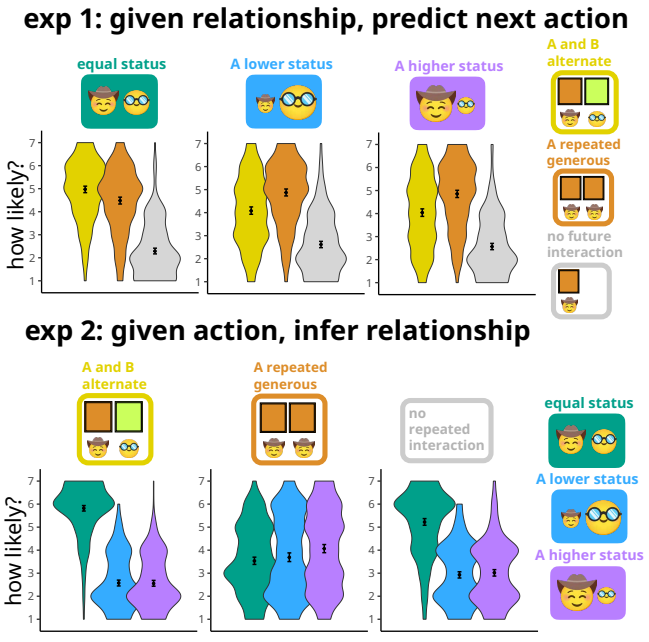

Figure 2: **(A)** Results of Experiment 1. **(B)** Results of Experiment 2. Error bars are bootstrapped 95% confidence intervals.

have just interacted for the first time, so there is no existing social relationship between them. For each scenario, participants evaluated how much the character performing the second generous action was motivated to (1) increase their own benefits in the long run ('self benefit'), (2) increase their partner's benefits ('other benefit'), (3) make sure their partner's benefits are as high as their own ('inequity aversion'), (4) communicate about wanting an equal relationship ('communicate equality'), and (5) communicate about wanting a hierarchical relationship ('communicate hierarchy').

The assignment of scenarios to conditions was balanced across participants. All participants provided judgments using a 7 point Likert scale (from "Extremely unlikely" (1) to "Extremely likely" (7) for Experiments 1 and 2; from "Not at all (1) to "Extremely" (7) for Experiment 3). To test our predictions, we used linear mixed-effects regression models, including categorical effects of the experimental condition and response, along with their interaction, and random intercepts for each scenario and participant.

Results

Experiment 1: People in equal relationships are expected to alternate generosity When the two characters were described as being in an equal status relationship, participants expected that the two characters would alternate generous actions (M = 4.98, SE = 0.094, CI: [4.79, 5.16]) rather than one person being repeatedly generous (M = 4.48, SE = 0.094, CI: [4.30, 4.67], $t(3090) = 4.764$, $p < 0.001$, or the interactions ceasing altogether (M = 2.29, SE = 0.094, CI: [2.10, 2.47]), $t(3090) = 26.089$, $p < 0.001$ (Figure 2). By contrast, when the relationship was hierarchical, no matter whether the generous character was 'higher' or 'lower' status, participants expected the same character to be repeatedly generous (M = 4.86, SE = 0.079, CI: [4.71, 5.02]), more than the characters to alternate (M = 4.06, SE = 0.078, CI: [3.90, 4.22]), $t(3090) = 11.039$, $p < 0.001$.

Experiment 2: People who alternate generosity are perceived to be in equal relationships When the two characters were described as alternating generous actions, participants inferred that their relationship was equal status (M = 5.82, SE = 0.089, CI: [5.64, 5.99]), rather than either character having more or less power, status, or influence (M = 2.56, SE = 0.073, CI: [2.42, 2.71]), $t(3041) = 25.090$, $p < 0.001$ (Figure 2). By contrast, when one character was repeatedly generous, participants inferred that the relationship was unequal (M = 3.88, SE = 0.073, CI: [3.73, 4.02]), rather than equal (M = 3.53, SE = 0.089, CI: [3.36, 3.71]), $t(3041) = 3.858$, $p = 0.002$. The generous character was more likely to be higher power/status/influence (M = 4.07, SE = 0.089, CI: [3.89, 4.24]), than lower power/status/influence (M = 3.69, SE = 0.089, CI: [3.51, 3.86]), t(3041) = 3.664, $p = 0.008$.

Experiment 3: Motives for alternating roles We predicted that social interactions are expected to communicate about desired relationships, and that alternating generosity specifically communicates about wanting an equal relationship. Participants judged that a character who sequentially reciprocated a generous action was motivated to communicate a desire for an equal relationship over a hierarchical relationship, both when averaged across all relationship conditions (estimated mean difference = 2.442, $z = 31.370$, $p < 0.001$) and for each relationship considered separately (all $p < 0.001$) (Figure 3). These motivations were significantly different from those of a character who was repeatedly generous (interaction of 'alternating' vs. 'repeating', and communicate desired 'equality' vs 'hierarchy') both across all prior relationship conditions (estimated mean difference = 4.198, SE = 0.11, $z = 38.140$, $p < 0.001$) and for each prior relationship considered separately (all $p < 0.001$; Figure 3, bottom).

We tested whether communicative motives are perceived selectively when the desired relationship implied by the action is different from the stated relationship. For example, alternating generosity, which implies equality (as shown in Experiments 1 and 2), might seem especially communicative when the scenario states that the two people are in a hierarchical relationship. However, for 'alternating' generosity, the results suggest the opposite: people gave higher ratings to 'communicate equality' when the relationship was equal, compared to when the relationship was hierarchical (diff = 0.714, SE = 0.135, z = 5.296, $p < 0.001$).

Finally, we specifically predicted that communicating

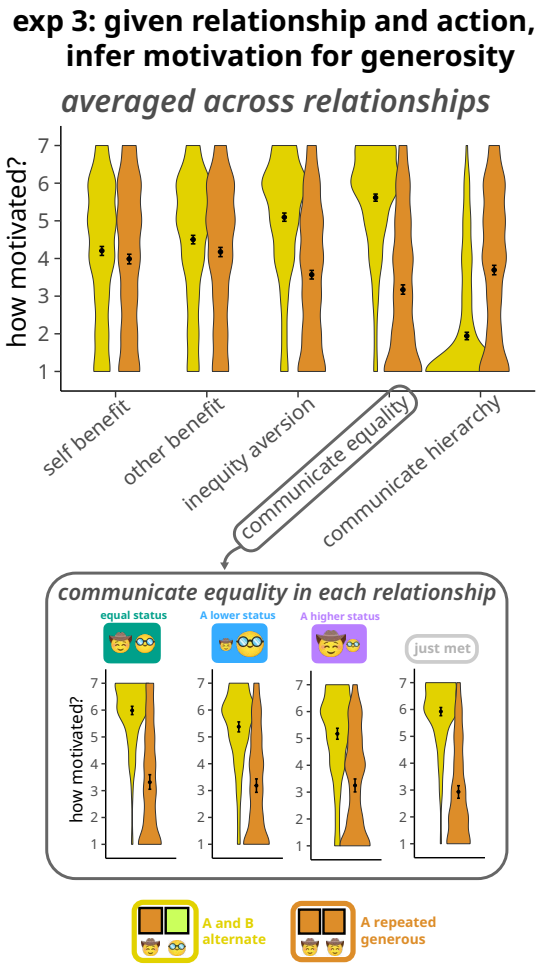

Figure 3: Experiment 3 results. **Top:** Ratings of five motives for generosity, when alternating versus repeating, averaged across the four relationship conditions. **Bottom:** Ratings of 'communicate equality', for each interaction and relationship condition separately. Error bars are bootstrapped 95% confidence intervals.

about wanting an equal relationship is a primary motivation for alternating, relative to other motivations. Indeed, 'communicate equality' was the highest rated motivation for alternating generosity, compared to each of the four other motivations (all $p < 0.001$). In general, this held within each of the relationship conditions considered separately (all $p < 0.003$), with one exception. When the second generous character was higher status, participants did not have a preference for 'communicate equality' over 'inequity aversion' ($p = 0.199$). Additionally, we counted the frequency with which one of the five motivations was rated higher than any other. 'Communicate equality' had the highest rating in the majority of individual trials describing alternation (250 out of 365 trials in which any single motivation had a unique highest rating).

Computational framework

Participants say that a primary motivation for alternating generosity is to communicate about wanting an equal relationship. Next, we developed a formal model of social action planning, to test whether reciprocal generosity could emerge from communicative goals. Our model follows the rational communicative social action (RCSA) framework, analogous to Rational Speech Act (RSA) models of language (Goodman & Frank, 2016; Radkani, Tenenbaum, & Saxe, 2022; Hung, Thomas, Radkani, Tenenbaum, & Saxe, 2022). The premise of the RCSA framework is that in social interactions, people choose actions in part to communicate about their relationships, based on rational recursive reasoning about the inferences observers will make from their actions. Here our goal is to investigate how recursive reasoning about observers' inferences can systematically influence the choice of whether to alternate or repeat roles in a sequence of asymmetric generous interactions.

We model a 'speaker' S and a 'listener' L interacting with each other in two successive similar asymmetric interactions, like those in the experiments' vignettes. Across conditions, we stipulate whether S or L did the generous action in the first interaction, and then we model the planning of S, in the subsequent second interaction. S can choose between two actions $a \in [\texttt{self}, \texttt{other}]$: performing the generous action, or receiving the generous action from L.[2]

Base speaker

We begin by modeling a 'base speaker' S_0 who chooses an action proportional to its corresponding total utility, but does not consider communicative value. The base speaker's utility consists the three non-communicative components that prior research has shown are potential sources of utility from generous actions:

- one's own utility $U_{\text{self}}(a)$
- their partner's utility $U_{\text{other}}(a)$
- inequity aversion, a cost term penalizing unequal outcomes $c(a; a_{\text{first}}) = |U_{\text{self}}(a; a_{\text{first}}) - U_{\text{other}}(a; a_{\text{first}})|$[3], where $U(a; a_{\text{first}})$ is the cumulative utility of the first and second action.

The relative weights of self and partner utility are modulated by $w_{\text{other}} \in [0, 1]$, and the inequity aversion cost term is modulated by w_c. Thus, the utility for S is

$$U_{S_0}(a; a_{\text{first}}, w_{\text{other}}) = (1 - w_{\text{other}}) \cdot U_{\text{self}}(a) + w_{\text{other}} \cdot U_{\text{other}}(a) - w_c \cdot c(a; a_{\text{first}}) \quad (1)$$

[2] We make the (simplified) assumption that the interaction will happen, and either the speaker or listener will be generous, so that the choice is binary.

[3] For simplicity, we display this term as U_{IA}, and w_c as w_{IA}, in Figure 4.

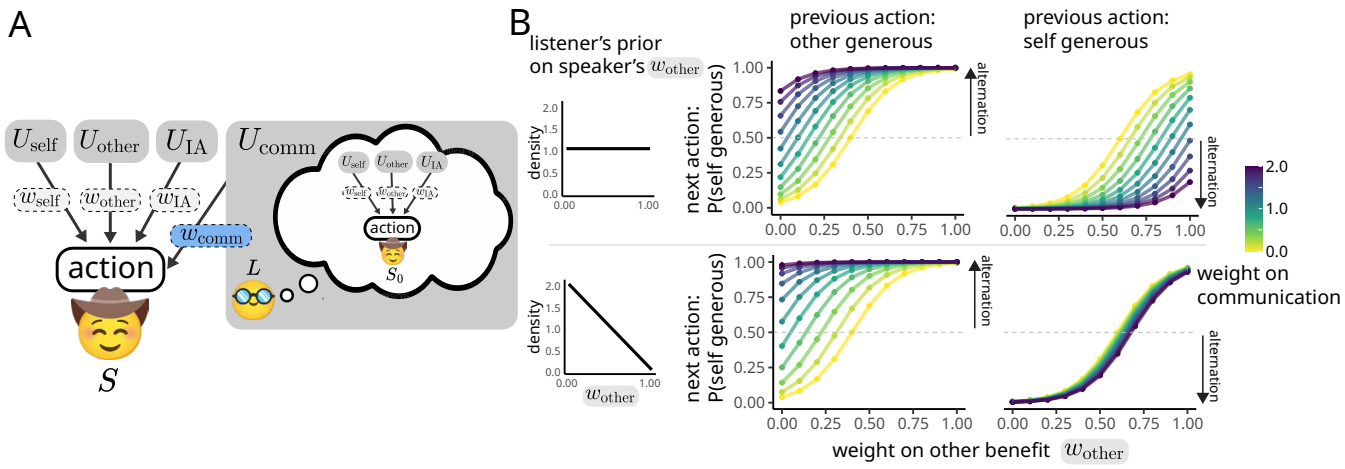

Figure 4: **(A)** Schematic of a model of a communicative speaker who recursively represents a rational listener. **(B)** Model predicted likelihoods of generosity for the second action, faceted by who was generous the first time (columns) and the listener's prior on w_{other} (rows). Increasing the value of communicating desired equality (darker colors) increases the probability of taking turns, and particularly of reciprocating generosity (left column).

The base speaker then uses a softmax decision rule to select an action, modulated by the temperature parameter β:

$$S_0(a|a_{first}, w_{other}) \propto \exp(\beta \cdot U_{S_0}(a; a_{first}, w_{other})) \quad (2)$$

Listener

The listener L observes an action from the base speaker, and uses Bayesian inference and what they observed for the first action to infer what the speaker thinks about w_{other}.

$$L(w_{other}|a, a_{first}) \propto S_0(a|a_{first}, w_{other}) \cdot \exp\left((1 - w_{other}) \cdot U_{self}(a_{first}) + w_{other} \cdot U_{other}(a_{first})\right) \cdot P(w_{other}) \quad (3)$$

assuming w_{other} and the past action a_{first} are independent.

Communicative speaker

Now, we have all the components to build a model of a communicative speaker. In addition to their own utility, their partner's utility, and inequity aversion, the communicative speaker S derives utility from L's inferences about the relative weight of their own and their partner's utility. In other words, we model the communicative speaker as trying to induce an inference of a target weight w_{target} in L. If the communicative speaker is trying to communicate about desired equality between weights, then $w_{target} = 0.5$ (i.e. so that L thinks $w_{other} = w_{target}$). Thus, the communicative speaker's utility is:

$$U_S(a; a_{first}, w_{other}) = \overbrace{U_{S_0}(a; a_{first}, w_{other})}^{\text{base speaker utility}} + w_{comm} \cdot \underbrace{\log L_1(w_{target} = 0.5|a, a_{first})}_{\text{communicative value}} \quad (4)$$

where w_{comm} is the speaker's weight on communicative value; higher weights mean the speaker prioritizes communication more. The communicative speaker chooses an action based on this utility function:

$$S(a|a_{first}, w_{other}) \propto \exp(\beta \cdot U_S(a; a_{first}, w_{other})) \quad (5)$$

Implementation details

We conducted our simulations using the probabilistic programming language WebPPL (Goodman & Stuhlmüller, 2014). Our primary interest was the qualitative effects of the model, so we chose reasonable parameter values. We chose $\beta = 2$, fixed w_c to 0.2, and discretized w_{other} in 100 increments ranging from 0 to 1, and used enumeration for exhaustive inference.

For $a =$ `self` (own generous), we set $U_{self} = 3$ and $U_{other} = 5$. For $a =$ `other` (other generous), we set $U_{self} = 5$ and $U_{other} = 3$, because generosity benefits the other more than oneself. The qualitative results of the simulations are robust to varying the values of action utilities.

Simulation results

In our simulations (Figure 4), we varied w_{other} (the speaker's weight on the listener's utility) and w_{comm} (the speaker's weight on communication), and kept the inequity aversion

term w_c fixed. We show model predictions for listeners with two different priors on w_{other}: a uniform prior, and a Beta$(1,2)$ prior, which puts a higher probability on lower values of w_{other}, capturing a plausible situation where the speaker thinks the listener thinks the speaker likely values themself more than the listener.

Figure 4B shows the model-predicted likelihoods of the speaker choosing to be generous for the next action, given who was generous for the first action. As shown by positive slopes in the plots, increasing w_{other} increases the probability of the speaker being generous (i.e. choosing to benefit the listener), regardless of who was generous in the first interaction. Increasing w_{comm} increases the probability of alternation, particularly if the listener was generous the first time and the speaker is reciprocating the generosity (left column). When the speaker was generous the first time, then increasing w_{comm} increases the probability of alternation (allowing the listener to be generous next time) if the speaker thinks the listener has a flat prior on the speaker's w_{other}, but only weakly if communicating a desire for equality must overcome a prior that the speaker is biased. These results are robust across a variety of values for β. Thus, increasing the speaker's weight on communicating equality can induce alternation, and specifically reciprocal generosity, across a range of plausible conditions.

Discussion

People reported that a primary motivation for alternating generosity is to communicate a desire for a (more) equal relationship. Communicating desired equality was reported to be expected over other classically accepted reasons for reciprocal generosity based on people's distributional preferences (Trivers, 1971; Fehr & Schmidt, 1999). A formal cognitive model, based on inverse planning using an intuitive theory of mind (Baker, Saxe, & Tenenbaum, 2009; Jara-Ettinger, Gweon, Schulz, & Tenenbaum, 2016), revealed that alternating generosity is more likely when a rational planner recursively represents a rational observer, and wants the observer to infer that the planner has equally weighted utilities for themselves and the observer. Thus, given people's distributional preferences, turn taking can emerge from a goal of rationally communicating desired equality.

The experiment and model we developed differ substantially from traditional approaches to studying reciprocal generosity (e.g. Axelrod & Hamilton, 1981; Nowak & Sigmund, 1994; Lau & Mui, 2008). While standard paradigms measure agents' behaviour in abstract anonymous interactions where the outcomes of the game are restricted to the value of the payoffs, we collected human inferences about naturalistic vignettes describing a range of familiar dyadic interactions in which one person benefits more than the other, and which could plausibly recur over time in sustained relationships (Adams & Miller, 2022). Thus the experimental stimuli mimic quotidian situated social interactions, and our model is designed to computationally recapitulate the cognitive processes of humans observing such interactions.

Our cognitive model can be contrasted with evolutionary models that emphasize the credibility of costly signals of generosity, but are agnostic to the proximate mechanism of producing or receiving the signals (Spence, 1973; Zahavi, 1977; Roberts, 1998; Smith & Bird, 2000; Gintis, Smith, & Bowles, 2001; Roberts, 2020). Rational observers likely do indeed consider the costs of communicative actions. In our experiments, people rated alternating actions as more communicative than repeating actions. Why? Consistent with signaling theory, alternation may be more effortful (Schweinfurth & Call, 2019). Accurately alternating requires remembering what happened last time in order to choose what happens next time, whereas following a precedent can be accomplished by a habit or simple rule. The cognitive cost of turn taking may increase its communicative credibility.

A key feature of our model is the focus on relations (as vicarious utility weights), rather than traits, as the content of the communication. We represented this goal as, the planner wants the observer to infer that the planner places equal weight on the self and observer (similar to a high welfare tradeoff ratio, Delton, Krasnow, Cosmides, & Tooby, 2011; Cosmides & Tooby, 2013; Powell, 2022). A plausible alternative formulation, to which our model could be extended, is that the planner aims to create shared mutual knowledge that both partners value the other equally, and both want the other to know it, and both want the other to know that they want them to know it (De Freitas, Thomas, DeScioli, & Pinker, 2019). Another view is that an equal status relationship is one of four innate categories of relationship models (Fiske, 1992), and turn taking is a way to evoke the Equality Matching relationship category. What is shared by all of these views, by contrast to models of reputation signaling (Jones, Pittman, et al., 1982; Schlenker & Pontari, 2000), is that turn taking communicates about a relationship, and not (just) about the traits or abilities of the person being generous (Kaufmann & Clément, 2014; Thomas, Woo, Nettle, Spelke, & Saxe, 2022).

One limitation of our approach is that we isolated a single plan and removed strategic uncertainty. In our experiments and model, there is no possibility of miscoordination (e.g. both people bring two coffees to the meeting, or zero coffees). We also didn't consider additional intentions (e.g. kindness) underlying reciprocity (Falk & Fischbacher, 2006), and bundled all of the strategic motivations for generosity (e.g. increasing future direct reciprocity, communal norms and reputation; Trivers, 1971; Wedekind & Braithwaite, 2002) into a single measure of "long-run benefit" for the self. Finally, over repeated interactions, reciprocity may become a norm, and people may prefer it not because of complex mental-state reasoning over payoffs, but because people prefer it as a social norm (Bicchieri, 2005; Kimbrough & Vostroknutov, 2016). In this sense, the model we present here is incomplete. Future cognitive models should incorporate planners' beliefs about the observer's actions, and about the future consequences of the planner's actions.

Acknowledgements

Thanks to Tony Chen and Josh Tenenbaum for helpful discussions, and to Simon Columbus, Young-eun Lee, Haoyu Du, Jordan Wilke, and Shengyi Wu for comments on an earlier draft of this paper.

References

Adams, M. M., & Miller, J. G. (2022). The flexible nature of everyday reciprocity: reciprocity, helping, and relationship closeness. *Motivation and Emotion*, *46*(4), 461–475.

Andreoni, J., & Miller, J. H. (1993). Rational cooperation in the finitely repeated prisoner's dilemma: Experimental evidence. *The economic journal*, *103*(418), 570–585.

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *science*, *211*(4489), 1390–1396.

Baker, C. L., Saxe, R., & Tenenbaum, J. B. (2009). Action understanding as inverse planning. *Cognition*, *113*(3), 329–349.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Ben-Ner, A., Putterman, L., Kong, F., & Magan, D. (2004). Reciprocity in a two-part dictator game. *Journal of Economic Behavior & Organization*, *53*(3), 333–352.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms*. Cambridge University Press.

Blake, P. R., & McAuliffe, K. (2011). "i had so much it didn't seem fair": Eight-year-olds reject two forms of inequity. *Cognition*, *120*(2), 215–224.

Bolton, G. E., & Ockenfels, A. (2000). Erc: A theory of equity, reciprocity, and competition. *American economic review*, *91*(1), 166–193.

Bowles, S., & Gintis, H. (1998). The moral economy of communities: Structured populations and the evolution of pro-social norms. *Evolution and Human Behavior*, *19*(1), 3–25.

Cosmides, L., & Tooby, J. (1992). Cognitive adaptations for social exchange. *The adapted mind: Evolutionary psychology and the generation of culture*, *163*, 163–228.

Cosmides, L., & Tooby, J. (2013). Evolutionary psychology: New perspectives on cognition and motivation. *Annual review of psychology*, *64*, 201–229.

De Freitas, J., Thomas, K., DeScioli, P., & Pinker, S. (2019). Common knowledge, coordination, and strategic mentalizing in human social life. *Proceedings of the National Academy of Sciences*, *116*(28), 13751–13758.

De Leeuw, J. R. (2015). jspsych: A javascript library for creating behavioral experiments in a web browser. *Behavior research methods*, *47*(1), 1–12.

Delton, A. W., Krasnow, M. M., Cosmides, L., & Tooby, J. (2011). Evolution of direct reciprocity under uncertainty can explain human generosity in one-shot encounters. *Proceedings of the National Academy of Sciences*, *108*(32), 13335–13340.

Falk, A., & Fischbacher, U. (2006). A theory of reciprocity. *Games and economic behavior*, *54*(2), 293–315.

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The quarterly journal of economics*, *114*(3), 817–868.

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism–experimental evidence and new theories. *Handbook of the economics of giving, altruism and reciprocity*, *1*, 615–691.

Fiske, A. P. (1992). The four elementary forms of sociality: framework for a unified theory of social relations. *Psychological review*, *99*(4), 689.

Gintis, H., Smith, E. A., & Bowles, S. (2001). Costly signaling and cooperation. *Journal of theoretical biology*, *213*(1), 103–119.

Goodman, N. D., & Frank, M. C. (2016). Pragmatic language interpretation as probabilistic inference. *Trends in cognitive sciences*, *20*(11), 818–829.

Goodman, N. D., & Stuhlmüller, A. (2014). *The design and implementation of probabilistic programming languages.*

Harris, M. B. (1970). Reciprocity and generosity: Some determinants of sharing in children. *Child Development*, 313–328.

Hung, M. S., Thomas, A. J., Radkani, S., Tenenbaum, J., & Saxe, R. (2022). Modeling risky food sharing as rational communication about relationships. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Jara-Ettinger, J., Gweon, H., Schulz, L. E., & Tenenbaum, J. B. (2016). The naïve utility calculus: Computational principles underlying commonsense psychology. *Trends in cognitive sciences*, *20*(8), 589–604.

Jones, E. E., Pittman, T. S., et al. (1982). Toward a general theory of strategic self-presentation. *Psychological perspectives on the self*, *1*(1), 231–262.

Kaufmann, L., & Clément, F. (2014). Wired for society: Cognizing pathways to society and culture. *Topoi*, *33*, 459–475.

Kimbrough, E. O., & Vostroknutov, A. (2016). Norms make preferences social. *Journal of the European Economic Association*, *14*(3), 608–638.

Kraft-Todd, G. T., Kleiman-Weiner, M., & Young, L. (2023). Assessing and dissociating virtues from the 'bottom up': A case study of generosity vs. fairness. *The Journal of Positive Psychology*, *18*(6), 894–905.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmertest package: tests in linear mixed effects models. *Journal of statistical software*, *82*, 1–26.

Lau, S.-H. P., & Mui, V.-L. (2008). Using turn taking to mitigate coordination and conflict problems in the repeated battle of the sexes game. *Theory and Decision*, *65*, 153–183.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2018). Emmeans: Estimated marginal means, aka least-squares means. *R package version*, *1*(1), 3.

McAuliffe, K., Blake, P. R., Kim, G., Wrangham, R. W., &

Warneken, F. (2013). Social influences on inequity aversion in children. *PloS one*, *8*(12), e80966.

Molm, L. D. (2010). The structure of reciprocity. *Social psychology quarterly*, *73*(2), 119–131.

Nowak, M. A., & Sigmund, K. (1994). The alternating prisoner's dilemma. *Journal of theoretical Biology*, *168*(2), 219–226.

Powell, L. J. (2022). Adopted utility calculus: Origins of a concept of social affiliation. *Perspectives on Psychological Science*, *17*(5), 1215–1233.

Radkani, S., Tenenbaum, J., & Saxe, R. (2022). Modeling punishment as a rational communicative social action. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 44).

Rai, T. S., & Fiske, A. P. (2011). Moral psychology is relationship regulation: moral motives for unity, hierarchy, equality, and proportionality. *Psychological review*, *118*(1), 57.

Roberts, G. (1998). Competitive altruism: from reciprocity to the handicap principle. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *265*(1394), 427–431.

Roberts, G. (2020). Honest signaling of cooperative intentions. *Behavioral Ecology*, *31*(4), 922–932.

Roberts, G., Raihani, N., Bshary, R., Manrique, H. M., Farina, A., Samu, F., & Barclay, P. (2021). The benefits of being seen to help others: indirect reciprocity and reputation-based partner choice. *Philosophical Transactions of the Royal Society B*, *376*(1838), 20200290.

Schlenker, B. R., & Pontari, B. A. (2000). The strategic control of information: Impression management and self-presentation in daily life.

Schweinfurth, M. K., & Call, J. (2019). Reciprocity: Different behavioural strategies, cognitive mechanisms and psychological processes. *Learning & behavior*, *47*(4), 284–301.

Shaw, A., Choshen-Hillel, S., & Caruso, E. M. (2016). The development of inequity aversion: Understanding when (and why) people give others the bigger piece of the pie. *Psychological Science*, *27*(10), 1352–1359.

Smith, E. A., & Bird, R. L. B. (2000). Turtle hunting and tombstone opening: Public generosity as costly signaling. *Evolution and human behavior*, *21*(4), 245–261.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, *87*(3), 355–374.

Tatone, D., & Csibra, G. (2024). The representation of giving actions: Event construction in the service of monitoring social relationships. *Current Directions in Psychological Science*, 09637214241242460.

Thibaut, J. W., & Kelley, H. H. (1959). The social psychology of groups.

Thomas, A. J., Woo, B., Nettle, D., Spelke, E., & Saxe, R. (2022). Early concepts of intimacy: Young humans use saliva sharing to infer close relationships. *Science*, *375*(6578), 311–315.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, *46*(1), 35–57.

Wedekind, C., & Braithwaite, V. A. (2002). The long-term benefits of human generosity in indirect reciprocity. *Current biology*, *12*(12), 1012–1015.

Zahavi, A. (1977). Reliability in communication systems and the evolution of altruism. In *Evolutionary ecology* (pp. 253–259). Springer.