Title

Understanding the representation and dynamics of welfare tradeoff ratios

Permalink

https://escholarship.org/uc/item/8fc7p205

Author

Qi, Wenhao

Publication Date

2024

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA SAN DIEGO

Understanding the representation and dynamics of welfare tradeoff ratios

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy

in

Experimental Psychology

by

Wenhao Qi

Committee in charge:

Professor Lindsey J. Powell, Chair
Professor Judith E. Fan
Professor Chujun Lin
Professor Michael E. McCullough
Professor Joel Sobel

2024

The Dissertation of Wenhao Qi is approved, and it is acceptable in quality and form for publication on microfilm and electronically.

University of California San Diego

2024

DEDICATION

To my love, Peipei, for always inspiring, challenging, and supporting me to be a better person.

TABLE OF CONTENTS

Appendices

Appendices

Appendices

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to thank my first advisor, Ed Vul. He was the reason I came to UCSD and had been the shepherd of my academic career for the first three years of graduate school. Whenever I got stuck in writing, focused too much on details to see the big picture, or had an overly optimistic model of the participants or readers (which, regrettably, happened a lot), all it took was a brief meeting with him to put me back on the right track. I am especially grateful for his appreciation of my nerdiness and gentle pushback when I took it too far.

I would like to thank my second advisor, Lindsey Powell. Although I joined her lab in my fourth year and did not work on anything developmental, she treated me as a native lab member and mentee, and has supported me in every step as I explored the somewhat turbulent transition into a different field. I learned a lot from her kindness and respect for everyone, her sense of responsibility, and her balance of priorities. Without her, my papers would look more like boring technical reports than social psychological research.

I would like to thank the other members of my committee: Mike McCullough, Judy Fan, Joel Sobel, and Chujun Lin. Mike is the first adopter and an ardent advocator of the Lambda Slider we developed in Chapter 1, even before the paper was written. Judy has always expressed interest in whatever I work on and plan to work on, and challenged me to think deeper. Joel is extremely reliable, and has provided invaluable feedback on the game theoretical aspects of this dissertation. Chujun joined my committee without hesitation when I was in desperate need of an additional member. Thank you all for supporting me through this journey.

I would like to thank my former mentee and collaborator, Boyu Wang. He was the smartest undergraduate student I have worked with, and played a critical role in developing and validating the ideas in Chapter 3. I will forever remember our joyful discussions in the former Vul lab.

I would like to thank my lab mates in the EVul and SoCal labs: Erik, Lauren O., Isabella, Cameron, Alexis, Lauren S., and Bill. You gave me the feeling of an academic family despite the solitude of doing research.

I would like to thank my SD friends: Leg Wang, God Liang and God Yao. My PhD life would be utterly boring without you.

I would like to thank my parents for supporting me, emotionally and financially, to do whatever makes me happy.

Finally, I would like to thank my girlfriend, wife, and soul mate, Peipei, for being the light in my life.

Chapter 1, in full, has been submitted for publication of the material as it may appear in “Qi, W., Vul, E., & Powell, L. J. (under review). An accurate and efficient measure of welfare tradeoff ratios. *PLOS One*.” The dissertation author was the primary investigator and author of this paper.

Chapter 2, in full, is a reprint of the material as it appears in “Qi, W., & Vul, E. (2022). The evolution of theory of mind on welfare tradeoff ratios. *Evolution and Human Behavior*, *43*(5), 381–393.” The dissertation author was the primary investigator and author of this paper.

Chapter 3, in full, is currently being prepared for submission for publication of the material. “Qi, W., Wang, B., & Powell, L. J. (in preparation). The joint evolution of theory of mind and reciprocity in noisy games.” The dissertation author was the primary investigator and author of this paper.

VITA

2019 Bachelor of Engineering, Automation, Tsinghua University

2024 Doctor of Philosophy, Experimental Psychology, University of California San Diego

ABSTRACT OF THE DISSERTATION

Understanding the representation and dynamics of welfare tradeoff ratios

by

Wenhao Qi

Doctor of Philosophy in Experimental Psychology

University of California San Diego, 2024

Professor Lindsey J. Powell, Chair

Human society is built on cooperative relationships. Humans have evolved sophisticated mental representations and mechanisms to cooperate effectively in a noisy and changing world. One such representation is the welfare tradeoff ratio (WTR)—the weight one places on another person's welfare relative to her own. People not only use WTRs to make social decisions consistently in varying circumstances, but can also infer others' WTRs from their actions, allowing them to reciprocate by adjusting their own WTRs. In this dissertation, I investigate the mechanisms underlying the dynamics of WTRs and build tools to support such investigation. In Chapter 1, I develop an accurate and efficient measure of WTRs, called the Lambda Slider, which, in contrast to previous measures, requires only one response from participants. In three experiments, I show that the Lambda Slider has high reliability and validity, enabling fine-grained investigation of the dynamics of WTRs over time or space. In Chapter 2, I explore the evolutionary origins of the capacity of WTR inference, a basic form of "theory of mind" that supports reciprocity and partner choice. Through evolutionary game simulations, I identify two environmental requirements for WTR inference to evolve—stable opponents and variable payoff structures. Using behavioral experiments, I show that people do perform WTR inference in such an environment. In Chapter 3, I strengthen the argument in Chapter 2 by focusing on games without strong interdependence, a more common occurrence in the real world, and considering reciprocity in the form of WTR inference and adjustment. I found that a reciprocal agent with WTR inference performs the best in both noiseless and noisy environments, but has a unique advantage only in a noisy environment with uncertainty about the payoffs perceived by the opponent. This points to the role uncertainty plays in the evolution of theory-of-mind capacities. In Chapter 4, I discuss some future directions. Overall, this work lays the groundwork for studying the dynamics of WTRs in a fine-grained way and understanding the evolution of sophisticated mental representations underlying people's social life.

Introduction

Human society is built on cooperative relationships, in which one person is willing to sacrifice some personal welfare to benefit another person. Mutual cooperation allows a group of people to be better off than selfish individuals who only care about their own welfare. The ideal situation is that each individual tries to maximize the total welfare of the group, which, in a changing world, only requires the ability to calculate the total welfare embedded in each social decision.

However, a constant threat to cooperation is defectors who enjoy the benefit of others' cooperation but do not contribute, or contribute less, to the public good. The maintenance of mutual cooperation requires identifying and punishing defectors, either by not cooperating with them in future interactions or by ostracizing them (Axelrod, 1984; Nowak, 2006b; Trivers, 1971). It is relatively easy to identify a defector who is entirely selfish, but much harder to identify a subtler defector who is only slightly less cooperative than the ideal level of cooperation, especially given the noise and uncertainty in the real world. This motivates partial defection. Such subtle defection requires a sophisticated cognitive ability—to calculate a weighted sum of one's own and others' potential payoffs, incorporating the degree to which the defector wants to discount others' rewards. This calculation can be expressed in terms of a welfare tradeoff ratio (WTR), the weight one places on another person's welfare relative to her own (Tooby & Cosmides, 2008). Perfect cooperation corresponds to a WTR of 1, complete defection corresponds to a WTR of 0, and subtle defection corresponds to a WTR between 0 and 1.

Subtle defectors can thrive in a population of perfect cooperators as long as they are not detected, gradually eroding the cooperation. Detecting subtle defection requires an even more sophisticated cognitive ability—to infer another person's WTR from their actions—which enables reciprocity in the form of adjusting one's own WTR in light of such inference. A central hypothesis of this dissertation is that the interplay between subtle defection and detecting subtle defection gives rise to a graded representation of WTRs shared between these two functions. The result of this evolutionary process is a complex dynamics of setting, inferring, and adjusting WTRs in response to a variety of information in the social environment (Ackermann et al., 2016; Jones & Rachlin, 2006; Lim, 2012; Piff et al., 2010; Qi et al., under review; Van Lange et al., 1997), which we have only started to uncover.

Two complementary ways of studying WTR psychology are behavioral and evolutionary. Behavioral studies examine how people set, infer, and adjust WTRs, usually in controlled experiments. Evolutionary studies model social interactions as games and examine what behavioral and cognitive strategies evolve in different game environments. This dissertation fills gaps in each of these approaches.

Behavioral studies rely on a good measure of participants' WTRs toward specific partners. Such a measure should be accurate and sensitive so that we can detect small differences in WTRs. It should also be efficient so that we can track changes in a participant's WTR toward a partner throughout an experiment or variations in a participant's WTRs toward many different targets. Previous measures of WTRs often require multiple responses from participants to obtain a single measurement, and have limited sensitivity because they are inherently discrete (Kirkpatrick et al., 2015; Liebrand, 1984; Murphy et al., 2011; Van Lange et al., 1997). In Chapter 1, I develop a new measure, called the Lambda Slider, which requires only one response and is inherently continuous. The key innovation is using a nonlinear payoff function instead of the traditional linear ones. In three experiments, I show that the Lambda Slider has highly reliability and validity, ready to be used in a variety of WTR research.

The concept of WTR is largely missing in previous evolutionary models of cooperation because most models only consider games with fixed payoff structures, such as fixed Prisoner's Dilemmas or public goods games, where making decisions based on WTRs can be reduced to behavioral strategies, e.g., to cooperate or defect (Axelrod, 1984; Nowak, 2006b). This is not the case in the real world, where each social interaction has a different payoff structure—lending an extra umbrella to someone on a rainy day and jumping in the sea to save a drowning person involve vastly different costs and benefits. Effective cooperation, subtle defection, and detecting subtle defection in such varying circumstances likely require online computation based on WTRs that takes into account the payoffs specific to each decision. But without directly modeling variable games, it is unclear under what conditions WTR computation and inference can evolve—e.g., if every game is a Prisoner's Dilemma with slightly different payoffs but clear labels for cooperation and defection, WTR computation is still likely unnecessary. In Chapters 2 and 3, I show that modeling variable games allows us to identify the conditions for the evolution of more sophisticated mental representations and mechanisms, such as graded WTR inference and adjustment.

In Chapter 2, inspired by the structures of one-shot games and fixed repeated games, I expand the space of game environments along two dimensions: the variability of opponents and the variability of payoff structures. I find that the only environment where WTR inference creates a unique evolutionary advantage is the one with stable opponents and variable payoff structures. When the opponents are variable (i.e., two agents never interact more than once, or at least they do not track each other's identity), inferring the opponent's WTR has no influence on future interactions. When the payoff structures are fixed or stable, action-level strategies work well, including a reinforcement-learning agent that learns the best action sequence under each particular opponent–payoff structure combination. In an environment with stable opponents and variable payoff structures, the agent with WTR inference performs the best. This is because having a higher-level model of the opponent allows the agent to better predict the opponent's actions. In games with strong interdependence, where one player's best action depends on the other player's simultaneous action, better predictions lead to higher utilities for the agent with WTR inference.

However, games with strong interdependence may be rare in the real world. Most of our social decisions are not influenced by a simultaneous decision by someone else; we make each decision by evaluating the costs and benefits created by that decision alone. In Chapter 3, I explore the conditions under which WTR inference can evolve using games without strong interdependence. I hypothesize that the evolution of WTR inference is driven by the need to reciprocate in a way similar to tit-for-tat in the iterated Prisoner's Dilemma (Axelrod, 1984). I show that an agent with graded WTR inference and reciprocity performs the best in both noiseless and noisy environments, but has a unique advantage only in a noisy environment with some uncertainty about the payoffs perceived by the opponent. This form of uncertainty is ubiquitous in the real world—when someone lends me an umbrella, I may be unsure whether it is her only umbrella or if she has a spare one, which entails different costs for her. In the noiseless environment, a much simpler strategy with a binary conception of cooperation and defection performs almost as well, reducing the likelihood that graded WTR inference can evolve there. These results suggest that specific forms of uncertainty in the world might have driven the evolution of more sophisticated mental representations and processes.

In Chapter 4, I conclude with a brief discussion of some future directions for addressing questions unanswered by this dissertation. Overall, this work lays the groundwork for studying the dynamics of WTRs in a fine-grained way and understanding the evolution of sophisticated mental representations underlying people's social life.

Chapter 1

An accurate and efficient measure of welfare tradeoff ratios

People's decisions are affected by their interest in others' welfare. They can be motivated both to help and to harm others. The direction and magnitude of these motivations can be quantified relative to a person's self-interest as a welfare tradeoff ratio (WTR). This construct is valuable for testing quantitative theories of social motivation. However, most existing measures of WTRs are based on multiple choices between discrete sets of payoffs, which forces a tradeoff between the accuracy and efficiency of the measures. Here we introduce the Lambda Slider, a WTR measure that is simultaneously accurate and efficient. A participant uses a linear slider to choose from a continuous range of payoff allocations for herself and her social partner. The underlying payoff functions for self and other create a one-to-one correspondence between the participant's potential WTR values and the slider positions that she could choose, which enables accurate measurements of WTR from a single response. Across three experiments, we show that a single response on the Lambda Slider has high reliability, high convergent validity with other measures of social motivation, and high external validity for an altruistic decision with real-world consequences. The Lambda Slider is easy to implement and can be applied in a wide variety of studies on the forces that shape social motivation.

1.1 Introduction

People's lives are full of choices that affect both their own welfare and others' welfare. For example, the decision to give your coat to another person on a cold winter night decreases your own welfare but increases that person's welfare. People's decisions in such interdependent situations are driven by a variety of social motivations, including an interest in social norms and reputation along with direct concern for others' well-being (Camerer, 2011; Fehr & Schmidt, 1999).

The sum of these social forces results in an overall motivation to increase or decrease another person's welfare; i.e., to benefit or harm that person. The direction and magnitude of this motivation can be captured as a welfare tradeoff ratio (WTR): the amount of personal welfare one is willing to give up in order to increase or decrease another person's welfare by a specified amount (Tooby & Cosmides, 2008). Formally, if Alice has a relationship with Bob, then we can express Alice's utility for a given decision as

$$u = w_{\mathrm{s}} + \lambda w_{\mathrm{t}}, \tag{1.1}$$

where w_{s} ("s" stands for "self") is Alice's resulting welfare (her actual or expected payoff from the decision), w_{t} ("t" stands for "target"[1]) is Bob's welfare (his actual or expected payoff), and λ is Alice's welfare tradeoff ratio toward Bob. For conciseness, we will use λ to represent welfare tradeoff ratios throughout the paper. A higher λ indicates stronger altruism or friendliness on the part of Alice toward Bob, as it means that Alice will favor actions or situations that are good for Bob, even at the expense of some of her own welfare. In contrast, a lower λ indicates stronger selfishness or dislike. A negative λ would mean that Alice could perceive utility in sacrificing some of her own welfare in order to harm Bob.

[1] We use "target" instead of the usual "other" due to the confusability between the letter o and the number 0 as subscripts, and that the letter t happens to be the alphabetical successor of the letter s.

λ varies across people and their social partners; some people feel more concern for others in general, and people also care more about some social partners than others (Jones & Rachlin, 2006; Lim, 2012; Van Lange et al., 1997). Higher λs promote cooperation, generosity, and altruism, while lower or even negative λs can lead to spiteful actions that are costly for both the actor and their target. Given the significance of interdependent decisions for the well-being of individuals and societies, an important goal for social psychology and behavioral economics has been to understand the factors that impact people's concern for others' welfare (Almlund et al., 2011; Henrich et al., 2001; Thielmann et al., 2020; Van Lange et al., 1997). Such research is predicated on good measurements of the dependent variable, which can usually and fruitfully be conceptualized as a λ value since it is well-defined and generalizable across situations.

1.1.1 Binary allocation tasks

An ideal tool for measuring λ would be both accurate[2] and efficient. This would make it feasible for researchers to measure meaningful differences or changes in λ across many people, partners, or situations. For example, an accurate and efficient measure of λ could allow researchers to study how λ changes as social partners build a history of reciprocation by quickly and repeatedly sampling across interactions (Ackermann et al., 2016). Or it could allow researchers to study how λ reflects positions and connections among many people in a social network (Leider et al., 2009).

How can we measure λ? The simplest way is through a "binary allocation task" (Messick & McClintock, 1968). If we want to measure Alice's λ toward Bob, we can give Alice two allocation options to choose from (Fig. 1.1A). Option A results in \$5 for Alice and \$0 for Bob ($w_s = 5$, $w_t = 0$), while Option B results in \$0 for Alice and \$10 for Bob

[2] For simplicity, in most of this paper, "accuracy", "accurate" or "accurately" entails the technical concepts of both accuracy (or unbiasedness) and precision (or reliability); i.e., a measure needs to be both accurate and precise in order to be called "accurate".

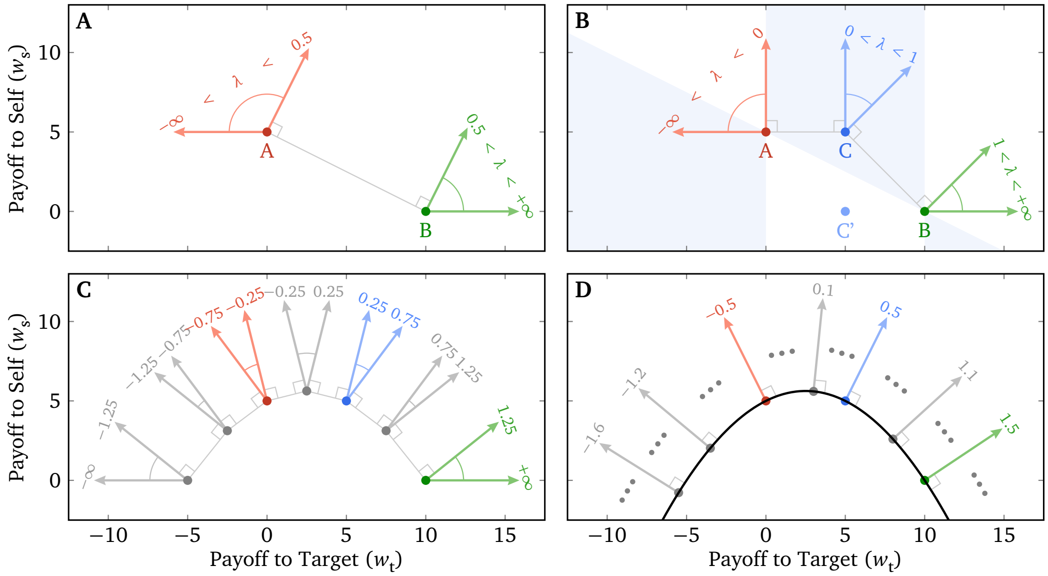


Figure 1.1. From binary allocation tasks to the Lambda Slider. (**A**) A binary allocation task, where the threshold of λ for switching between Options A and B is $\hat{\lambda} = 0.5$. The arrows point in the direction of the gradient of the utility function (Eq. (1.1)) for a given λ. (**B**) Adding a third option C to create a triple-dominance task, which is equivalent to two binary allocation tasks with $\hat{\lambda} \in \{0, 1\}$, but requires only one response. The shaded areas are all the locations where Option C can be placed in order for the task to be triple-dominance. C′ is a hypothetical third option that does not form a triple-dominance task with A and B because they do not fall along a strictly concave function $w_s = f(w_t)$. (**C**) An illustration of a hypothetical "septuple-dominance task" in which each option would be preferred for some range of λ. (**D**) A possible option space for a Lambda Slider, where a participant can choose any point on the curve (via a slider; Fig. 1.2A). Each point on the curve corresponds to a unique λ whose corresponding utility gradient is perpendicular to the tangent of the curve at that point.

($w_s = 0$, $w_t = 10$)[3]. If $\lambda > 0.5$, Alice will choose Option B since it leads to a higher overall utility for herself than Option A. If $\lambda < 0.5$, then the overall utility of Option A is higher, and Alice will be more likely to choose that option instead. Therefore, Alice's decision on this allocation task tells us whether her λ toward Bob is above or below the threshold $\hat{\lambda} = 0.5$. This threshold tested by the task can be adjusted by changing the payoff values involved in the two allocation options.

However, a single binary allocation task has very low sensitivity, defined as the inverse of the smallest change that can be detected by the measure. A binary allocation task with $\hat{\lambda} = 0.5$ cannot distinguish among different λs above, or below, 0.5. By analogy, refusing to give your coat to another person could reflect any λ ranging from valuing your own warmth just a little more than theirs to actively wishing for them to be cold. The accuracy of a measure is upper-bounded by its sensitivity. To gain a higher sensitivity in our overall measurement of λ, we can give Alice *multiple* binary allocation tasks with different $\hat{\lambda}$s. For instance, if we assume λ falls between -2 and 3 and want a measure that gets within 0.5 of the correct value, we need 9 tasks with $\hat{\lambda} \in \{-1.5, -1, \ldots, 2.5\}$. If we aim to get within 0.1 of the correct value, then the number of tasks goes up to 49. This illustrates the inevitable tradeoff between sensitivity and efficiency when we measure λ with binary allocation tasks.

Most existing measures of λ (Delton et al., 2023; Jones & Rachlin, 2006; Kirkpatrick et al., 2015), or related constructs such as social value orientation (SVO; Liebrand, 1984; Liebrand and McClintock, 1988; Messick and McClintock, 1968; Murphy et al., 2011; Sonnemans et al., 2006; Van Lange et al., 1997), share the logic of narrowing down λ with multiple binary allocation tasks, thus sharing the tradeoff between sensitivity and efficiency. This can result in study designs in which many participants must be

[3] Here we assume a linear relationship between monetary payoffs and welfare, and that the same increase in payoff leads to the same increase in welfare for both oneself and the target. In practice this may not be exactly true (Kahneman & Tversky, 1979), but in our experiments we try to minimize sources of nonlinearity.

recruited to study the effects of only a few factors on social motivation (e.g., Hall et al., 2021). (For a comprehensive review of the measures in the SVO literature, see Murphy and Ackermann (2014).)

1.1.2 Triple-dominance tasks

Can we achieve the level of sensitivity of many binary allocation tasks with only a few responses or even one response from the participant? We can draw inspiration from the Triple-Dominance Measure (Van Lange et al., 1997). Although a triple-dominance task is equivalent to *two* binary allocation tasks, a participant only needs to make *one* response on the measure. We can create a triple-dominance task by adding a third option to the binary allocation task in Fig. 1.1A: Option C results in \$5 for Alice and \$5 for Bob (Fig. 1.1B). With these three options, Alice will choose A if $\lambda < 0$, B if $\lambda > 1$, and C if λ is between 0 and 1. Therefore, this triple-dominance task is equivalent to *two* binary allocation tasks but Alice only needs to make *one* decision by choosing the best option among the three.

Allocation options in a triple-dominance task must be selected such that for any given λ, one option dominates (i.e., results in a higher utility than) the other two, and for each option there exists some λ such that the given option is dominant. In order to maintain these features, the three options need to fall along a *strictly concave* function $w_{\mathrm{s}} = f(w_{\mathrm{t}})$. This means that Options A and B constrain the possible payoffs offered in Option C, as illustrated by the shaded areas in Fig. 1.1B. As a counterexample, consider Option C′ in Fig. 1.1B, which corresponds to $w_{\mathrm{s}} = 0$ and $w_{\mathrm{t}} = 5$. Given Options A, B and C′, Alice will choose A if $\lambda < 0.5$ and B if $\lambda > 0.5$, but she will never choose C′, so the task does not satisfy the criterion that for each option there exists some λ such that the given option is dominant.

1.1.3 Lambda Slider

By the same logic, we can add more allocation options to a single-choice task to gain a higher sensitivity in the measurement of λ. Fig. 1.1C is a hypothetical example of a “septuple-dominance task” with 7 options, and the sensitivity of the measurement is $\frac{1}{0.5}$ for $\lambda \in (-1.25, 1.25)$. The options still need to fall along a strictly concave function $w_\text{s} = f(w_\text{t})$ to ensure that each option corresponds to the best choice given some λ. If we keep adding options to the task, we can create a smooth, continuous curve in the w_s–w_t space (Fig. 1.1D), with each point on the curve corresponding to a *single* λ. This one-to-one correspondence (bijection) between potential λs and points on the curve results in a (theoretically) infinite sensitivity of the measurement, which makes it possible to accurately measure a participant's λ toward a particular social partner from a single choice.

We can present such a continuous set of allocations to the participant with a slider (Fig. 1.2A), and we call it the Lambda Slider (see Appendix 1.A for a formal definition and Appendix 1.B for comparison with a related measure, the Circle Test (Sonnemans et al., 2006)). The rewards allocated to the participant and target, w_s and w_t, are both continuous functions of the slider position x, and we call $w_\text{s}(x)$ and $w_\text{t}(x)$ the payoff functions of the slider.

We can choose the payoff functions such that the slider position x that a utility-maximizing participant chooses (denoted x^*) is an *identity function* of her λ toward the social partner in question. Consequently, the slider position is a direct measure of λ and no additional calculation is required. One class of such payoff functions (and arguably the simplest class; see Appendix 1.A) is

$$w_\text{s}(x) = -ax^2 + b_\text{s}, \tag{1.2}$$

$$w_\text{t}(x) = 2ax + b_\text{t}, \tag{1.3}$$

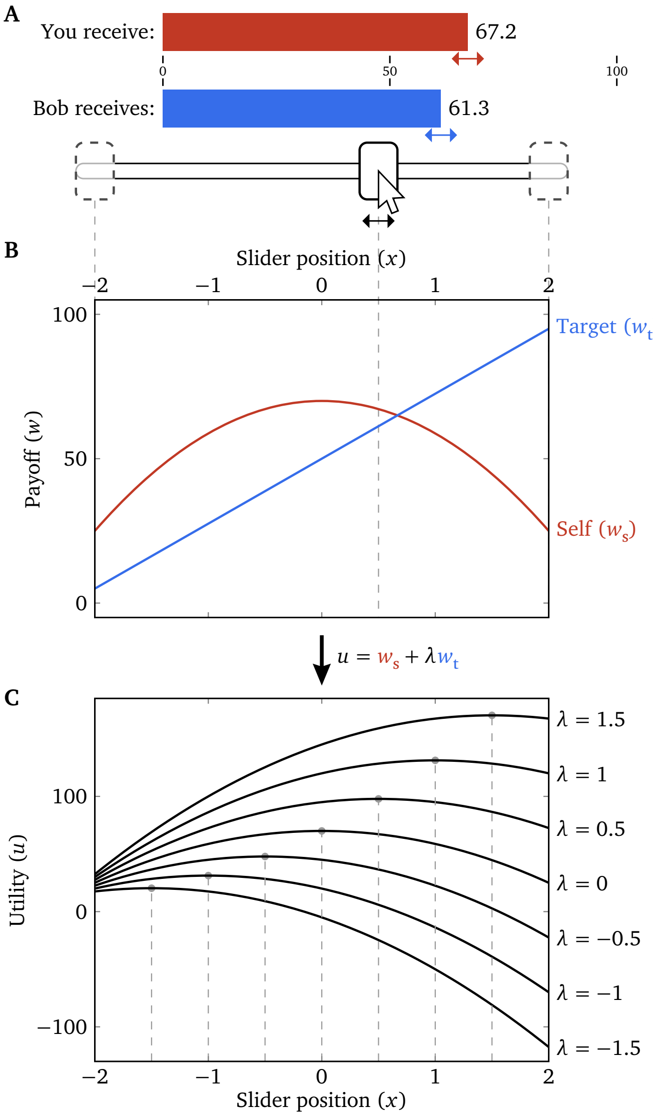


Figure 1.2. The (quadratic) Lambda Slider. (**A**) The interface of the slider. The payoff to oneself (red bar) and payoff to the target (blue bar) change continuously as the participant moves the slider. (**B**) The payoff functions of the quadratic Lambda Slider used in Experiment 1 ($a = 11.25$, $b_s = 70$, $b_t = 50$, $x_{\min} = -2$ and $x_{\max} = 2$). If we plot w_s and w_t against each other, we get a parabolic curve similar to Fig. 1.1D. (**C**) Examples of the participant's utility function for different λs. The slider position that maximizes each utility function is marked, which is an identity function of the participant's λ.

$$x \in [x_{\min}, x_{\max}]\,, \tag{1.4}$$

where $a > 0$ is an arbitrary scale parameter that expands or shrinks the range of payoff values, b_{s} and b_{t} are arbitrary shift parameters that can offset the participant and target's payoff ranges from one another, and $x_{\min}$ and $x_{\max}$ are boundaries of the slider (Fig. 1.2B). When we apply the utility definition of Eq. (1.1), we get

$$\begin{aligned} u(x) &= w_{\mathrm{s}}(x) + \lambda w_{\mathrm{t}}(x) \\ &= -ax^2 + b_{\mathrm{s}} + \lambda(2ax + b_{\mathrm{t}}) \\ &= -a(x-\lambda)^2 + a\lambda^2 + b_{\mathrm{s}} + b_{\mathrm{t}}\lambda\,, \end{aligned}$$

which is a concave parabola with a peak at $x = \lambda$ (Fig. 1.2C), so it satisfies the criterion

$$x^* = \underset{x \in [x_{\min}, x_{\max}]}{\arg\max}\; u(x) = \lambda\,, \quad \forall \lambda \in \left(x_{\min}, x_{\max}\right). \tag{1.5}$$

In other words, the participant will choose the slider position that is equal to her λ (as long as it falls between $x_{\min}$ and $x_{\max}$) in order to maximize her utility, and this single response on the Lambda Slider gives a measurement of the participant's λ with theoretically infinite sensitivity, though of course there will be some limits imposed by the implementation of the task.

We call a Lambda Slider with payoff functions given by Eqs. (1.2) and (1.3) the quadratic Lambda Slider. When plotted on the w_{s}–w_{t} plane, the quadratic Lambda Slider is still a parabola, and $w_{\mathrm{s}} = f(w_{\mathrm{t}})$ is a strictly concave function, similar to Fig. 1.1D. The (quadratic) Lambda Slider shares the logic with mechanism design (Hurwicz & Reiter, 2006), i.e., we design the payoff structure such that the player's rational action directly reveals her hidden preferences (λ in our case).

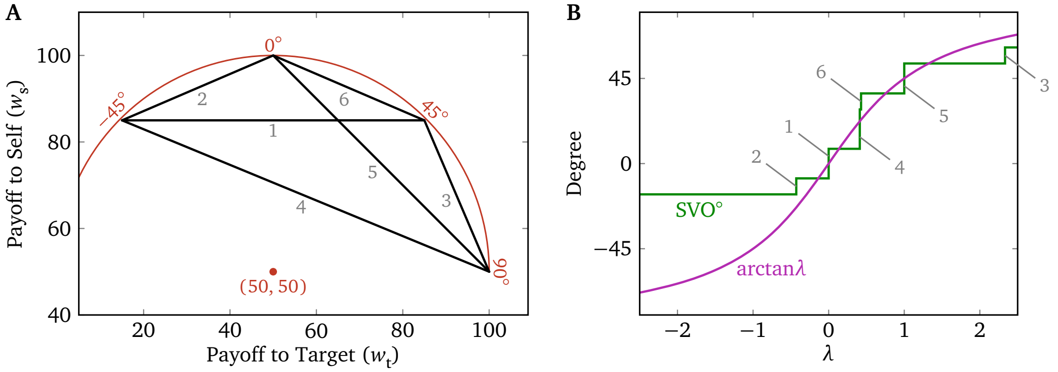


Figure 1.3. The SVO Slider Measure (Murphy et al., 2011). (**A**) The payoff functions of the 6 primary items of the measure (black lines). Each segment represents the linear relationship between w_s and w_t on one of the items, and they are labeled in the same order as in Murphy et al. (2011). The red arc and point $(50, 50)$ provide an intuitive explanation (but not formal justification) for the calculation of SVO°. (**B**) The theoretical step function (green curve) between the output of the measure (SVO°) and λ. The labeled vertical segments correspond to the $\hat{\lambda}$s (the thresholds of λ at which a utility-maximizing participant switches from one end to the other on the sliders) of the 6 items. The theoretical response on the circular Lambda Slider ($\arctan \lambda$; Eq. (1.15)) is also plotted for comparison.

1.1.4 SVO Slider Measure

One apparent difference between the Lambda Slider and the measures based on binary allocation tasks is that the set of possible responses is continuous for the former but discrete for the latter. There is a measure, the SVO Slider Measure, that employs continuous sliders to assess social value orientation, a theoretical construct similar to λ (Murphy & Ackermann, 2014; Murphy et al., 2011). This measure consists of 6 sliders, each involving linear payoff functions for the participant, w_s, and another person, w_t. Each slider connects two points on a circular arc, centered on $(w_s, w_t) = (50, 50)$ (Fig. 1.3A). The points represent the choices most aligned with four categorical social value orientations: competitive, selfish, prosocial, and altruistic. After calculating the average chosen payoffs for self and target across the 6 sliders ($\overline{w_s}$ and $\overline{w_t}$), a summary output is calculated as:

$$\text{SVO}^\circ = \arctan\left(\frac{\overline{w_t} - 50}{\overline{w_s} - 50}\right), \tag{1.6}$$

This "angle" can be interpreted as the angle of the point a participant would choose among payoff values aligned along the arc in Fig. 1.3A, with larger angles corresponding to higher values of λ. (In fact, this arc can be used to create a "circular" Lambda Slider; see Appendix 1.B.)

The SVO Slider Measure is relatively efficient, requiring 6 responses for one measurement, which is fewer than previous measures such as the 9-item Triple-Dominance Measure (Van Lange et al., 1997) and the Ring Measure (Liebrand, 1984; Liebrand & McClintock, 1988). However, the linear nature of the slider payoff functions effectively results in binary allocation tasks, which create the familiar tradeoff between sensitivity and efficiency. For instance, the first slider has payoff functions

$$\begin{aligned} w_{\mathrm{s}}(x) &= 85\,, \\ w_{\mathrm{t}}(x) &= -70x + 85\,, \end{aligned}$$

where $x \in [0, 1]$ is the slider position. Then the utility function is

$$\begin{aligned} u(x) &= w_{\mathrm{s}}(x) + \lambda w_{\mathrm{t}}(x) \\ &= -70\lambda x + 85(1+\lambda)\,. \end{aligned}$$

A utility-maximizing participant would choose $x = 0$ if $\lambda > 0$, choose $x = 1$ if $\lambda < 0$, and be indifferent if $\lambda = 0$. Therefore, this slider is equivalent to a binary allocation task with $\hat{\lambda} = 0$. Similarly, the $\hat{\lambda}$s for the remaining 5 sliders are $-\frac{3}{7}$, $\frac{7}{17}$, $\frac{3}{7}$, 1, and $\frac{7}{3}$. The measure has no way of distinguishing among different λs between two adjacent $\hat{\lambda}$s (e.g., 0 from slider 1 and $\frac{7}{17}$ from slider 4). For any given λ, we can derive the output of the measure, SVO°, from the choices that the participant would make on the 6 sliders, which is plotted in Fig. 1.3B. The relationship between λ and SVO° is not one-to-one, but many-to-many (i.e., different λs between two adjacent $\hat{\lambda}$s lead to the same responses, and for a given λ that is equal to one of the $\hat{\lambda}$s, all positions on one of the sliders are equally preferable). Technically speaking, the SVO Slider Measure has high resolution but low sensitivity. The Lambda Slider has the potential to provide both higher sensitivity and higher efficiency, though when implemented as a single item measure it may have somewhat lower reliability.

1.1.5 Current research

In Experiment 1, we compare the Lambda Slider to the SVO Slider Measure in terms of test–retest reliability and convergent validity, because (a) the SVO Slider Measure performs relatively well in practice and is regarded as the state-of-the-art measure of λ, and (b) it can share an interface with the Lambda Slider (Fig. 1.2A), allowing us to easily mix them in a single experiment. In Experiment 2, we rule out an alternative hypothesis that participants use a heuristic to make decisions on the Lambda Slider. In Experiment 3, we test the external validity of the Lambda Slider using a social decision with real-world consequences, and explore the effects of inequity aversion on measurements of λ.

1.2 Experiment 1

We have formally shown above that the one-shot Lambda Slider has infinite sensitivity. However, how much such theoretical *sensitivity* translates to empirical *accuracy* is limited by the degree to which participants perfectly maximize a utility function in the form of Eq. (1.1).

In Experiment 1, we evaluate the reliability and validity of the quadratic Lambda Slider, and compare it with the SVO Slider Measure[4]. To evaluate the psychometric properties of the Lambda Slider, we need to elicit as wide a range of λs as possible from each participant. It has been shown that a person's λ toward another person decreases as

[4] In all three experiments, we report all measures, manipulations and exclusions.

their social distance increases (Jones & Rachlin, 2006). Therefore, we asked participants to each generate a list of 10 known people (subsequently called "targets") occupying a range of social distances from themselves. We then had participants make hypothetical allocation decisions between themselves and each of those 10 targets. Such a manipulation not only helps elicit a wide range of λs, but also tests the measure's convergent validity with social distance, based on an expected negative correlation between a participant's measured λs toward the targets and her reported social distances from the targets.

1.2.1 Methods

Participants

40 participants were recruited on Prolific and completed the experiment online[5]. The participants were drawn from the "standard sample", were located in the USA, were fluent in English, had an approval rate of at least 95%, and had at least 10 previous submissions on the platform. The participants gave informed consent to participate in the experiment. The experiment was approved by the UCSD institutional review board. Each participant received US$2 for completing the experiment. 30 participants (7 female, 23 male) passed at least 8 out of the 9 attention checks (see below) and only these participants are included in the analyses below.

Design

The experiment is implemented as a web page and can be viewed at https://experiments.evullab.org/qi-games-2/. There are three stages in the experiment: List, Rank, and Slide.

In the List stage, participants are asked to list the first names of 10 people they

[5] The sample sizes in all experiments were determined before any data analysis, although this is not strictly necessary because all data analyses are fully Bayesian. The sample sizes of Experiments 1 and 2 were determined heuristically, while the sample size of Experiment 3 was determined based on a frequentist power analysis as preregistered.

know, 2 in each of 5 categories: family+, friends, neighbors and colleagues, acquaintances, and adversaries. These categories are designed to maximize the range of social distances between a participant and the targets and, presumably, of the participant's λs toward the targets.

In the Rank stage, participants are asked to rank the 10 names they input in the List stage "based on how close you are to them (in terms of relationship, not physical distance)" by dragging the 10 names in a vertical list. The order of the names is initially randomized. The final order of the names is recorded.

In the Slide stage, each participant completes 72 allocation trials using an interface similar to Fig. 1.2A. In each trial, participants drag the horizontal slider, and the payoffs to the participant (w_s) and to the target (w_t), depicted both numerically and as horizontal bars, change continuously according to the underlying payoff functions, which are bounded at 0 and 100 in an arbitrary unit. The bars are labeled "You receive:" and "[Target] receives:", where "[Target]" is replaced by the name of the target in the current trial. Participants are told that the payoffs are hypothetical and are asked to move the slider until the settings look the best to them. The initial position of the slider is randomized in each trial.

In order to evaluate the test–retest reliability of the Lambda Slider and the SVO Slider Measure, we need two measurements for each target for each measure, which amounts to 2 quadratic Lambda Slider trials and 12 SVO Slider Measure trials (twice for each of the 6 primary items) per target. If we measured each participant's λs toward all the targets on both measures, there would be 6 times as many SVO Slider Measure trials as Lambda Slider trials and too many trials in total. Therefore, we measure each participant's λs toward all the 10 targets on the Lambda Slider (20 trials in total), but only targets whose social distance rankings are 1, 4, 7 or 10 on the SVO Slider Measure (48 trials in total).

A participant's response on each quadratic Lambda Slider trial is directly used as the measured λ by virtue of Eq. (1.5). A participant's responses on the 6 different SVO Slider Measure items are aggregated to an SVO° according to Eq. (1.6). The first occurrence of each item is treated as part of the first measurement of SVO°, and the remaining items compose the second measurement.

We also include 4 "catch trials" as attention checks, in which "[Target]" is replaced by "Left" or "Right". Participants are instructed that on these trials they should move the slider to the far left (right) regardless of the payoffs. A participant is considered to pass a catch trial if the slider position she chooses satisfies $x < -1.9$ ($x > 1.9$) when the target is "Left" ("Right"). These 72 trials are randomized in order. Immediately after Trials 2, 6, 14, 30 and 62 (called "memory trials"), participants are asked to type the target's name (or "Left" or "Right") they just saw as attention checks. Participants are considered to pass a memory trial if the name they type is the same as the target's name they just saw, after transforming both names to lowercase and removing whitespaces. The combined "catch" and "memory" trials result in 9 attention checks altogether.

The quadratic Lambda Slider trials have payoff functions defined by Eqs. (1.2)–(1.4) with $a = 11.25$, $b_\mathrm{s} = 70$, $b_\mathrm{t} = 50$, $x_\mathrm{min} = -2$ and $x_\mathrm{max} = 2$, such that $w_\mathrm{s} \in [25, 70]$, $w_\mathrm{t} \in [5, 95]$, and the range of λ that can be accurately measured is $(-2, 2)$ (Fig. 1.2B). We make the range of w_s and w_t narrower than the full range $[0, 100]$ because (a) allowing the payoffs to reach extreme values creates salient points that may bias participants' responses (Thomas & Kyung, 2019), and (b) the welfare participants perceive for themselves and the targets with respect to the raw payoffs is likely to be more nonlinear when the payoffs are close to 0 (Kahneman & Tversky, 1979). The catch trials depict payoff functions in the same manner as the Lambda Slider trials. The SVO Slider Measure trials have the same payoff functions as in Murphy et al. (2011), as shown in Fig. 1.3A.

1.2.2 Results

Test–retest reliability

We evaluate the test–retest reliability of the quadratic Lambda Slider by estimating the correlation between the two measurements of λ of each participant–target combination, and compare it to the correlation between the two measurements of SVO°. We do not expect the correlation of λs to be as high as the correlation of SVO°s because (a) one measurement of SVO° is an aggregation of 6 responses, which almost certainly has less noise than 1 response on the Lambda Slider, and (b) the Lambda Slider has a nonlinear payoff structure, which might be harder to understand than the linear payoff structures of the SVO Slider Measure. However, researchers using the Lambda Slider have the flexibility to select the number of repeated measurements to achieve the desired tradeoff between precision and efficiency[6]. We will first compare the test–retest reliability of the 1-response λ with the 6-response SVO°, and then estimate the reliability of the multiple-response λ.

Figs. 1.4A and B plot the relationship between the two measurements of each participant-target combination. We fit a bivariate normal distribution to the Lambda Slider data and to the SVO Slider Measure data (see Appendix 1.D for details). The two measurements on the quadratic Lambda Slider have a high correlation ($\rho = 0.859$ $(0.823, 0.888)$ [7], Fig. 1.4A), indicating that a single measurement on the Lambda Slider has high test–retest reliability. As predicted, the two measurements of SVO° have an even higher correlation ($\rho = 0.952$ $(0.930, 0.967)$, Fig. 1.4B).

[6] This is different from the tradeoff between sensitivity and efficiency involved in the binary allocation tasks, mentioned in the Introduction. For the binary allocation tasks, the tradeoff arises from a theoretical limitation which even applies to a noiseless decision maker, while the current tradeoff is only due to noise in the decisions.

[7] Here, 0.859 is the posterior median of ρ, and $(0.823, 0.888)$ is the 95% (equal-tailed) credible interval of ρ. The same notation is used for the rest of the paper. Besides, we do not report the probability of direction (p_{d}) or the Bayes factor (relative to a null model) of a parameter if p_{d} calculated using the "direct" method (Makowski et al., 2019) is 100%, in which case the true p_{d} is expected to be at least 99.975% because we take at least 4000 posterior samples in our models.

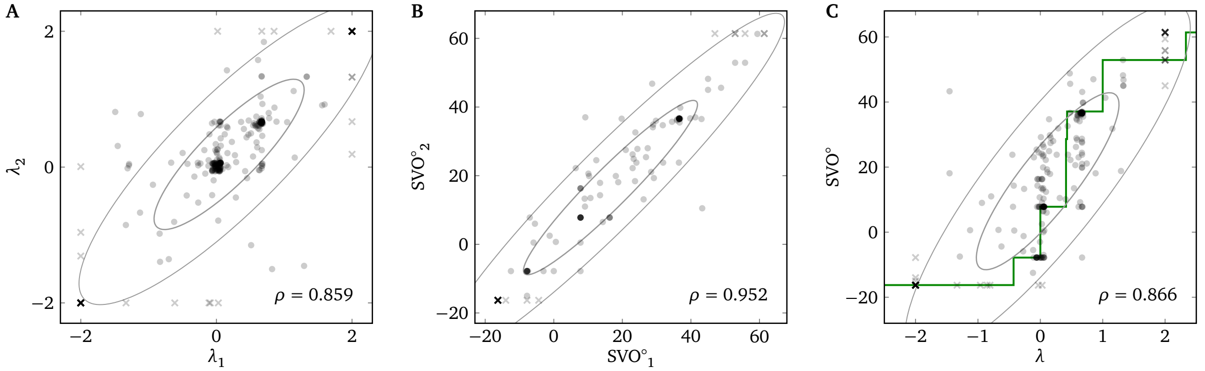


Figure 1.4. Test–retest reliability of (**A**) the quadratic Lambda Slider and (**B**) the SVO Slider Measure, and (**C**) convergent validity between the two measures. In (**A**) and (**B**), each data point represents one participant–target combination. In (**C**), for each participant–target combination, there are two data points representing the first λ paired with the first SVO°, and the second λ paired with the second SVO°. The green line is the theoretical relationship between λ and SVO°, same as Fig. 1.3B. Data points on the boundaries, which are treated as censored data, are represented as crosses (same for all figures below). The ellipses indicate the 1-σ and 2-σ iso-density loci of the fitted bivariate normal distributions with parameters set to their posterior medians.

To assess how many administrations of the Lambda Slider would be required to make the reliability scores of the two measures comparable, we estimate the reliability score of the average of multiple measurements on the Lambda Slider. According to the classical test theory (Lord & Novick, 1968), the test–retest correlation is equal to the reliability score, and

$$\rho = \frac{\sigma_\mathrm{t}^2}{\sigma_\mathrm{t}^2 + \sigma_\mathrm{e}^2},$$

where σ_t^2 is the variance of the true score and σ_e^2 is the variance of the error (of one measurement) on the Lambda Slider. Let ρ' be the reliability score of the average of n measurements on the Lambda Slider. Averaging n measurements shrinks the variance of the error by a factor of n, so we have

$$\rho' = \frac{\sigma_\mathrm{t}^2}{\sigma_\mathrm{t}^2 + \frac{\sigma_\mathrm{e}^2}{n}},$$

and therefore

$$\frac{\rho'}{1-\rho'} = n\,\frac{\rho}{1-\rho}\,.$$

The same result can be obtained by first assuming a multivariate normal distribution over $2n$ variables with the same mean and standard deviation and a fixed pairwise correlation ρ, and then deriving the correlation between the mean of the first n variables and the mean of the other variables.

For a baseline of $\rho = 0.858\ (0.823, 0.888)$, if we increase the number of measurements to $n = 3$, we have $\rho' = 0.948\ (0.933, 0.960)$, which indicates that the reliability of the average of 3 measurements on the Lambda Slider is expected to be comparable to the reliability of the SVO Slider Measure, which requires 6 measurements.

Convergent validity: Lambda Slider vs. SVO Slider Measure

Fig. 1.4C plots the relationship between λ as measured by the quadratic Lambda Slider and by the SVO Slider Measure (SVO°). We fit a 4-variate normal distribution (2 measurements × 2 measures for each participant–target combination) to the data (see Appendix 1.D for details). The two measures are highly correlated ($\rho_{\lambda\nu} = 0.866$ $(0.821, 0.902)$), indicating that the Lambda Slider has high convergent validity with the SVO Slider Measure.

In Fig. 1.4C, there seem to be more responses of SVO° between 7.82° and 36.61° than λ between 0 and 0.667. This does not indicate that the SVO Slider Measure has a higher sensitivity for measuring λ in this range than the Lambda Slider, because (a) it is inconsistent with theoretical predictions, and (b) it can be explained away by assuming that a participant probabilistically chooses between self-gain maximization and perfect inequity aversion for *each decision*, which we do not explicate here but can be investigated by future work.

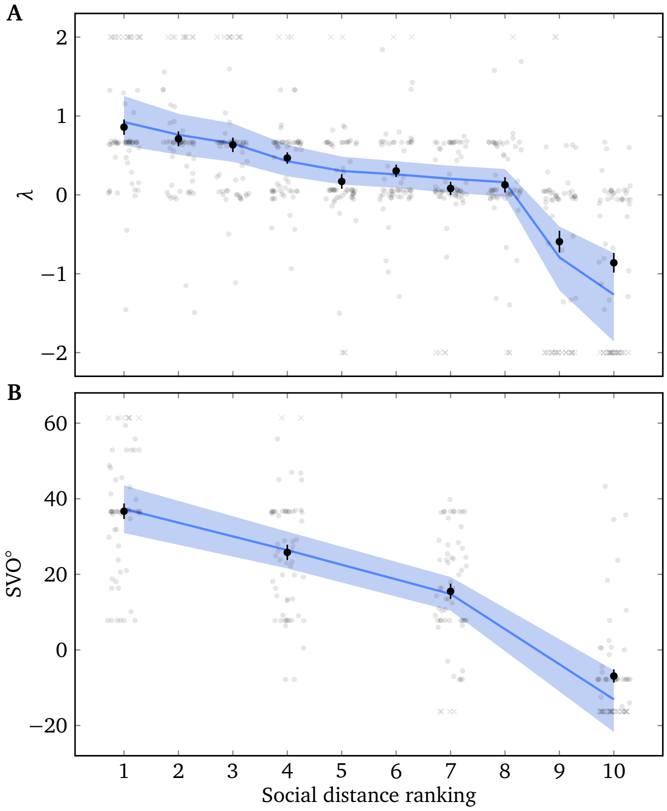


Figure 1.5. Relationship between λ and social distance, for the quadratic Lambda Slider (**A**) and the SVO Slider Measure (**B**). Each raw data point is one of the two measurements of a participant–target combination. Black points and ranges represent the means and standard errors of data in each group. Blue lines and ranges represent the conditional effects (also called marginal effects; Bürkner, 2017) of the social distance ranking as a monotonic predictor, with 95% credible intervals.

Convergent validity: λ vs. social distance

Fig. 1.5 shows participants' measured λs from the Lambda Slider and their SVO° measurements toward targets with different social distances from the participants. We fit a Bayesian mixed effects model to the data with the social distance ranking as a monotonic predictor and λ or SVO° as the dependent variable (see Appendix 1.D for details). As predicted, λ as measured by the quadratic Lambda Slider decreases as the target's social distance ranking increases (mean slope $b_3 = -0.25$ $(-0.34, -0.16)$; this corresponds to how much λ decreases on average as social distance ranking increases by 1). The output of the SVO Slider Measure (SVO°) also decreases as the target's social distance ranking increases (mean slope $b_3 = -16.9$ $(-20.9, -13.0)$; this corresponds to how much SVO° decreases on average as social distance ranking increases by 3).

It is worth noting that participants' λs spanned a wide range (Fig. 1.5A). The *mean* λ toward the socially closest person is 0.86, which means that participants value the target's welfare almost as much as their own. The *mean* λ toward the socially most distant person is -0.86, which means that participants are almost willing to give up $1 to take $1 away from the target. The SVO Slider Measure has very low sensitivity for $\lambda < -0.43$ or $\lambda > 1$ (Fig. 1.3B), and thus cannot measure a large subset of plausible λs accurately.

1.3 Experiment 2

Experiment 1 provided evidence that the Lambda Slider is a valid and reliable measure of λ. However, it is possible that instead of making decisions by incorporating the relevant λs into a utility function like the one in Eq. (1.1) (we call this hypothesis H_λ), participants use the slider position as a qualitative representation of kindness/spitefulness and make decisions based on this representation (we call this hypothesis H_χ). For instance, after getting an intuitive idea of how the two payoffs change as a function of the raw slider position $\chi \in [0, 1]$ [8], a participant might treat $\chi = 0$, 0.25, 0.5, 0.75 and 1 as "very mean", "somewhat mean", "neutral", "somewhat nice", and "very nice", respectively. Then she may choose to be "very nice" to Alice, "somewhat mean" to Bob, etc., and choose slider positions accordingly.

H_λ and H_χ make different predictions when we alter the relationship between λ and the raw slider position χ. For example, suppose that in an initial trial in which $x \in [-2, 2]$, Alice chooses $\chi = 0.75$ on the quadratic Lambda Slider, corresponding to a λ of 1 for that target. If we then have Alice make a decision for the same target on

[8] Note that χ is different from x above. $\chi = 0$ ($\chi = 1$) always corresponds to the left (right) end of the slider.

a different quadratic Lambda Slider with $x \in [-1, 1]$, H_λ predicts that she will choose $\chi = 1$ (corresponding to $\lambda = 1$) but H_χ predicts that she will still choose $\chi = 0.75$ (corresponding to $\lambda = 0.5$).

In general, suppose we have two quadratic Lambda Sliders, Slider A and Slider B. Let $x \in [x_{\text{minA}}, x_{\text{maxA}}]$ on Slider A and $x \in [x_{\text{minB}}, x_{\text{maxB}}]$ on Slider B. Let the raw slider position that the participant chooses be χ_A on Slider A and χ_B on Slider B. For simplicity, suppose neither χ_A nor χ_B is at the boundaries of the slider. Let λ_A (λ_B) be the λ derived from χ_A (χ_B). We have

$$\begin{aligned}\lambda_\text{A} &= (1 - \chi_\text{A}) x_{\text{minA}} + \chi_\text{A} x_{\text{maxA}},\\ \lambda_\text{B} &= (1 - \chi_\text{B}) x_{\text{minB}} + \chi_\text{B} x_{\text{maxB}}.\end{aligned}$$

Given H_λ, since λ on the two sliders should be the same, we have $\lambda_\text{A} = \lambda_\text{B}$, and therefore

$$\chi_\text{B} = \frac{x_{\text{maxB}} - x_{\text{minB}}}{x_{\text{maxA}} - x_{\text{minA}}} \chi_\text{A} + \frac{x_{\text{minA}} - x_{\text{minB}}}{x_{\text{maxA}} - x_{\text{minA}}}. \tag{1.7}$$

Given H_χ, we have

$$\chi_\text{B} = \chi_\text{A}. \tag{1.8}$$

To adjudicate between H_λ and H_χ, in Experiment 2, we let participants make decisions for each target on three different quadratic Lambda Sliders with different ranges of x, and see which hypothesis best predicts the responses.

1.3.1 Methods

Participants

20 participants were recruited on Prolific and completed the experiment online. The participant consent, experiment approval, prescreening, payments, and attention check criteria were the same as Experiment 1. 16 participants (4 female, 12 male) passed

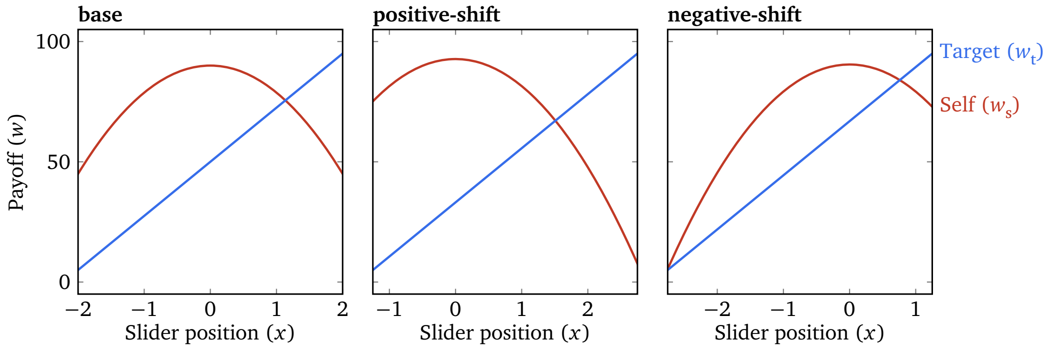


Figure 1.6. Payoff functions of the three quadratic Lambda Sliders in Experiment 2. The ranges on the x axes reflect the ranges of the sliders.

at least 8 out of the 9 attention checks and are included in the analyses below.

Design

Similar to Experiment 1, Experiment 2 is implemented as a web page and can be viewed at https://experiments.evullab.org/qi-games-4/. It also has three stages: List, Rank and Slide, and the List and Rank stages are identical to Experiment 1.

In the Slide stage, there are three quadratic Lambda Sliders: a "base" slider with $x \in [-2, 2]$, same as Experiment 1; a "**pos**itive-shift" slider with $x \in [-1.25, 2.75]$; a "**neg**ative-shift" slider with $x \in [-2.75, 1.25]$ (Fig. 1.6; how we select these ranges and the payoff functions is detailed in Appendix 1.C). For each participant, each target is measured twice on each of the three sliders, with a total of 60 Lambda Slider trials. There are 4 "Left"/"Right" catch trials, similar to Experiment 1, whose payoff functions are the same as the base slider. These 64 trials are randomized in order. The memory trials are at the same locations as in Experiment 1, so there are still 9 attention checks altogether.

We will compare the responses on the three sliders. From Eqs. (1.7) and (1.8), we see that H_λ predicts

$$\chi_{\text{base}} = \chi_{\text{pos}} + 0.1875 = \chi_{\text{neg}} - 0.1875\,, \tag{1.9}$$

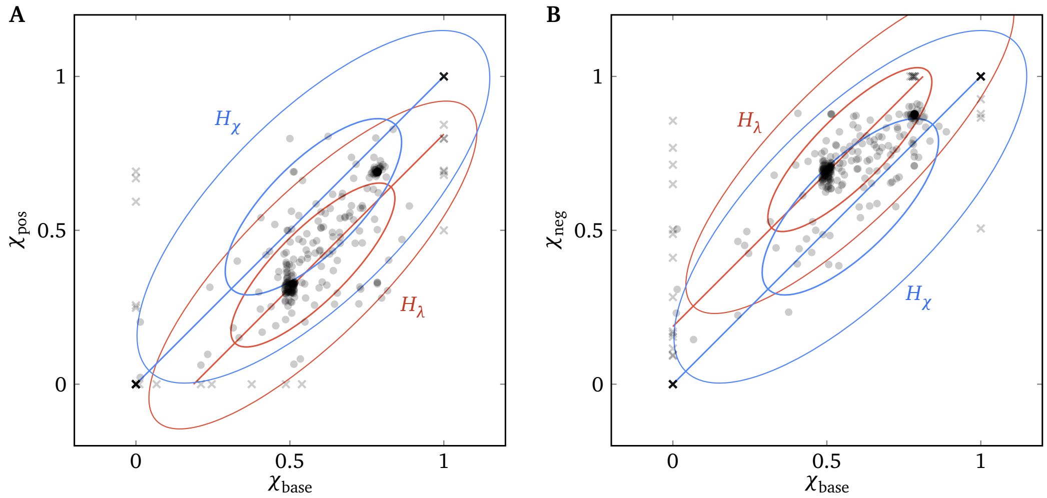


Figure 1.7. Responses in Experiment 2 compared to predictions of H_λ and H_χ. The axes are raw slider positions (χ) for the base, **pos**itive-shift and **neg**ative-shift sliders. For each participant–target combination, there are two raw data points in each panel representing the two measurements on either slider. The diagonal lines indicate the predictions of the two hypotheses without noise. The ellipses indicate the bivariate normal distributions representing the two fitted models (see Appendix 1.D).

while H_χ predicts

$$\chi_{\text{base}} = \chi_{\text{pos}} = \chi_{\text{neg}}\,. \tag{1.10}$$

1.3.2 Results

Fig. 1.7 plots the comparisons of the responses on the base slider versus the positive-shift or negative-shift slider, and compares them to the predictions of H_λ and H_χ. For either hypothesis, we fit a 6-variate normal distribution (2 measurements × 3 sliders for each participant–target combination) to the data, with the constraint that the means satisfy either Eq. (1.9) or Eq. (1.10) depending on the hypothesis (see Appendix 1.D for details). The logarithm of the Bayes factor between H_λ and H_χ is 101.6, indicating decisive evidence in favor of H_λ compared to H_χ. This confirms that participants likely made decisions based on λ and utility maximization rather than based on a qualitative representation of kindness.

As further evidence for the test–retest reliability of the Lambda Slider under different configurations, the within-slider correlations under H_λ are $\rho_{\text{base}} = 0.876$ (0.833, 0.908), $\rho_{\text{pos}} = 0.903$ $(0.870, 0.927)$, and $\rho_{\text{neg}} = 0.906$ $(0.871, 0.931)$. We also examine the relationship between λ and the social distance ranking with a model similar to Experiment 1 (see Appendix 1.D), and the mean slope is $b_3 = -0.22$ $(-0.34, -0.10)$, $p_{\text{d}} = 99.98\%$.

1.4 Experiment 3

So far, the decisions participants made in the experiments were all hypothetical. However, the utility of the Lambda Slider in practice also depends on its external validity (also called predictive validity by some); i.e., whether hypothetical decisions on the Lambda Slider predict real-world altruistic behavior. Despite theoretical concerns about whether decisions with hypothetical payoffs can predict decisions with real payoffs (Kahneman & Tversky, 1979), experiments using matched designs have generally found good alignment between the two settings (Bostyn et al., 2018; FeldmanHall et al., 2012; Johnson & Bickel, 2002; Locey et al., 2011; Wiseman & Levin, 1996). However, most decisions people make in the lab, such as making monetary tradeoffs in an economic game, are so different from real-life decisions that it is unclear whether behavior in these decisions can generalize to real-life situations. Therefore, for the best test of the external validity of a measure, we need to use real-life decisions with real payoffs. Höglinger and Wehrli (2017) examined the external validity of the SVO Slider Measure using a standard dictator game. Using anonymous targets in all measures, they found that the correlation between the SVO Slider Measure (in terms of SVO°) and amount given in the dictator game was 0.42. Likewise, in Experiment 3, we let participants make a real-life decision of how much money to donate, with an underlying structure of a dictator game, and examine its relationship with hypothetical decisions on the Lambda Slider. We also examine the robustness of the Lambda Slider under different configurations and the effects of inequity aversion on the measurements. This experiment was preregistered at https://osf.io/zbw8f.

1.4.1 Inequity aversion

A basic assumption of any study on λ is that a person's utility function is a *linear* combination of w_s and w_t, at least within the range of payoffs in that study. In other words, the only motivations under consideration are the motivations to increase or decrease one's own and the other person's welfare. However, another relevant social motivation is inequity aversion, which is the desire to decrease the absolute difference between w_s and w_t (Fehr & Schmidt, 1999). We can see inequity aversion at play in the previous experiments. In Figs. 1.4A, 1.4C and 1.5A, instead of forming a smooth distribution between $\lambda = 0$ and $\lambda = 2$, many Lambda Slider responses were concentrated at $\lambda = 0.667$, which leads to $w_s = w_t$ given the parameters in Experiment 1. Likewise, in Figs. 1.4B, 1.4C and 1.5B, instead of forming a smooth distribution between SVO° = 7.82° (corresponding to a λ slightly greater than 0) and SVO° = 61.39° (the maximum possible value), many SVO Slider Measure responses were concentrated at SVO° = 36.61°, which is consistent with the responses of a perfectly inequity-averse decision maker.

Formally, we can add an inequity-aversion term to the utility function of Eq. (1.1):

$$u = w_s + \lambda w_t - \kappa |w_s - w_t|, \tag{1.11}$$

where $\kappa \in [0, 1)$ captures the strength of inequity aversion[9].

The measurement of λ may be biased and may lose sensitivity around the equal-payoff point if the participant has a nonzero κ. This problem is shared by all the existing measures of λ, including our Lambda Slider. Trying to counter this problem, Murphy et

[9] It can be shown algebraically that this utility function is equivalent to a utility function with separate advantageous- and disadvantageous-inequality terms but no λ term, as in Fehr and Schmidt (1999). In fact, the utility function in Fehr and Schmidt (1999) can always be rewritten in the form of Eq. (1.11), but not vice versa.

al. (2011) describe a set of secondary linear-payoff sliders that are used to distinguish between inequality aversion and "joint gain maximization". Participants' responses to these secondary items can be used to calculate an "inequality aversion index" ranging from 0 (pure inequality aversion) to 1 (pure joint gain maximization). However, this approach can only be used for participants whose responses on the "primary items" (Fig. 1.3A) are consistent with a "prosocial" orientation (i.e., $\lambda \approx 1$), as the index assesses the degree to which a participant's *responses* are closer to a perfectly consistent decision maker with $\lambda = 1, \kappa = 0$ versus $\lambda = 1, \kappa = 1$.

Using a range of payoff configurations, the Lambda Slider can simultaneously measure λ and κ with no additional restriction on the value of λ. To see how, we can consider the cases where $w_\mathrm{s} \geq w_\mathrm{t}$ and $w_\mathrm{s} < w_\mathrm{t}$ separately and substitute Eqs. (1.2) and (1.3) into Eq. (1.11):

$$\begin{aligned}
u(x) &= \begin{cases} (1-\kappa)w_\mathrm{s} + (\lambda+\kappa)w_\mathrm{t} & w_\mathrm{s} \geq w_\mathrm{t} \\ (1+\kappa)w_\mathrm{s} + (\lambda-\kappa)w_\mathrm{t} & w_\mathrm{s} < w_\mathrm{t} \end{cases} \\
&= \begin{cases} (1-\kappa)\big(-ax^2 + b_\mathrm{s}\big) + (\lambda+\kappa)\big(2ax + b_\mathrm{t}\big) & w_\mathrm{s} \geq w_\mathrm{t} \\ (1+\kappa)\big(-ax^2 + b_\mathrm{s}\big) + (\lambda-\kappa)\big(2ax + b_\mathrm{t}\big) & w_\mathrm{s} < w_\mathrm{t} \end{cases} \\
&= \begin{cases} -a(1-\kappa)\left(x - \dfrac{\lambda+\kappa}{1-\kappa}\right)^2 + \text{const} & w_\mathrm{s} \geq w_\mathrm{t} \\ -a(1+\kappa)\left(x - \dfrac{\lambda-\kappa}{1+\kappa}\right)^2 + \text{const} & w_\mathrm{s} < w_\mathrm{t} \end{cases}.
\end{aligned} \tag{1.12}$$

For any $\lambda \in \mathbb{R}$ and $\kappa \in [0, 1)$, we can make the difference between the shift parameters of the payoff functions, $b_\mathrm{s} - b_\mathrm{t}$, positive enough such that $w_\mathrm{s} > w_\mathrm{t}$ when $x = \frac{\lambda+\kappa}{1-\kappa}$ and thus $x^* = \frac{\lambda+\kappa}{1-\kappa}$; we can also make $b_\mathrm{s} - b_\mathrm{t}$ negative enough such that $w_\mathrm{s} < w_\mathrm{t}$ when $x = \frac{\lambda-\kappa}{1+\kappa}$ and thus $x^* = \frac{\lambda-\kappa}{1+\kappa}$ (proof omitted). Assuming that the participant maximizes their utility perfectly, these two different values of x^* allow us to solve for λ and κ independently. In Experiment 3 we define a likelihood function based on Eq. (1.12) and perform Bayesian inference on λ and κ. Estimating κ also allows us to examine *its* external validity, similar to the external validity of λ, by looking at the relationship between the estimated κ and participants' decisions in the dictator game with real payoffs.

1.4.2 Methods

Participants

90 participants were recruited on Prolific and completed the experiment online. The participant consent, experiment approval, prescreening, payments, and attention check criteria were the same as Experiment 1. 76 participants (39 female, 36 male, 1 unknown) passed at least 4 out of the 5 attention checks and only these participants are included in the analyses below.

Design

Similar to Experiments 1 and 2, Experiment 3 is implemented as a web page and can be viewed at https://experiments.evullab.org/qi-games-7/. It has three stages: List, Slide, and Bonus & Donation.

The List stage is the same as Experiment 1, except that participants list only one target in each of the five categories. We use the categories as proxies for the social distance rankings and do not ask the participants to rank the targets. After participants list these targets, we introduce an additional target described as a victim in the wildfires of Maui, Hawaii in 2023. The name of the target was extracted from a non-paywalled news article on the wildfires, and we provide a link to the article as well as a description of the victim's circumstances.

In the Slide stage, there are three quadratic Lambda Sliders: a "balanced" slider, where w_s can be either greater than, less than, or equal to w_t depending on the slider position; a "self-more" slider, where $w_\text{s} > w_\text{t}$ holds regardless of the slider position (within the allowed range); and a "target-more" slider, where the inverse holds (Fig. 1.8; see

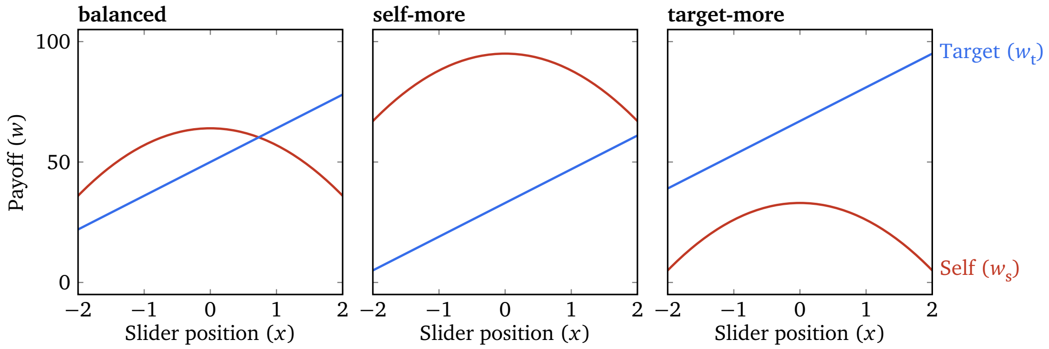


Figure 1.8. Payoff functions of the three quadratic Lambda Sliders in Experiment 3. The ranges on the x axes reflect the ranges of the sliders.

Appendix 1.C for the exact parameters). The range of λ on the sliders is always $[-2, 2]$. These three sliders allow us to estimate the inequity aversion parameter κ as described above.

For each participant, a slider allocation to each of the 6 targets is measured twice on each of the 3 sliders, with a total of 36 Lambda Slider trials, which are randomized in order. A "Left" catch trial and a "Right" catch trial (as in Experiment 1) are added, which become Trials 5 and 20, respectively. Trials 2, 11 and 32 are memory trials, so there are 5 attention checks altogether.

In the Bonus & Donation stage, participants are asked to use a slider to split US$2 between a monetary bonus to themselves and a donation to the Maui Strong Fund, a fund created by the Hawaii Community Foundation to support recovery from the Maui wildfires. Participants essentially play a dictator game between themselves and the fund. The slider has a precision of $0.01. Participants are assured that there is no deception involved and that we will actually donate the amount they specify to the Maui Strong Fund. We also tell participants that after we have collected all the data, we will send them a spreadsheet documenting the donation from each participant and a receipt of the total donation. We tell them that in the spreadsheet the participants will only be identified by the last 5 characters of their Prolific IDs, to prevent them from taking into account others' perception of them.

After a participant completed the experiment, we sent them the monetary bonus they specified in the Bonus & Donation stage through Prolific. Donations from the participants totaled \$65.01, and we donated this amount to the Maui Strong Fund and sent the participants a message through Prolific with links to the spreadsheet and receipt as we promised them.

1.4.3 Results

Robustness

We first fit a 6-variate normal distribution (2 measurements $\times$ 3 sliders for each participant–target combination) to the data from the Slide stage to examine the within-slider and between-slider correlations (see Appendix 1.D for details). The within-slider correlations are ("b" for balanced, "s" for self-more, "t" for target-more) $\rho_{\mathrm{b}} = 0.853$ $(0.825, 0.879)$, $\rho_{\mathrm{s}} = 0.889$ $(0.863, 0.911)$, and $\rho_{\mathrm{t}} = 0.868$ $(0.841, 0.890)$, confirming that the Lambda Slider has high test–retest reliability for a variety of configurations, even though its scale is smaller in this experiment than previous ones ($a = 7$ vs. $a = 11.25$; Figs. 1.2B, 1.6 and 1.8). The Bayes factor between the full model and an alternative model where $\rho_{\mathrm{b}} = \rho_{\mathrm{s}} = \rho_{\mathrm{t}}$ is roughly 1.6, indicating inclusive evidence about whether the test–retest reliabilities of the three sliders are meaningfully different. The between-slider correlations are $\rho_{\mathrm{bs}} = 0.802$ $(0.768, 0.831)$, $\rho_{\mathrm{bt}} = 0.784$ $(0.751, 0.814)$, and $\rho_{\mathrm{st}} = 0.681$ $(0.629, 0.728)$, indicating that measurements of λ are relatively robust to different shift parameters b_{s} and b_{t}.

We examined the relationship between λ and the social distance ranking (excluding the Maui wildfire victim) with the same model as in Experiment 2 (see Appendix 1.D), and the mean slope is $b_3 = -0.60$ $(-0.73, -0.47)$ [10]. Given the larger sample size com-

[10] Since there are only 5 targets here, the value of this slope is roughly comparable to the previous experiments after being divided by 2.

pared to Experiments 1 and 2, we also conducted an exploratory analysis of the effect of sex on λ and the interaction between sex and social distance ranking, and found no evidence toward the existence or nonexistence of these two effects, meaning that the sample size is still not large enough to reach a conclusion (see Appendix 1.D).

External validity

To examine the relationship between measurements on the Lambda Slider and real-world altruistic behavior, we fit a 3-variate normal distribution to the participants' measured λs toward the Maui wildfires victim and their actual donations to the Maui Strong Fund (see Appendix 1.D). We fit the model separately for the three different sliders. For the balanced slider, the correlation between λ and the donation is $\rho_{xd} = 0.448\ (0.231, 0.618)$, and the Bayes factor between the full model and a null model where $\rho_{xd} = 0$ is BF = 640, indicating extreme evidence that the measured λ and the donation are positively correlated and that the Lambda Slider has good external validity. This correlation is close to the correlation of 0.42 between the SVO Slider Measure (in terms of SVO°) and a standard dictator game (Höglinger & Wehrli, 2017), despite the Lambda Slider only depending on 1 response instead of 6. For the self-more slider, $\rho_{xd} = 0.453\ (0.245, 0.623)$, BF = 491. For the target-more slider, $\rho_{xd} = 0.349\ (0.139, 0.529)$, $p_\mathrm{d} = 99.95\%$, BF = 33.3.

Inequity aversion

If participants are inequity-averse, i.e., they have a non-zero κ, the slider position they choose on average would be highest on the self-more slider, lowest on the target-more slider, and in-between on the balanced slider. In the fitted 6-variate normal distribution described above, we have $\mu_\mathrm{s} = 1.04\ (0.85, 1.23)$, $\mu_\mathrm{t} = 0.08\ (-0.02, 0.18)$, and $\mu_\mathrm{b} = 0.22\ (0.10, 0.33)$, suggesting that participants are indeed inequity-averse to some extent.

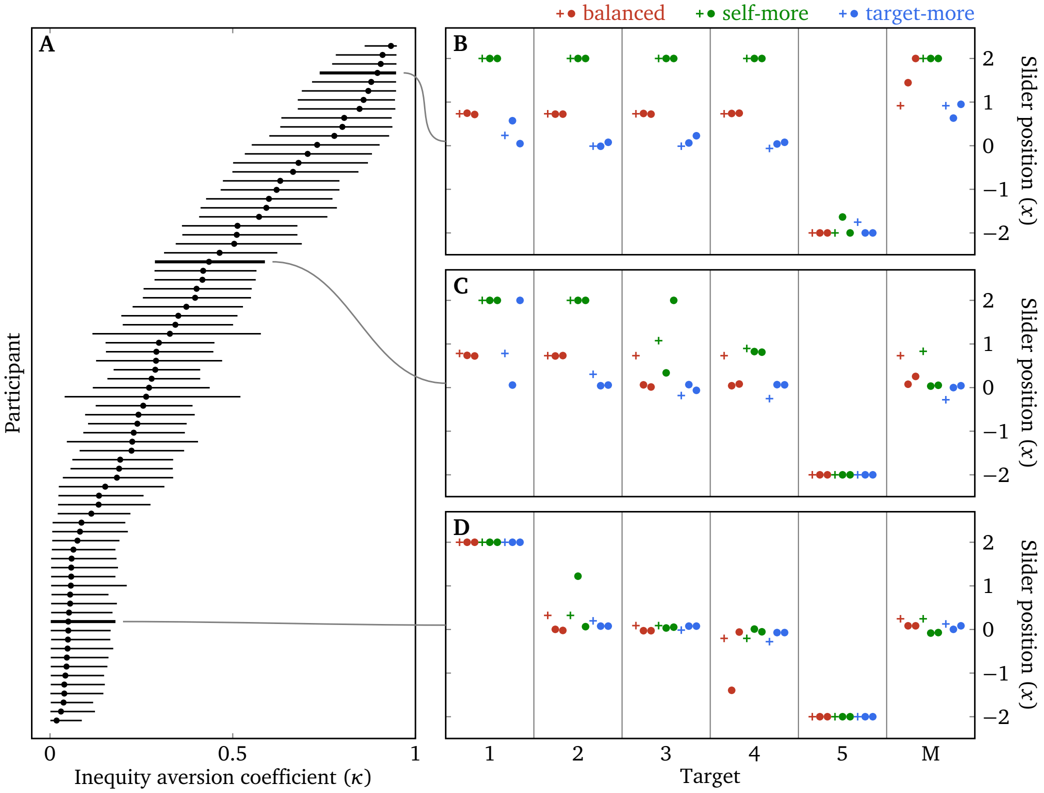


Figure 1.9. Inequity aversion in Experiment 3. (**A**) Posterior distributions of κ for each participant sorted by posterior median. The dots indicate the posterior medians and the lines indicate the 95% credible intervals. Three participants are highlighted, whose raw responses are plotted in (**B**)–(**D**). Targets 1–5 are the targets listed by the participants in the List stage, in increasing order of social distance. The target "M" is the Maui wildfires victim. For each participant–target–slider combination, the cross represents the predicted utility-maximizing response given the posterior medians of λ and κ, while the two dots are the actual responses.

We fit a hierarchical model to jointly estimate λ and κ for each participant–target combination. We assume that each participant has a fixed κ, but their λ varies across targets. We restrict the range of κ to $[0, 0.95]$ because the model becomes unstable when κ gets too close to 1 (see Appendix 1.D for details and other assumptions).

Fig. 1.9A plots the estimates of κ for each participant, which span a wide range. Figs. 1.9B–D plot raw responses of three participants with high, medium and low estimates of κ. We see that the higher κ is, the more slider positions are influenced by the relative offsets of the sliders.

To examine the external validity of κ, we look at the relationship between a participant's estimated κ and how far the participant's donation d is from the equal-payoff point: $|d-1|$. The two variables are negatively correlated ($\rho = -0.343$ $(-0.527, -0.129)$, $p_{\mathrm{d}} = 99.92\%$), indicating that participants with a higher κ are more likely to choose equal payoffs between themselves and another person in real-world decisions.

These data suggest that there is considerable variation among participants in terms of the degree of inequity aversion and, although a single response on the Lambda Slider is highly correlated with a participant's true λ, it may be biased toward the equal-payoff point, especially for participants with high degrees of inequity aversion. In many research programs such biases do not affect the validity of the conclusions, but if and when such biases are a concern, we recommend that researchers jointly estimate λ and κ using multiple Lambda Sliders. We also recommend fitting a complete model like we did for the benefits of having uncertainty estimates and easy integration of prior and global information. But a quick point estimate of λ and κ is also possible by having one measurement x_1 where $w_{\mathrm{s}} > w_{\mathrm{t}}$ and another measurement x_2 where $w_{\mathrm{s}} < w_{\mathrm{t}}$ on two sliders with different relative offsets, and then solving

$$\begin{aligned} x_1 &= \frac{\lambda + \kappa}{1 - \kappa}, \\ x_2 &= \frac{\lambda - \kappa}{1 + \kappa} \end{aligned}$$

for λ and κ by virtue of Eq. (1.12):

$$\begin{aligned} \lambda &= \frac{x_1 + x_2}{2 + x_1 - x_2}, \\ \kappa &= \frac{x_1 - x_2}{2 + x_1 - x_2}. \end{aligned}$$

There is a solution for $\kappa \in [0, 1)$ as long as $x_1 \geq x_2$. In case $x_1 < x_2$, we can assume $\kappa = 0$ and use the average of x_1 and x_2 as a point estimate of λ.

Now we can have a refined understanding of the tradeoff between accuracy and efficiency discussed in the Introduction. Accuracy entails both unbiasedness and reliability. There is a straightforward tradeoff between reliability and efficiency for any measure; the more administrations of a measure are averaged to get a single measurement, the more reliable the measurement will be. On the other hand, biases are trickier to deal with, and none of the correlation metrics we reported in the experiments really deals with biases. In the context of measuring λ, biases are prominently introduced in two ways: (a) through discreteness in the underlying measure, such as the measures based on binary allocation tasks; and (b) through the failure of accounting for inequity aversion. The Lambda Slider, unlike most other measures of λ, is free of the first kind of biases. The second kind of biases can be mitigated by administering multiple Lambda Sliders with different relative offsets of the payoff functions for each participant–target combination and jointly estimating λ and κ, assuming that they are stable across the multiple measurements. Of course, one has to sacrifice some efficiency for this joint estimation. In general, the more prior information one has about λ and/or κ, the less efficiency one has to sacrifice to achieve the same level of accuracy.

1.5 Discussion

We have developed the Lambda Slider, an accurate and efficient measure of λ that is theoretically rigorous. We have shown that the Lambda Slider has high reliability, convergent validity, and external validity for real-world decisions. We have also demonstrated how multiple Lambda Sliders can be used to correct the biases in the measurements of λ caused by inequity aversion.

The Lambda Slider can be straightforwardly implemented using any dynamic graphical user interface. To make it easier for other researchers to use the Lambda Slider, we have created a standalone version of the quadratic Lambda Slider with the same payoff functions as in Experiment 1 (https://experiments.evullab.org/lambda-slider/). It can be directly embedded into web-based survey platforms such as Qualtrics; instructions can be found at https://github.com/jameswhqi/wtr-slider-data/blob/main/README.md.

Although the Lambda Slider is efficient and has good psychometric properties, it may not be the best measure to use under some conditions. The nonlinear payoff structures may be difficult for people to quickly familiarize themselves with. It is also inapplicable to projects relying on paper-based measures. Under these circumstances, it may be preferable to use another measure such as the SVO Slider Measure (Murphy et al., 2011) or the Welfare Trade-Off Task (Delton et al., 2023; Kirkpatrick et al., 2015). The SVO Slider Measure may also better align with personality scale measures designed to assess the same four social strategies that serve as endpoints for the SVO items.

One potential future direction is to use the Lambda Slider to study social perception. People not only make social decisions based on their λs toward other people, but can represent, infer, and predict others' λs toward themselves or someone else and react accordingly e.g., Ackermann et al., 2016; Delton and Robertson, 2012; Krasnow et al., 2016; Lim, 2012; Qi and Vul, 2022; Quillien et al., 2023; Sell et al., 2017. Because of the drawbacks of the existing measures of λ based on binary allocation tasks, the processes of (a) conveying another person's λ to the participant, and (b) measuring the participant's *prediction* of another person's λ, have had a relatively low ceiling on the product of accuracy and efficiency, limiting the study of the dynamics of such inference and prediction over time or space. The Lambda Slider can potentially be used to make these processes more accurate and/or efficient. Using the Lambda Slider to measure participants' predictions of another person's λ seems straightforward—Alice could imagine that she adopts Bob's λ and makes decisions on the Lambda Slider in the same way she makes her own decisions. Participants should also be able to infer others' λs from observations of Lambda Slider choices, so long as the observing participants have a good understanding of the underlying payoff functions. This understanding could potentially be achieved by allowing the participant to manipulate the slider, or by depicting the relationship between the payoff functions using a 2D curve, as in Fig. 1.1D. The validity and reliability of the (1D or 2D) Lambda Slider for either of these purposes need to be established by further research.

Data availability

All data and analysis code in the experiments can be found at https://github.com/jameswhqi/wtr-slider-data, with instructions for reproducing the results.

Acknowledgements

Chapter 1, in full, has been submitted for publication of the material as it may appear in “Qi, W., Vul, E., & Powell, L. J. (under review). An accurate and efficient measure of welfare tradeoff ratios. *PLOS One*.” The dissertation author was the primary investigator and author of this paper.

Appendices

1.A Formal derivation of the Lambda Slider

The general form

Suppose we have a curve on the w_s–w_t space defined by $w_\text{s} = f(w_\text{t})$, $w_\text{t} \in [w_\text{min}, w_\text{max}]$, and f is everywhere differentiable on $w_\text{t} \in (w_\text{min}, w_\text{max})$ and strictly concave (or, equivalently, f' is strictly decreasing), from which we can deduce that f' is continuous and invertible. We can rewrite the utility (Eq. (1.1)) as a function of w_t:

$$\begin{aligned} u(w_\text{t}) &= w_\text{s} + \lambda w_\text{t} \\ &= f(w_\text{t}) + \lambda w_\text{t}, \quad w_\text{t} \in [w_\text{min}, w_\text{max}]. \end{aligned}$$

Then we have

$$u'(w_{\mathrm{t}}) = f'(w_{\mathrm{t}}) + \lambda\,, \quad w_{\mathrm{t}} \in (w_{\min}, w_{\max})\,.$$

Since $f'(w_{\mathrm{t}})$ is continuous and strictly decreasing, we have

$$u'(w_{\mathrm{t}}) \begin{cases} > 0 & w_{\mathrm{t}} \in \big(w_{\min}, f'^{-1}(-\lambda)\big) \\ = 0 & w_{\mathrm{t}} = f'^{-1}(-\lambda) \\ < 0 & w_{\mathrm{t}} \in \big(f'^{-1}(-\lambda), w_{\max}\big) \end{cases},$$

as long as $w_{\min} < f'^{-1}(-\lambda) < w_{\max}$, or, equivalently, $-f'(w_{\min}) < \lambda < -f'(w_{\max})$. Therefore,

$$w_{\mathrm{t}}^{*} = \operatorname*{arg\,max}_{w_{\mathrm{t}} \in [w_{\min}, w_{\max}]} u(w_{\mathrm{t}}) = f'^{-1}(-\lambda)\,, \quad \forall \lambda \in \big(-f'(w_{\min}), -f'(w_{\max})\big)\,.$$

In other words, there is a one-to-one correspondence between $\lambda \in \big(-f'(w_{\min}), -f'(w_{\max})\big)$ and points on the curve that a utility-maximizing participant will choose.

To derive a slider and two payoff functions from this curve, we can parameterize the curve as

$$\begin{aligned} w_{\mathrm{t}} &= g(x)\,, \\ w_{\mathrm{s}} &= f\big(g(x)\big)\,, \\ x &\in [x_{\min}, x_{\max}]\,, \end{aligned}$$

where x is the slider position, $x_{\min}$ and $x_{\max}$ are the boundaries of the slider, and g is a continuous and strictly monotonic (and thus invertible) function. The slider (with the two payoff functions) derived in such a way is called a Lambda Slider. If g is strictly increasing, we have $x_{\min} = g^{-1}(w_{\min})$ and $x_{\max} = g^{-1}(w_{\max})$, and the relationship is reversed if g is strictly decreasing. Then the slider position that the participant (with

$\lambda \in \big(-f'(w_{\text{min}}), -f'(w_{\text{max}})\big)$ will choose is

$$\begin{aligned} x^* &= g^{-1}(w_{\text{t}}^*) \\ &= g^{-1}\big(f'^{-1}(-\lambda)\big). \end{aligned}$$

Let $h(\lambda) = g^{-1}\big(f'^{-1}(-\lambda)\big)$. Since both g^{-1} and f'^{-1} are continuous and strictly monotonic functions, h is also a continuous and strictly monotonic function, so there is a one-to-one correspondence between x^* and λ.

Quadratic Lambda Slider

If we select g such that $g(x) = f'^{-1}(-x)$, we have

$$\begin{aligned} x^* &= -f'\big(f'^{-1}(-\lambda)\big) \\ &= -(-\lambda) \\ &= \lambda, \end{aligned}$$

in which case h is the identity function.

What are the simplest f and g such that h is the identity function? Can f, g, or $f \circ g$ (the payoff function for “self”) be linear? Since f is strictly concave, it cannot be linear. In order to measure both positive and negative λs, $f'(w_{\text{min}})$ and $f'(w_{\text{max}})$ need to have different signs, which means f cannot be monotonic. Since g is monotonic, $f \circ g$ cannot be monotonic, and thus cannot be linear. Therefore, only g can be linear.

Let $w_{\text{t}} = g(x) = Ax + B$, $A > 0$. Given $g(x) = f'^{-1}(-x)$, we have

$$\begin{aligned} & Ax + B = f'^{-1}(-x) \\ \Rightarrow \quad & f'(Ax + B) = -x \\ \Rightarrow \quad & f'(w_{\text{t}}) = -\frac{w_{\text{t}} - B}{A} \end{aligned}$$

$$\begin{aligned}\Rightarrow \qquad f(w_{\mathrm{t}}) &= \int -\frac{w_{\mathrm{t}} - B}{A}\,\mathrm{d}w_{\mathrm{t}} \\ &= -\frac{1}{2A}(w_{\mathrm{t}} - B)^2 + C\,,\end{aligned}$$

where C is an arbitrary constant, and

$$\begin{aligned}w_{\mathrm{s}} &= f\big(g(x)\big) \\ &= -\frac{1}{2A}\big((Ax + B) - B\big)^2 + C \\ &= -\frac{A}{2}x^2 + C\,.\end{aligned}$$

Letting $A = 2a$, $B = b_{\mathrm{t}}$ and $C = b_{\mathrm{s}}$, we get the same payoff functions as Eqs. (1.2) and (1.3).

1.B Circle Test and circular Lambda Slider

Sonnemans et al. (2006) introduced a “Circle Test” to measure participants’ social value orientations (or λs). Participants are presented with a w_{s}–w_{t} plane and are asked to choose a point on a circle defined by $w_{\mathrm{s}}^2 + w_{\mathrm{t}}^2 = 1000$. The Circle Test has the same underlying logic as the Lambda Slider, and, theoretically, there is also a one-to-one correspondence between the participant’s potential λs and points on the (right half of the) circle.

To see this, we can create a Lambda Slider that is almost the same as the Circle Test, by selecting f that defines a half circle on the w_{s}–w_{t} plane and verifying that f satisfies the requirements of a Lambda Slider. Instead of writing out f directly, it is easier to parameterize the curve with θ:

$$\begin{aligned}w_{\mathrm{s}} &= a\cos\theta + b_{\mathrm{s}}\,, \\ w_{\mathrm{t}} &= a\sin\theta + b_{\mathrm{t}}\,,\end{aligned}$$

$$\theta \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right],$$

where $a > 0$, b_s and b_t are arbitrary scale and shift parameters. Then we have

$$\begin{aligned} f'(w_\mathrm{t}) &= \frac{\mathrm{d}w_\mathrm{s}}{\mathrm{d}w_\mathrm{t}} \\ &= \frac{\mathrm{d}w_\mathrm{s}}{\mathrm{d}\theta} \Big/ \frac{\mathrm{d}w_\mathrm{t}}{\mathrm{d}\theta} \\ &= \frac{-\sin\theta}{\cos\theta} \\ &= -\tan\theta, \end{aligned}$$

which confirms that f is everywhere differentiable on $(-a+b_\mathrm{t}, a+b_\mathrm{t})$ and strictly concave. Since the relationship between w_t and θ is bijective, we can define a unique θ^* according to

$$w_\mathrm{t}^* = a\sin\theta^* + b_\mathrm{t}$$

and have

$$\begin{aligned} & & w_\mathrm{t}^* &= f'^{-1}(-\lambda) \\ &\Rightarrow & f'(w_\mathrm{t}^*) &= -\lambda \\ &\Rightarrow & -\tan\theta^* &= -\lambda \\ &\Rightarrow & \theta^* &= \arctan\lambda\,. \end{aligned}$$

Based on this curve, we can define a Lambda Slider by letting $x = \theta$:

$$w_\mathrm{s}(x) = a\cos x + b_\mathrm{s}\,, \tag{1.13}$$

$$w_\mathrm{t}(x) = a\sin x + b_\mathrm{t}\,, \tag{1.14}$$

$$x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right].$$

We call such a Lambda Slider the “circular Lambda Slider”. Applying the utility definition of Eq. (1.1), it can be easily verified that

$$x^* = \underset{x \in [-\frac{\pi}{2}, \frac{\pi}{2}]}{\arg\max}\, u(x) = \arctan \lambda\,, \quad \forall \lambda \in \mathbb{R}. \tag{1.15}$$

We see that the circular Lambda Slider can measure an infinite range of λ, while the quadratic Lambda Slider cannot (due to the constraint of an identity function between λ and x).

Although the Circle Test has the same underlying logic as the Lambda Slider, there are three limitations in Sonnemans et al.’s presentation of the Circle Test. First, they developed the Circle Test as an intuitive extension of the Ring Measure (Liebrand, 1984) without linking the measured angle to λ itself. Second, Sonnemans et al. (2006) only used the Circle Test as a tool without testing its psychometric properties. Third, the Circle Test involves negative payoffs because the payoff structure is defined by $w_\text{s}^2 + w_\text{t}^2 = 1000$. This seems to result from a direct influence of the Ring Measure (Liebrand, 1984). It is well known that people interpret gains and losses differently (Kahneman & Tversky, 1979) and mixing positive and negative payoffs might exacerbate the nonlinearity in the relationship between perceived welfare and payoffs, biasing the measurements. In our Lambda Slider, we can select the shift parameters b_s and b_t such that the payoffs are always positive or always negative, and we only used positive payoffs in our experiments.

An important difference between the Circle Test and the circular Lambda Slider is that the payoff structure of the Circle Test is a full circle, while the payoff structure of the circular Lambda Slider is a half circle. This raises two related questions: (a) How can we explain choices (if any) made on the left half of the circle in the Circle Test? (b) What happens when we extend the range of the circular Lambda Slider to $x > \frac{\pi}{2}$ and/or $x < -\frac{\pi}{2}$ while keeping the functional forms of Eqs. (1.13) and (1.14)[11]? If

[11] In this case the slider is no longer a Lambda Slider as defined in Appendix 1.A, because the curve on

we restrict the utility function to be a linear combination of w_s and w_t, the participant must have a zero or negative coefficient on w_s in her utility function in order to choose $x \in \left(-\pi, -\frac{\pi}{2}\right] \cup \left[\frac{\pi}{2}, \pi\right]$. In reality and in experimental data e.g., Sonnemans et al., 2006, it is unlikely for someone to have a zero or negative coefficient on w_s (i.e., all else being equal, the person is indifferent about her own payoff or prefers a lower payoff for herself). Hence in this paper we mostly restrict ourselves to the utility function in the form of Eq. (1.1), which entails that any extension of the circular Lambda Slider beyond a half circle is useless because those points do not correspond to any λ.

However, it has been shown that people can perceive and make predictions based on social value orientations of altruism ($x = \frac{\pi}{2}$), martyrdom ($x = \frac{3\pi}{4}$), masochism ($x = \pi$), sadomasochism ($x = -\frac{3\pi}{4}$) and aggression ($x = -\frac{\pi}{2}$), which involve zero or negative coefficients on w_s, although their ability to understand these motivations is generally worse than motivations with positive coefficients on w_s (Maki et al., 1979). We can capture such "abnormal" motivations with a different parameterization of the utility function, such as

$$u = w_\mathrm{s} \cos\phi + w_\mathrm{t} \sin\phi\,, \tag{1.16}$$

where $\phi \in (-\pi, \pi]$ is a parameter analogous to λ, and this utility function is equivalent to Eq. (1.1) (up to a scaling factor) given $\lambda = \tan\phi$ for $\phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. ϕ is also equivalent to θ_M in Griesinger and Livingston Jr. (1973). Then we can define the "Phi Slider", which is a (potentially) accurate and efficient measure of ϕ, and the "circular Phi Slider", which has the same payoff functions as Eqs. (1.13) and (1.14) but a wider range over x. For the circular Phi Slider, we have

$$x^* = \underset{x \in (-\pi, \pi]}{\arg\max}\, u(x) = \phi\,, \quad \forall \phi \in (-\pi, \pi]\,,$$

the w_s–w_t plane cannot be written in the form of $w_\mathrm{s} = f(w_\mathrm{t})$. We need a more general definition of a "ϕ slider" that can measure ϕ as defined in Eq. (1.16), which we do not elaborate in the current paper.

so there is a one-to-one correspondence (an identity function) between the slider position a participant chooses and her ϕ.

On the other hand, extending the range of the circular Lambda Slider might be useful even if we are committed to the utility function of Eq. (1.1). Participants are often drawn to the boundaries of a finite-length slider, even when those points do not strictly maximize their utility. If the circular Lambda Slider has a range of $x \in \left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$, participants' choices near the boundaries of the slider are likely to be biased toward the boundaries, and many responses would correspond to $\lambda = \pm\infty$. To eliminate the salient points of $x = \pm\frac{\pi}{2}$ on the slider, the experimenter can extend the range of the slider such that the boundaries (e.g., $x = \pm\frac{2\pi}{3}$) are sufficiently discouraged for typical social motivations (i.e., positive coefficient on w_s) and responses near $x = \pm\frac{\pi}{2}$ are minimally biased.

Another difference between the Circle Test and the circular Lambda Slider is that the Circle Test is presented as a circle on the computer screen, and participants are asked to choose a point on the circle, while the Lambda Slider is presented as a linear slider. In general, a Lambda Slider (or a Phi Slider) can be presented either as a linear slider with two bars of varying lengths indicating the payoffs (1D presentation), or as a curve on the w_s–w_t plane (2D presentation). We think that neither of these two presentations is intrinsically better, but for the Phi Slider, the 2D presentation seems more intuitive when the range of ϕ we want to measure is $(-\pi, \pi]$, while the 1D presentation seems more intuitive when the range of ϕ we want to measure is smaller, such as $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$ (in which case we can measure λ instead of ϕ). It is possible that one of these two presentations has better psychometric properties than the other, which future research can investigate.

1.C Payoff functions in Experiments 2 and 3

In Experiment 2, we would like the payoff functions of the base, positive-shift (pos), and negative-shift (neg) sliders (in the form of Eqs. (1.2)–(1.4)) to have the following properties:

1. The range of x on the base slider is $[-2, 2]$ (same as Experiment 1).
2. The range of x on the pos (neg) slider is a constant shift $d > 0$ upward (downward) from the base slider, and the larger d is, the better.
3. The range of w_t is $[5, 95]$ on any slider.
4. The range of w_s is narrower than $[5, 95]$ on any slider.
5. Let χ^*_base, χ^*_pos and χ^*_neg be the raw slider positions corresponding to $w_\mathrm{s} = w_\mathrm{t}$ on the three sliders, respectively (the equal-payoff points). H_λ predicts $\chi_\mathrm{pos} = \chi_\mathrm{base} - d$ and $\chi_\mathrm{neg} = \chi_\mathrm{base} + d$ while H_χ predicts $\chi_\mathrm{pos} = \chi_\mathrm{neg} = \chi_\mathrm{base}$. We would like χ^*_base, χ^*_pos and χ^*_neg to be halfway between the predictions of H_λ and H_χ so that inequity-averse responses do not bias toward one of the hypotheses, and thus $\chi^*_\mathrm{pos} = \chi^*_\mathrm{base} - \frac{d}{2}$ and $\chi^*_\mathrm{neg} = \chi^*_\mathrm{base} + \frac{d}{2}$.

We find that 0.75 is almost the maximum value d can have to satisfy all these constraints, so we set $d = 0.75$ and use the following parameters:

$$\begin{aligned}
\text{base:}\quad & a = 11.25\,, & b_\mathrm{s} &= 90\,, & b_\mathrm{t} &= 50\,,\\
\text{pos:}\quad & a = 11.25\,, & b_\mathrm{s} &= 92.716\,, & b_\mathrm{t} &= 33.125\,,\\
\text{neg:}\quad & a = 11.25\,, & b_\mathrm{s} &= 90.448\,, & b_\mathrm{t} &= 66.875\,.
\end{aligned}$$

Fig. 1.6 confirms that these sliders satisfy constraints 1–4. In Fig. 1.7, a tight cluster of points halfway between the predictions of the two hypotheses corresponds to the inequity-averse responses, which confirms that the sliders satisfy constraint 5.

In Experiment 3, all three sliders have the same range $[-2, 2]$ and scale $a = 7$, but different offsets:

$$\begin{aligned} \text{balanced:} \quad & b_\text{s} = 64\,, \quad b_\text{t} = 50\,, \\ \text{self-more:} \quad & b_\text{s} = 95\,, \quad b_\text{t} = 33\,, \\ \text{target-more:} \quad & b_\text{s} = 33\,, \quad b_\text{t} = 67\,. \end{aligned}$$

1.D Model specifications

Experiment 1

Test–retest reliability

The two measurements for each participant–target combination i form a data vector $\boldsymbol{x}_i$. The data vectors are assumed to be sampled i.i.d. from a bivariate normal distribution where the two variables have the same mean and standard deviation:[12]

$$\begin{aligned} &\boldsymbol{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})\,, \\ &\boldsymbol{x}_i = (x_{i1}, x_{i2})^\top\,, \\ &\boldsymbol{\mu} = (\mu, \mu)^\top\,, \\ &\boldsymbol{\Sigma} = \sigma^2 \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\,, \end{aligned}$$

[12] In this paper, univariate normal distributions, denoted by $N(\mu, \sigma)$, are parameterized by their standard deviations instead of variances; multivariate normal distributions, denoted by $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, are parameterized by their covariance matrices as usual. Log-normal distributions, denoted by Lognormal(μ, σ), are parameterized by their means and standard deviations in the log space.

where ρ is the correlation parameter representing the test–retest reliability. We set the priors to be

$$\begin{aligned}
&\mu \sim N(0, 2), \\
&\sigma \sim \text{Lognormal}\big(\ln(1), 1\big), \\
&\rho \sim \text{Uniform}(-1, 1)
\end{aligned}$$

for the Lambda Slider data, and

$$\begin{aligned}
&\mu \sim N(20, 40), \\
&\sigma \sim \text{Lognormal}\big(\ln(20), 1\big), \\
&\rho \sim \text{Uniform}(-1, 1)
\end{aligned}$$

for the SVO Slider Measure data. Elements of $\boldsymbol{x}_i$ are restricted to a range ($[-2, 2]$ for λ and $[-16.26°, 61.39°]$ for SVO°), so we treat data points that lie on the boundaries as censored data. For example, if $\lambda_{i1} = 2$ and $\lambda_{i2} = 1.8$, λ_{i2} will be used as data in the model, but λ_{i1} will be a parameter with a lower bound of 2, which will be included in the posterior samples along with the other model parameters. We fit the model using RStan (Stan Development Team, 2023, 2024) with the default sampling parameters.

Lambda Slider vs. SVO Slider Measure

For each participant–target combination i, the two measurements on the Lambda Slider (denoted by λ) and the two measurement on the SVO Slider Measure (denoted by ν) form a data vector $\boldsymbol{x}_i$. The data vectors are assumed to be sampled i.i.d. from a 4-variate normal distribution with certain constraints:

$$\boldsymbol{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$

$$
\begin{aligned}
\boldsymbol{x}_i &= (\lambda_{i1}, \lambda_{i2}, \nu_{i1}, \nu_{i2})^\top ,\\
\boldsymbol{\mu} &= (\mu_\lambda, \mu_\lambda, \mu_\nu, \mu_\nu)^\top ,\\
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_\lambda^2 & \sigma_{\lambda\lambda} & \sigma_{\lambda\nu} & \sigma_{\lambda\nu} \\ \sigma_{\lambda\lambda} & \sigma_\lambda^2 & \sigma_{\lambda\nu} & \sigma_{\lambda\nu} \\ \sigma_{\lambda\nu} & \sigma_{\lambda\nu} & \sigma_\nu^2 & \sigma_{\nu\nu} \\ \sigma_{\lambda\nu} & \sigma_{\lambda\nu} & \sigma_{\nu\nu} & \sigma_\nu^2 \end{pmatrix} ,\\
\sigma_{\lambda\lambda} &= \rho_\lambda \sigma_\lambda^2 ,\\
\sigma_{\nu\nu} &= \rho_\nu \sigma_\nu^2 ,\\
\sigma_{\lambda\nu} &= \rho_{\lambda\nu} \sigma_\lambda \sigma_\nu ,
\end{aligned}
$$

where ρ_λ and ρ_ν are within-measure correlations representing the test–retest reliability of either measure, and $\rho_{\lambda\nu}$ is the between-measure correlation representing the convergent validity. Since not all values of ρ_λ, ρ_ν and $\rho_{\lambda\nu}$ in the range $[-1, 1]$ result in a valid covariance matrix $\boldsymbol{\Sigma}$, instead of parameterizing $\boldsymbol{\Sigma}$ directly, we parameterize the Cholesky factor of the correlation matrix[13]:

$$
\begin{aligned}
\boldsymbol{\Sigma} &= \operatorname{diag}(\boldsymbol{\sigma}) \boldsymbol{P} \operatorname{diag}(\boldsymbol{\sigma}) ,\\
\boldsymbol{\sigma} &= (\sigma_\lambda, \sigma_\lambda, \sigma_\nu, \sigma_\nu)^\top ,\\
\boldsymbol{P} &= \boldsymbol{L}\boldsymbol{L}^\top ,\\
\boldsymbol{L} &= \begin{pmatrix} 1 & 0 & 0 & 0 \\ a_1 & A_1 & 0 & 0 \\ a_2 & A_2 a_3 & A_2 A_3 & 0 \\ a_2 & A_2 a_3 & A_2 A_3 a_4 & A_2 A_3 A_4 \end{pmatrix} ,\\
a_j &\in [-1, 1] ,
\end{aligned}
$$

[13] The correlation matrix is written $\boldsymbol{P}$, the Greek capital letter of ρ. Since it looks identical to the Latin letter P, it can be pronounced as either rho or P.

$$
\begin{aligned}
A_j &= \sqrt{1-a_j^2}\,, \\
a_3 &= \frac{(1-a_1)a_2}{A_1A_2}\,,
\end{aligned}
$$

where $a_{\{1,2,4\}}$ are the true parameters for the correlation matrix $\boldsymbol{P}$. Such parameterization guarantees that $\boldsymbol{\Sigma}$ is a valid covariance matrix and that its constraints are satisfied.

We set the priors to be

$$
\begin{aligned}
\mu_\lambda &\sim N(0,2)\,, \\
\mu_\nu &\sim N(20,40)\,, \\
\sigma_\lambda &\sim \text{Lognormal}\big(\ln(1),1\big)\,, \\
\sigma_\nu &\sim \text{Lognormal}\big(\ln(20),1\big)\,, \\
\boldsymbol{P} &\sim \text{LKJ}(1)\,.
\end{aligned}
$$

Data points that lie on the boundaries are treated as censored data. We fit the model using RStan with the default sampling parameters.

The ellipses in Fig. 1.4C correspond to a bivariate normal distribution with mean vector $(\mu_\lambda, \mu_\nu)^\top$ and covariance matrix

$$
\begin{pmatrix}
\sigma_\lambda^2 & \sigma_{\lambda\nu} \\
\sigma_{\lambda\nu} & \sigma_\nu^2
\end{pmatrix}
$$

where the parameters are set to their posterior medians.

λ vs. social distance

Let i index the participants; n be the number of distinct social distance rankings minus 1 (9 for the Lambda Slider and 3 for the SVO Slider Measure); $0 \leq t \leq n$ index the targets sorted by their social distance rankings ($t = 0$ corresponds to the target with the smallest social distance); $r_t \in [0, n]$ be the linear predictor term derived from the monotonic predictor t and the simplex parameters ζ_i, $1 \leq i \leq n$ (Bürkner & Charpentier, 2020); and y_{it} be the dependent variable (λ or SVO°) for participant i and target t. The model is

$$y_{it} \sim N\big(b_1 + b_{2i} + (b_3 + b_{4i})r_t, \sigma_0\big),$$
$$\begin{pmatrix} b_{2i} \\ b_{4i} \end{pmatrix} \sim N\left(\mathbf{0}, \begin{pmatrix} \sigma_2^2 & \rho\sigma_2\sigma_4 \\ \rho\sigma_2\sigma_4 & \sigma_4^2 \end{pmatrix}\right).$$

Values of y_{it} that lie on the boundaries are treated as censored data. The priors for the Lambda Slider data are

$$\begin{aligned} b_1 &\sim N(1, 2), \\ b_3 &\sim N(0, 0.5), \\ \sigma_0 &\sim \text{Lognormal}(0, 1), \\ \sigma_2 &\sim \text{Lognormal}(0, 1), \\ \sigma_4 &\sim \text{Lognormal}(-1, 1), \\ \rho &\sim \text{Uniform}(-1, 1), \\ \boldsymbol{\zeta} &\sim \text{Dirichlet}(\mathbf{1}). \end{aligned}$$

The priors for the SVO Slider Measure data are

$$\begin{aligned} b_1 &\sim N(40, 40), \\ b_3 &\sim N(0, 30), \\ \sigma_0 &\sim \text{Lognormal}(3, 1), \\ \sigma_2 &\sim \text{Lognormal}(3, 1), \end{aligned}$$

$$
\begin{aligned}
\sigma_4 &\sim \text{Lognormal}(2,1)\,,\\
\rho &\sim \text{Uniform}(-1,1)\,,\\
\boldsymbol{\zeta} &\sim \text{Dirichlet}(\mathbf{1})\,.
\end{aligned}
$$

We fit the model using brms (Bürkner, 2017) with the default sampling parameters.

Experiment 2

Adjudicating between the two hypotheses

For each participant–target combination i, the two raw slider positions on each of the three sliders ("b" for base, "p" for positive-shift, "n" for negative-shift) form a data vector $\boldsymbol{x}_i$. The data vectors are assumed to be sampled i.i.d. from a 6-variate normal distribution with certain constraints:

$$
\begin{aligned}
\boldsymbol{x}_i &\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})\,,\\
\boldsymbol{x}_i &= (\chi_{i\mathrm{b}1},\chi_{i\mathrm{b}2},\chi_{i\mathrm{p}1},\chi_{i\mathrm{p}2},\chi_{i\mathrm{n}1},\chi_{i\mathrm{n}2})^\top\,,\\
\boldsymbol{\mu} &= (\mu,\mu,\mu-\mu^\Delta,\mu-\mu^\Delta,\mu+\mu^\Delta,\mu+\mu^\Delta)^\top\,,\\
\boldsymbol{\Sigma} &= \sigma^2\begin{pmatrix}
1 & \rho_\mathrm{b} & \rho_\mathrm{bp} & \rho_\mathrm{bp} & \rho_\mathrm{bn} & \rho_\mathrm{bn}\\
\rho_\mathrm{b} & 1 & \rho_\mathrm{bp} & \rho_\mathrm{bp} & \rho_\mathrm{bn} & \rho_\mathrm{bn}\\
\rho_\mathrm{bp} & \rho_\mathrm{bp} & 1 & \rho_\mathrm{p} & \rho_\mathrm{pn} & \rho_\mathrm{pn}\\
\rho_\mathrm{bp} & \rho_\mathrm{bp} & \rho_\mathrm{p} & 1 & \rho_\mathrm{pn} & \rho_\mathrm{pn}\\
\rho_\mathrm{bn} & \rho_\mathrm{bn} & \rho_\mathrm{pn} & \rho_\mathrm{pn} & 1 & \rho_\mathrm{n}\\
\rho_\mathrm{bn} & \rho_\mathrm{bn} & \rho_\mathrm{pn} & \rho_\mathrm{pn} & \rho_\mathrm{n} & 1
\end{pmatrix},
\end{aligned}
$$

where $\rho_{\{\mathrm{b,p,n}\}}$ are within-slider correlations, and $\rho_{\{\mathrm{bp,bn,pn}\}}$ are between-slider correlations. According to Eqs. (1.9) and (1.10), we have $\mu^\Delta = 0.1875$ for H_λ and $\mu^\Delta = 0$ for H_χ.

Like above, not all values of $\rho_{\{\mathrm{b,p,n,bp,bn,pn}\}}$ in the range $[-1,1]$ result in a valid covariance matrix $\boldsymbol{\Sigma}$, so again we parameterize the Cholesky factor of the correlation matrix:

$$\begin{aligned}
\boldsymbol{\Sigma} &= \sigma^2 \boldsymbol{P} = \sigma^2 \boldsymbol{L}\boldsymbol{L}^\top , \\
\boldsymbol{L} &= \begin{pmatrix}
1 & 0 & 0 & 0 & 0 & 0 \\
a_1 & A_1 & 0 & 0 & 0 & 0 \\
a_2 & A_2 a_4 & A_2 A_4 & 0 & 0 & 0 \\
a_2 & A_2 a_4 & A_2 A_4 a_6 & A_2 A_4 A_6 & 0 & 0 \\
a_3 & A_3 a_5 & A_2 A_5 a_7 & A_2 A_5 A_7 a_8 & A_2 A_5 A_7 A_8 & 0 \\
a_3 & A_3 a_5 & A_2 A_5 a_7 & A_2 A_5 A_7 a_8 & A_2 A_5 A_7 A_8 a_9 & A_2 A_5 A_7 A_8 A_9
\end{pmatrix} , \qquad (1.17) \\
a_j &\in [-1, 1] , \\
A_j &= \sqrt{1 - a_j^2} , \\
a_4 &= \frac{(1 - a_1) a_2}{A_1 A_2} , \\
a_5 &= \frac{(1 - a_1) a_3}{A_1 A_3} , \\
a_8 &= \frac{(1 - a_6) a_7}{A_6 A_7} ,
\end{aligned}$$

where $a_{\{1,2,3,6,7,9\}}$ are the true parameters for the correlation matrix $\boldsymbol{P}$. Such parameterization guarantees that $\boldsymbol{\Sigma}$ is a valid covariance matrix and that its constraints are satisfied.

We set the priors to be

$$\begin{aligned}
\mu &\sim N(0.5, 0.5) , \\
\sigma &\sim \mathrm{Lognormal}\big(\ln(0.25), 1\big) , \\
\boldsymbol{P} &\sim \mathrm{LKJ}(1) .
\end{aligned}$$

Data points that lie on the boundaries are treated as censored data. We fit both models (H_λ and H_χ) using RStan with the default sampling parameters, except that the number of total iterations is set to 5000, half of which are warm-up samples. We estimate the marginal likelihoods of the two models using bridge sampling (Gronau et al., 2020).

In Fig. 1.7, using the base-pos comparison as an example (panel A), the ellipses correspond to a bivariate normal distribution with mean vectors $(\mu, \mu - \mu^{\Delta})^\top$ and covariance matrix

$$\sigma^2 \begin{pmatrix} 1 & \rho_{\text{bp}} \\ \rho_{\text{bp}} & 1 \end{pmatrix}$$

where the parameters are set to their posterior medians.

λ vs. social distance

The model is similar to Experiment 1, but with an extra predictor representing the slider, which we assume only changes the intercept but not the slope on r_t. Let s_p and s_n be dummy variables corresponding to the positive- and negative-shift sliders. The model becomes

$$y_{it} \sim N\big(b_1 + b_{2i} + b_5 s_\text{p} + b_6 s_\text{n} + (b_3 + b_{4i}) r_t, \sigma_0\big),$$

and the priors are the same as the ones for the Lambda Slider data in Experiment 1, with the extra terms

$$b_5 \sim N(0, 1),$$
$$b_6 \sim N(0, 1).$$

Note that the dependent variable is slider position x, whose range depends on the slider, not the *raw* slider position χ, whose range is always $[0, 1]$.

Experiment 3

Robustness

For each participant–target combination i, the two slider positions on each of the three sliders (“b” for balanced, “s” for self-more, “t” for target-more) form a data vector $\boldsymbol{x}_i$. The data vectors are assumed to be sampled i.i.d. from a 6-variate normal distribution with certain constraints:

$$\begin{aligned}
\boldsymbol{x}_i &\sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \\
\boldsymbol{x}_i &= (x_{i\mathrm{b}1}, x_{i\mathrm{b}2}, x_{i\mathrm{s}1}, x_{i\mathrm{s}2}, x_{i\mathrm{t}1}, x_{i\mathrm{t}2})^\top, \\
\boldsymbol{\mu} &= (\mu_\mathrm{b}, \mu_\mathrm{b}, \mu_\mathrm{s}, \mu_\mathrm{s}, \mu_\mathrm{t}, \mu_\mathrm{t})^\top, \\
\boldsymbol{\Sigma} &= \begin{pmatrix}
\sigma_\mathrm{b}^2 & \sigma_\mathrm{bb} & \sigma_\mathrm{bs} & \sigma_\mathrm{bs} & \sigma_\mathrm{bt} & \sigma_\mathrm{bt} \\
\sigma_\mathrm{bb} & \sigma_\mathrm{b}^2 & \sigma_\mathrm{bs} & \sigma_\mathrm{bs} & \sigma_\mathrm{bt} & \sigma_\mathrm{bt} \\
\sigma_\mathrm{bs} & \sigma_\mathrm{bs} & \sigma_\mathrm{s}^2 & \sigma_\mathrm{ss} & \sigma_\mathrm{st} & \sigma_\mathrm{st} \\
\sigma_\mathrm{bs} & \sigma_\mathrm{bs} & \sigma_\mathrm{ss} & \sigma_\mathrm{s}^2 & \sigma_\mathrm{st} & \sigma_\mathrm{st} \\
\sigma_\mathrm{bt} & \sigma_\mathrm{bt} & \sigma_\mathrm{st} & \sigma_\mathrm{st} & \sigma_\mathrm{t}^2 & \sigma_\mathrm{tt} \\
\sigma_\mathrm{bt} & \sigma_\mathrm{bt} & \sigma_\mathrm{st} & \sigma_\mathrm{st} & \sigma_\mathrm{tt} & \sigma_\mathrm{t}^2
\end{pmatrix}, \\
\sigma_{ll} &= \rho_l \sigma_l^2, \quad l \in \{\mathrm{b}, \mathrm{s}, \mathrm{t}\}, \\
\sigma_{l_1 l_2} &= \rho_{l_1 l_2} \sigma_{l_1} \sigma_{l_2}, \quad l_1 l_2 \in \{\mathrm{bs}, \mathrm{bt}, \mathrm{st}\},
\end{aligned}$$

where ρ_l are within-slider correlations, and $\rho_{l_1 l_2}$ are between-slider correlations. The parameterization of $\boldsymbol{\Sigma}$ (in fact, the correlation matrix $\boldsymbol{P}$) is the same as in Experiment 2, which was not described in the preregistration because we were not aware of the issue that not all values of ρ result in a valid covariance matrix.

For the alternative model where $\rho_\mathrm{b} = \rho_\mathrm{s} = \rho_\mathrm{t}$, the Cholesky factor of $\boldsymbol{P}$ has the same format as Eq. (1.17), but with two additional constraints among the variables:

$$a_6 = \frac{a_1 - a_2^2 - A_2^2 a_4^2}{A_2^2 A_4^2},$$
$$a_9 = \frac{a_1 - a_3^2 - A_3^2 a_5^2 - A_3^2 A_5^2 a_7^2 - A_3^2 A_5^2 A_7^2 a_8^2}{A_3^2 A_5^2 A_7^2 A_8^2},$$

so $a_{\{1,2,3,7\}}$ are the true parameters for $\boldsymbol{P}$.

We set the priors to be

$$\mu_l \sim N(0,2),$$
$$\sigma_l \sim \text{Lognormal}(0,1),$$
$$\boldsymbol{P} \sim \text{LKJ}(1).$$

Data points that lie on the boundaries are treated as censored data. We fit the full model using RStan with the default sampling parameters, except that the number of total iterations is set to 3000, half of which are warm-up samples. For estimating the Bayes factor between the two models, we randomly sample 10% of the data (bridge sampling would not reliably converge for more data, likely because there are too many parameters in the model corresponding to the censored data points) and fit the two models with the default sampling parameters, except that the number of total iterations is set to 5000, half of which are warm-up samples. We repeat this process 10 times, producing 10 Bayes factors, and report their median in the main text.

For the relationship between λ and social distance, the model is similar to Experiment 2, but with a different set of sliders and two extra predictors—sex and the interaction between sex and social distance. Let s_{s} and s_{t} be dummy variables corresponding to the self-more and target-more sliders. Let s_{m} be dummy variable corresponding to being male (instead of female). The model becomes

$$y_{it} \sim N\big(b_1 + b_{2i} + b_5 s_\mathrm{s} + b_6 s_\mathrm{t} + b_7 s_\mathrm{m} + b_8 s_\mathrm{m} r_t + (b_3 + b_{4i}) r_t, \sigma_0\big),$$

and the priors are the same in Experiment 2, with the extra terms

$$b_7 \sim N(0, 0.5),$$
$$b_8 \sim N(0, 0.2).$$

The number of total sampling iterations is 5000, half of which are warm-up samples.

Call this full model M_1. There are two alternative models, the first one without the b_8 term (call it M_2), and the second one without both b_7 and b_8 (call it M_3). M_3 is essentially the same model as in Experiment 2. The mean slope b_3 reported in the main text is from M_3 fit to the full dataset. To examine the effect of sex, we fit the three models to a dataset where the only participant whose sex was "prefer not to say" is excluded. There is very weak evidence for the *nonexistence* of an interaction between sex and social distance ($b_8 = 0.05$ $(-0.16, 0.26)$, $\mathrm{BF}_{M_1/M_2} = 0.55$) and no evidence for the existence or nonexistence of an effect of sex ($b_7 = 0.16$ $(-0.16, 0.47)$, $\mathrm{BF}_{M_2/M_3} = 0.91$).

External validity

For each participant i, the two measurements on a particular Lambda Slider toward the Maui wildfires victim $x_{i\{1,2\}} \in [-2, 2]$ and the participant's donation to the Maui Strong Fund $d_i \in [0, 2]$ form a data vector $\boldsymbol{x}_i$. The data vectors are assumed to be sampled i.i.d. from a 3-variate normal distribution with certain constraints:

$$\boldsymbol{x}_i \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma}),$$
$$\boldsymbol{x}_i = (x_{i1}, x_{i2}, d_i)^\top,$$

$$
\begin{aligned}
\boldsymbol{\mu} &= (\mu_x, \mu_x, \mu_d)^\top , \\
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_x^2 & \rho_x \sigma_x^2 & \rho_{xd} \sigma_x \sigma_d \\ \rho_x \sigma_x^2 & \sigma_x^2 & \rho_{xd} \sigma_x \sigma_d \\ \rho_{xd} \sigma_x \sigma_d & \rho_{xd} \sigma_x \sigma_d & \sigma_d^2 \end{pmatrix} .
\end{aligned}
$$

The Cholesky factor of the correlation matrix is parameterized as

$$
\begin{aligned}
\boldsymbol{L} &= \begin{pmatrix} 1 & 0 & 0 \\ a_1 & A_1 & 0 \\ a_2 & A_2 a_3 & A_2 A_3 \end{pmatrix} , \\
a_j &\in [-1, 1] , \\
A_j &= \sqrt{1 - a_j^2} , \\
a_3 &= \frac{(1 - a_1) a_2}{A_1 A_2} ,
\end{aligned}
$$

where the true parameters are $a_{\{1,2\}}$. In the null model where $\rho_{xd} = 0$, there is an additional constraint $a_2 = 0$, and the only true parameter is a_1.

We set the priors to be

$$
\begin{aligned}
\mu_x &\sim N(0, 2) , \\
\mu_d &\sim N(1, 1) , \\
\sigma_x, \sigma_d &\sim \text{Lognormal}(0, 1) , \\
\boldsymbol{P} &\sim \text{LKJ}(1) .
\end{aligned}
$$

Data points that lie on the boundaries are treated as censored data. We fit both the full model and the null model using RStan with the default sampling parameters, except that the number of total iterations is set to 5000, half of which are warm-up samples. We estimate the marginal likelihoods of the two models using bridge sampling.

Inequity aversion

We assume that participants' choices on the sliders are noisy maximization of the utility function:

$$p(x) \propto \exp\big(\beta u(x)\big), \tag{1.18}$$

where $\beta > 0$ is a global softmax parameter (we do not have enough data to fit a β for each participant). Substituting Eq. (1.12) into Eq. (1.18), we see that $p(x)$ has the same form as the probability density function of a normal distribution on each segment of the slider where $w_{\mathrm{s}} - w_{\mathrm{t}}$ has the same sign. In other words, x is distributed according to a truncated normal distribution on each of these segments, with the constraint that $p(x)$ is continuous at the boundaries between segments.

Let $\phi_{\mu,\sigma}$ and $\Phi_{\mu,\sigma}$ be the density and cumulative functions of a normal distribution with mean μ and standard deviation σ. Let

$$\begin{aligned}
\mu_1 &= \frac{\lambda+\kappa}{1-\kappa}, \\
\mu_2 &= \frac{\lambda-\kappa}{1+\kappa}, \\
\sigma_1 &= \frac{1}{2\beta a(1-\kappa)}, \\
\sigma_2 &= \frac{1}{2\beta a(1+\kappa)},
\end{aligned}$$

where $a = 7$ is the scale of the slider. The likelihood functions for the self-more and target-more sliders are

$$\begin{aligned}
p_{\mathrm{s}}(x \mid \lambda, \kappa, \beta) &= \frac{\phi_{\mu_1,\sigma_1}(x)}{\Phi_{\mu_1,\sigma_1}(2) - \Phi_{\mu_1,\sigma_1}(-2)}, \\
p_{\mathrm{t}}(x \mid \lambda, \kappa, \beta) &= \frac{\phi_{\mu_2,\sigma_2}(x)}{\Phi_{\mu_2,\sigma_2}(2) - \Phi_{\mu_2,\sigma_2}(-2)}.
\end{aligned}$$

The likelihood function for the balanced slider is slightly more complex because there are two segments where $w_\mathrm{s} - w_\mathrm{t}$ has different signs:

$$p_\mathrm{b}(x \mid \lambda, \kappa, \beta) = \begin{cases} \dfrac{\phi_{\mu_1,\sigma_1}(x)}{A_1 + kA_2} & x \in [-2, \hat{x}] \\ \dfrac{\phi_{\mu_2,\sigma_2}(x)}{\frac{1}{k}A_1 + A_2} & x \in (\hat{x}, 2] \end{cases},$$

where

$$\begin{aligned} \hat{x} &= \sqrt{3} - 1\,, \\ A_1 &= \Phi_{\mu_1,\sigma_1}(\hat{x}) - \Phi_{\mu_1,\sigma_1}(-2)\,, \\ A_2 &= \Phi_{\mu_2,\sigma_2}(2) - \Phi_{\mu_2,\sigma_2}(\hat{x})\,, \\ k &= \frac{\phi_{\mu_1,\sigma_1}(\hat{x})}{\phi_{\mu_2,\sigma_2}(\hat{x})}\,. \end{aligned}$$

Unless $\lambda = -1$, both $\mu_1 \to \infty$ and $\sigma_1 \to \infty$ when $\kappa \to 1$, which makes the model unstable, so we restrict the range of κ to $[0, 0.95]$.

Let i index the participants; κ_i be a participant's κ parameter; $1 \le t \le 6$ index the targets; λ_{it} be the λ parameter for a participant–target combination; and y_{itl} be the response for participant i, target t, and slider l. The model is

$$\begin{aligned} \lambda_{it} &\sim N\left(\mu_t, 0.5\right) \\ y_{itl} &\sim p_l\left(\cdot \mid \lambda_{it}, \kappa_i, \beta\right)\,, \end{aligned}$$

where μ_t is the mean λ for a target across participants. We put a slightly strong prior on λ_{it} (but not unreasonable given prior data) because for extreme values of λ, the posterior distribution would be strongly degenerate and the sampling algorithm would have difficulty exploring the distribution efficiently. The other priors are

$$\begin{aligned}
\mu_t &\sim N(0,1)\,,\\
\beta &\sim \text{Lognormal}(-1,0.5)\,,\\
\kappa_i &\sim \text{Uniform}(0,0.95)\,.
\end{aligned}$$

The relatively strong prior on β is also set to prevent degeneracy. We fit the model using RStan with the default sampling parameters.

The correlation between κ_i and $|d_i - 1|$ (where d_i is the donation amount) is calculated by fitting a bivariate normal distribution to the data, without treating any data as censored:

$$\begin{aligned}
\boldsymbol{x}_i &\sim N(\boldsymbol{\mu},\boldsymbol{\Sigma})\,,\\
\boldsymbol{x}_i &= \left(\kappa_i, |d_i - 1|\right)^\top,\\
\boldsymbol{\mu} &= (\mu_1,\mu_2)^\top,\\
\boldsymbol{\Sigma} &= \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix},
\end{aligned}$$

and the priors are

$$\begin{aligned}
\mu_1,\mu_2 &\sim N(0.5,0.5)\,,\\
\sigma_1,\sigma_2 &\sim \text{Lognormal}\big(\ln(0.5),1\big)\,,\\
\rho &\sim \text{Uniform}(-1,1)\,.
\end{aligned}$$

We fit the model using RStan with the default sampling parameters.

Chapter 2

The evolution of theory of mind on welfare tradeoff ratios

People seem to attribute beliefs and desires to another person when interacting with them. Such a “theory of mind” capacity is essential for complex and uniquely human behavior such as language, but its evolutionary origin remains elusive. Using the formal tools of evolutionary game theory, we asked what environmental properties are necessary to select for a basic form of theory of mind—the ability to infer the prosociality, quantified by the welfare tradeoff ratio, of another person towards oneself. We found that none of the environments studied in classical or evolutionary game theory give an advantage to this form of theory of mind capacities; theory of mind is advantageous only in a new class of environments with stable opponents and variable payoff structures. In two behavioral experiments ($n = 91$) we verified that people can, and do use theory of mind in such an environment. These results suggest that some features of early humans’ social environment that were previously neglected in evolutionary game theory may be responsible for the evolution of people’s complex social capacities.

2.1 Introduction

Theory of mind (ToM) refers to one's ability to impute different mental states, such as beliefs and desires, to others, and use them to build causal theories that can explain, predict, and influence behavior (Premack & Woodruff, 1978). Although there is no consensus on whether non-human animals have ToM (Call & Tomasello, 2008; Krupenye et al., 2016; Penn & Povinelli, 2007), many researchers believe that ToM is a prerequisite for such uniquely human phenomena as language (Baron-Cohen, 2000), culture (Tomasello et al., 2005), and even consciousness (Baumeister & Masicampo, 2010). Thus a natural question is why humans evolved to have ToM. It has been suggested that ToM evolved to facilitate cheater detection in reciprocal altruism (Trivers, 1971), which increases the long-term fitness of every altruistic individual in a group-living setting (Brüne & Brüne-Cohrs, 2006). However, when Axelrod (1984) famously formalized reciprocal altruism with the iterated Prisoner's Dilemma tournament, the best-performing strategy was tit-for-tat, a simple behavior-level strategy that requires no ToM whatsoever. Since Axelrod, evolutionary game theory has helped us understand (or at least speculate about) the origins of a wide range of social behavior in animals and humans (Nowak, 2006b; Weibull, 1997), but the puzzle of the evolution of ToM remains largely untouched. Can evolutionary game theory explain the origins of human ToM abilities?

Evolutionary game theory explains behavioral tendencies by considering what social environment (modeled with games) creates an evolutionary pressure for that behavior (Maynard Smith & Price, 1973). For instance, in the Prisoner's Dilemma (Fig. 2.1A), although mutual cooperation is the globally optimal outcome, the payoff structure strongly encourages defection (Rapoport et al., 1965), which raises the question of how self-interested organisms can cooperate to achieve the mutually best payoffs. Axelrod (1984) showed that when the game is played *repeatedly* between two players, cooperation can be maintained through direct reciprocity, exemplified by the simple strategy of tit-for-tat

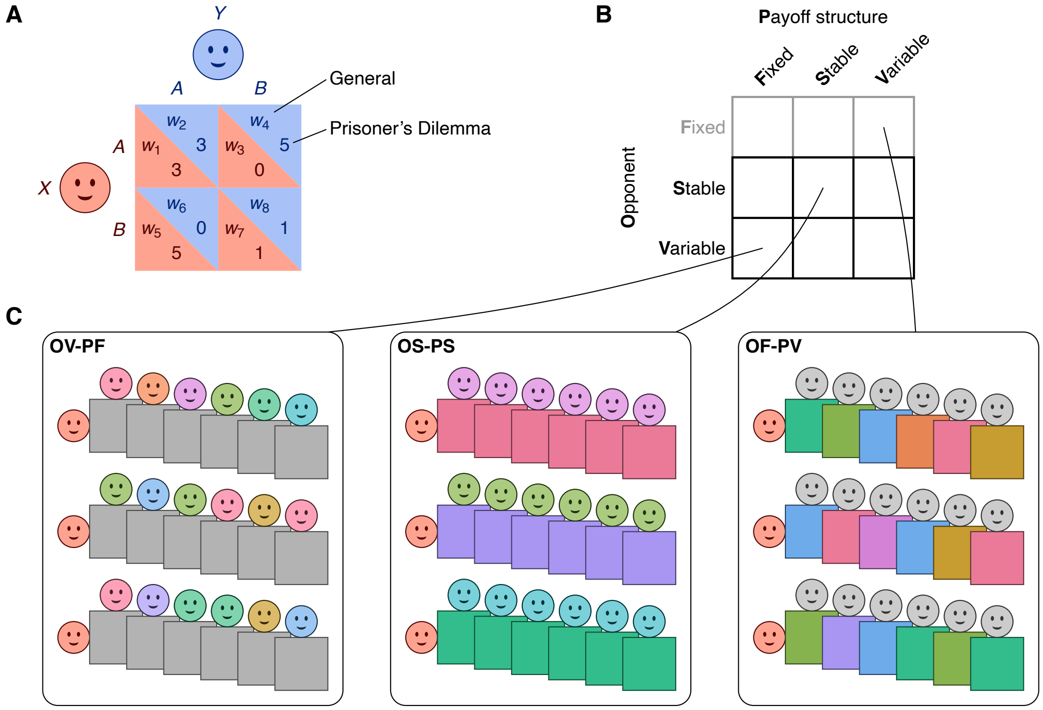


Figure 2.1. The game environments we study. (**A**) The 2×2 normal-form game. Either player chooses an action (A or B), and the payoffs in the resulting cell are given to the two players respectively (e.g., if X chooses A and Y chooses B, X gets w_3 and Y gets w_4). The payoff structure is determined by $w_1, \ldots, w_8$. The canonical payoffs of a Prisoner's Dilemma are also given (A is cooperation and B is defection). (**B**) The 3×3 grid of stability in the opponent and the payoff structure. We do not consider the "fixed opponent" row in the simulation because it makes no sense in an evolutionary setting (see Section 2.2.2). (**C**) Three examples of the environments in the 3×3 grid. Faces with different colors represent different agents, and squares with different colors represent different payoff structures. Each box illustrates the experience of a fixed row player, and every agent in the environment has a similar experience. Note that when the opponent or payoff structure is "fixed", it is the same for all agents in the environment.

(TfT). Here the critical condition for cooperation to arise is the stability of the opponent, which gives TfT players an opportunity to punish uncooperative behavior and thereby avoid exploitation. Using this same logic we ask: what social environment favors ToM abilities?

Most of the social environments that have thus far been studied in evolutionary game theory are insufficient for ToM to arise, in much the same way that one-shot Prisoner's Dilemma is insufficient for cooperation to arise. A central feature of ToM reasoning is that it allows us to respond uniquely to the behavior of distinct individuals by tracking and adapting to their idiosyncratic mental states. In one-shot games (Hauert et al., 2006) an agent never meets the same opponent again, so nothing learned about any individual is useful in the future and we would not expect ToM to be selected in such settings. In fixed repeated games (i.e., repeated games with a fixed payoff structure (Axelrod, 1984) or stochastic games with simple fixed specifications (Hilbe et al., 2018; Shapley, 1953)), simple action-level strategies like TfT are likely to be sufficient, so the additional complexity of ToM would not be advantageous.

We will instead consider a broader range of environments characterized by the stability of opponents and payoff structures, forming a 3×3 grid (Fig. 2.1B). A given feature of the environment may be fixed (invariant over all interactions), stable (repeated over many, but not all, interactions), or variable (unique for each interaction). We consider a grid of possible environments where the rows correspond to fixed/stable/variable opponents (abbreviated as OF/OS/OV) and the columns are fixed/stable/variable payoff structures (abbreviated as PF/PS/PV). This subsumes one-shot games (variable opponents and fixed payoffs: OV-PF), fixed repeated games (stable opponents and fixed payoffs: OS-PF) and variable games (stable opponents and variable payoffs: OS-PV) (Kleiman-Weiner, 2018; Qi & Vul, 2020). We ask if any region of this space creates an evolutionary advantage for ToM abilities.

ToM in human adults is extremely rich and flexible. There are a variety of mental states that ToM can impute to others, such as perceptions, knowledge, beliefs, desires, intentions, and emotions. Although we are expanding the scope of the game environments, they remain highly abstract and restricted compared to the real world environment, such that it is infeasible to study most of the mental states that human ToM can impute. For instance, the literature on ToM has largely focused on beliefs (about some objective state of the world), and false beliefs in particular (Baron-Cohen et al., 1985; Liu et al., 2008; Wimmer & Perner, 1983), but beliefs cannot play a role in a 2×2 normal-form game because the payoff structure is transparent and common knowledge (in a recursive sense).

However, there is one basic form of ToM that is subject to investigation in these slightly enriched environments—thinking about how nice or mean another agent is toward oneself, which can be formalized as a welfare tradeoff ratio (WTR; Tooby and Cosmides, 2008). The WTR reflects how an agent values another agent's welfare compared to her own, and is defined through the utility function

$$u_{\text{self}} = v_{\text{self}} + \lambda \cdot v_{\text{opp}}, \tag{2.1}$$

where λ is the WTR (for the rest of the paper we will mostly use λ to refer to WTR), v_{self} is the payoff for the agent itself and v_{opp} is the payoff for the "opponent", i.e., the other agent (in game theory "opponent" does not necessarily entail competition)[1]. Intuitively, an agent may be selfish ($\lambda = 0$), altruistic ($\lambda > 0$) or spiteful ($\lambda < 0$) toward the opponent. In a game theoretic setting, λ generalizes the notion of cooperation and defection—defined at the action level in classic social dilemma games like the Prisoner's Dilemma—to games with arbitrary payoff structures, and is tied to such concepts as social value orientation (Van Lange et al., 1997) and social discounting (Jones & Rachlin, 2006). There is a rich psychological literature revealing that people model others as utility-maximizers (Jara-Ettinger et al., 2016) and can infer another person's WTR toward themselves or someone else and react accordingly (e.g., through emotions including anger and gratitude and through reciprocity; Delton and Robertson, 2012; Krasnow et al., 2016; Lim, 2012; Monroe, 2020; Sell et al., 2017), highlighting the importance of this basic form of ToM ability.

Here, we focus on the evolution of the ToM ability to infer another agent's λ toward oneself. We conducted an evolutionary simulation based on repeated 2×2 normal-form games (Fig. 2.1A) to identify which environment in the 3×3 grid of stability favors

[1] In this work, following the standard assumptions in the WTR and evolutionary game theory literature, we assume that "payoff" only involves "objective" welfare, but not "subjective" welfare or happiness, and it is the "objective" welfare that gets translated to fitness in the evolutionary analysis below.

an agent with ToM abilities, and then tested whether people use ToM in such an environment through two behavioral experiments.

2.2 Simulation

In this section, we first introduce the agents we used in the simulation. Then we describe the tournaments in which the agents play the games against each other and the resulting pairwise mean payoffs (i.e., for each pair of agents, their respective average payoffs per round in the tournaments). Finally, we describe the evolutionary results based on the pairwise mean payoffs.

2.2.1 Agents

For simplicity, in this study we assume that the two λs (one for either direction) between a pair of agents are fixed throughout their lifetime interactions. It is plausible for agent A to adjust its λ towards agent B given the perceived λ of B towards A, which would constitute reciprocity (Fehr & Schmidt, 2006). If the capacity for such reciprocity endows an agent with an evolutionary advantage in an environment, having a ToM would be even more evolutionarily advantageous compared to agents without a ToM because reciprocity at this level of abstraction requires first representing the opponent's λ. In this work we examine the minimal case: in which environments does a ToM capacity to represent, and infer, an opponent's λ toward oneself—without reciprocity—have an evolutionary advantage?

In the simple case of fixed lifetime λs, a rational ToM agent would infer the opponent's λ towards itself, and make choices to optimize its utility in light of beliefs about the opponent's λ. Such an agent has a much more sophisticated decision policy than a simple action-level strategy like TfT. First, it maximizes utility in light of payoffs, rather than adopting a fixed decision rule over options. Second, it assumes a utility-maximizing model of the opponent to predict their choices. Third, it learns from past experience, rather than adopting a fixed policy that is insensitive to past outcomes. Finally, it learns the utility function of a specific opponent that can be used to predict behaviors in novel environments. While our ToM agent embodies the conjunction of these traits, some of these features in isolation may be beneficial. To identify which environment is necessary for the full set of ToM traits to arise, we first introduce several foils for a ToM agent, which also enriches the evolutionary environment, and then describe the ToM agent.

The first foil is a random agent who chooses one of the two actions with equal probability.

The second foil is "tit-for-tat (TfT)", an action-level strategy that has proved to be very effective in iterated Prisoner's Dilemmas (IPD) (Axelrod, 1984). It chooses A (which corresponds to "cooperation" in the Prisoner's Dilemma when the payoff structure is fixed; see Fig. 2.1A) in the first round, and, in all the other rounds, repeats the opponent's action from the previous round.

The third foil is a "reinforcement-learning (RL)" agent that can adapt to an opponent given sufficient opportunities to learn. However, like TfT, it learns policies over *actions*, does not build a flexible model of the opponent, and cannot immediately adapt to new payoff structures. Intuitively, it learns the best action (in terms of the highest total expected utility in the future) given the history of previous rounds. See Appendix 2.A for a detailed description of the RL agent and Section 2.4 for a discussion on the possibility of making it payoff structure–aware.

The fourth foil is a "naïve utility maximizer (NUM)", who assumes that the opponent will choose randomly between the two actions, and makes a choice by comparing the expected utilities of its own two actions in light of its λ towards the opponent (i.e., it deterministically chooses the action with the higher expected utility for itself[2]). This agent can immediately alter behavior in response to new payoff structures, but cannot

[2] We could have used a slightly noisy decision rule (e.g., a softmax function on the utilities), but the results would be analogous, because the ToM agent presumes some uncertainty in the opponent's λ (see below) and makes robust inferences of it.

adapt to opponents.

The fifth foil is a "fixed-belief maximizer (FBM)", who assumes that the opponent is a NUM whose λ towards the agent is sampled from some population distribution (we use a normal distribution). This agent has a richer understanding of the social world, but still does not adapt to each unique opponent. See Appendix 2.A for a detailed description of the FBM.

Finally, a ToM agent is one that adapts to each opponent by learning what they value (as reflected in their λ), and thus gains the ability to predict their behavior in a broad range of circumstances. Similar to the FBM, a ToM agent assumes that the opponent is a NUM whose λ is sampled from a normal distribution, but contrary to the FBM, this distribution is not fixed. Instead, when playing repeated games, a ToM agent does iterated Bayesian inference to learn a distribution over the opponent's λ. This is the simplest agent that can been seen as having a ToM. See Appendix 2.A for a detailed description of the ToM agent.

2.2.2 Tournaments

To find out which environmental structures favor which kinds of agents, we simulated an evolutionary process based on "tournaments" among the different agents for each environment in the 3×3 grid of stability of games and opponents (Figs. 2.1B and 2.1C). In the tournaments, each agent plays repeated games as specified by the environment with every other agent, and we use the mean payoffs per round for each pair of agents as the indicator of their fitness. First, we do not consider the "fixed opponent" row as it makes no ecological sense when simulating population behavior: it describes a world in which all agents in the environment only ever interact with a single central hub agent, but not each other. In the remaining 2×3 grid, when the payoff structure is fixed, we use the canonical payoffs of a Prisoner's Dilemma (Fig. 2.1A) to give TfT an advantage (other payoff structures likely favor other fixed action-level strategies and the results would be analogous). When the payoff structure is stable or variable, for each new payoff structure, we sample the 8 payoff values independently from a uniform distribution.

We include 6 types of agents in the environment: random, tit-for-tat (TfT), reinforcement-learning (RL), naïve utility maximizer (NUM), fixed-belief maximizer (FBM), and theory-of-mind (ToM). RL, NUM, FBM and ToM are called "utility-maximizing agents" because they maximize a utility function in the form of Eq. (2.1). Since λs are driven by social structure, they are not a property of an agent, but a property of the directional relation between two agents, and any one individual will have a distribution of inbound and outbound λs determined by the social structure. Consequently, we are not concerned with the evolutionary selection of λs themselves, but with the evolutionary selection of the ability to infer others' λs toward oneself. So we fix each utility-maximizing agent's distribution of outbound λ values to $\lambda = -1/0/1$ with probabilities 0.25/0.5/0.25, based on the heuristic that an agent is more likely to be selfish than altruistic or spiteful towards another agent, and assume that inbound and outbound λs are independent.

For each of the 6 environments in the 2×3 grid, we first simulate the mean payoffs for each pair of agents, and then use the pairwise mean payoffs to simulate the evolutionary process (described in Section 2.2.3 below). For the utility-maximizing agents (RL, NUM, FBM and ToM), we include 3 agents of each type in the environment, corresponding to $\lambda = -1/0/1$, so that we have 14 agents in total. In this way, we do not need to explicitly randomize each agent's λ. Instead, we get the pairwise payoffs for all the 14 agents and then collapse over the 3 agents of each type according to the probabilities 0.25/0.5/0.25.

In the stable-opponent (OS) environments, the basic unit of the tournament is 100 rounds of games played by two fixed agents, which we call a supergame ("supergame" in game theory is almost a synonym of "repeated game", but here we use it to denote this specific repeated game with 100 rounds). The history-dependent agents (TfT, RL, ToM) are reset at the start of each supergame. For each pair of agents (including two clones of the same agent), the supergame is repeated multiple times until the standard errors of the mean payoffs are small enough.

In the variable-opponent (OV) environments, we include two groups of agents. Agents in one group always play as the row player, and agents in the other group always play as the column player. In each round of the game, one agent from either group is randomly sampled and the two sampled agents play against each other. In either group there are 4 random agents, 4 TfT agents, and 1/2/1 agents of every other type for $\lambda = -1/0/1$ (24 agents in total). In other words, each agent has equal probability of encountering each type of agent as the opponent, and the distribution of λs for the opponents is 0.25/0.5/0.25. Note that in principle the distribution of the opponents should change as evolution progresses, but here we use the simplifying assumption that the distribution is fixed, so that we can use fixed pairwise mean payoffs in the evolutionary simulation. Each supergame includes 100×24 rounds of games, so that on average each agent plays 100 rounds of games. The TfT agents and RL agents are reset at the start of each supergame because they are defined in terms of raw actions and are insensitive to changes in the type of the opponent. However, the ToM agents are reset in every round of the game (i.e., they do not learn at all and are identical to the FBMs), because when the opponent is constantly changing, it is meaningless to try to learn a specific opponent's λ and the ToM agents may as well assume a fixed distribution of λ like the FBMs. In contrast, the RL and TfT agents might adapt in non-trivial ways to some aspects of the payoff structure or opponent distribution despite changing opponents. The supergame is also repeated multiple times until the standard errors of the all the pairwise mean payoffs are small enough.

In the fixed-payoff (PF) environments, the payoff structure is always a Prisoner's Dilemma with canonical payoffs (Fig. 2.1A), and the payoffs are normalized (i.e., rescaled to have a mean of 0 and a standard deviation of 1). In the stable-payoff (PS) environ-

Table 2.1. Number of repetitions in the simulation.

Environment	# of repetitions	Total # of rounds
OS-PF	100	$100 \times 100 \times 105$
OS-PS	10,000	$100 \times 10{,}000 \times 105$
OS-PV	100	$100 \times 100 \times 105$
OV-PF	500	$100 \times 24 \times 500$
OV-PS	50,000	$100 \times 24 \times 50{,}000$
OV-PV	50,000	$100 \times 24 \times 50{,}000$

ments, the payoff structure stays the same within each supergame, and varies across repetitions of supergames. Each new payoff structure is constructed by sampling 8 values independently from $U(0,1)$ and normalizing them. In the variable-payoff (PV) environments, the payoff structure is resampled in each round, and the sampling process is the same as above.

Table 2.1 shows the number of repetitions of supergames and the total number of rounds in the simulation of each environment. The number of repetitions for OS environments is for each pair of agents (there are $\binom{14}{2} + 14 = 105$ unique pairs). The pairwise mean payoffs are shown in Figs. 2.2A, 2.2B, and 2.5. The payoffs on the diagonal of the matrices are the average payoffs for the two clones of the same agent.

2.2.3 Evolution

To find out which type of agent has the most evolutionary advantage in each environment, we first simulate an evolutionary process that starts from equal proportions of the agents in the environment. Then we confirm the stability of the converged states and generalize the results to arbitrary initial conditions by conducting a formal analysis of the pairwise mean payoffs.

Dynamic simulation

The relative evolutionary advantage of the different types of agents can be characterized by how the population proportions of the agents change as evolution progresses.

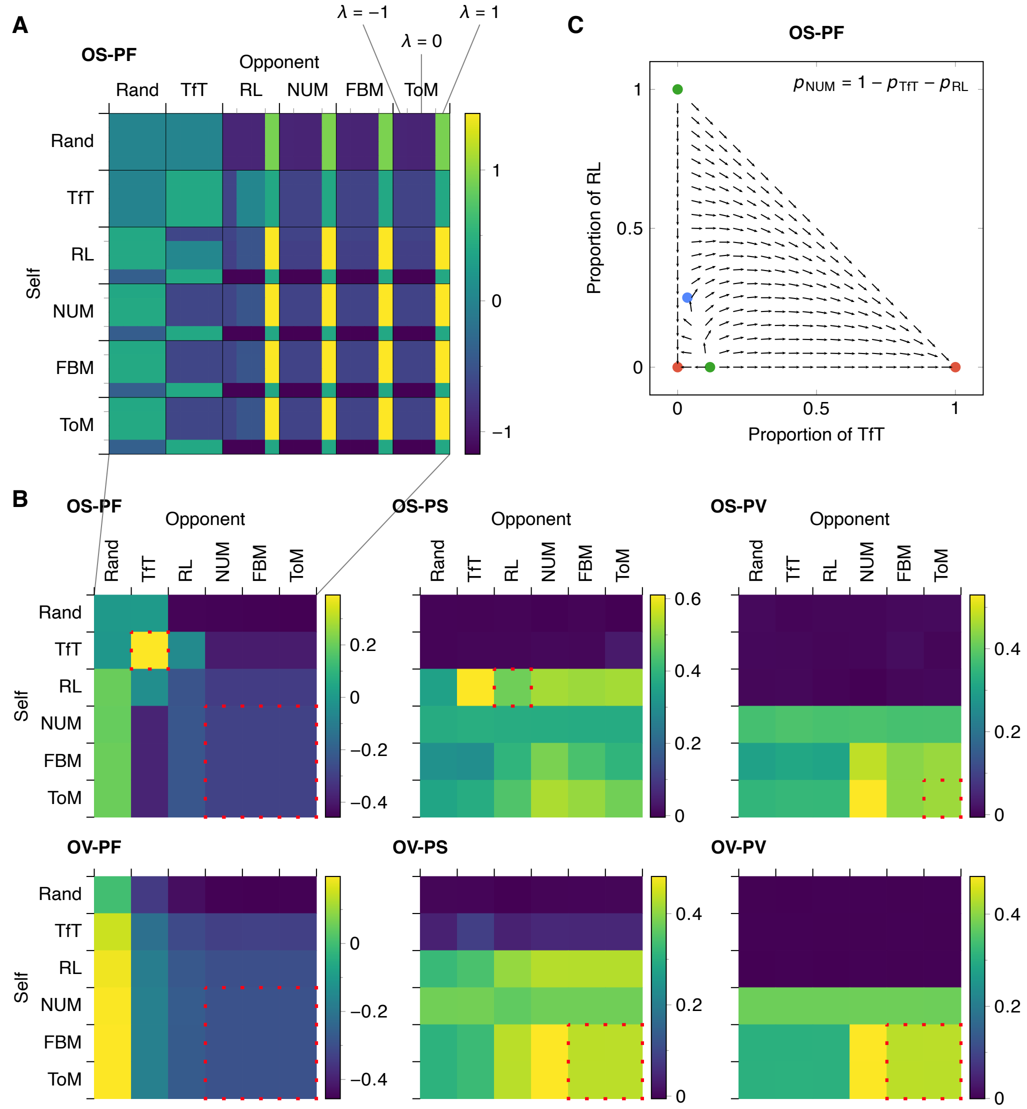


Figure 2.2. (**A**) The raw pairwise mean payoffs for all the agents used in the simulation, including different λs for the same type of agent. Here the environment with stable opponents and a fixed payoff structure (OS-PF) is used as an example (see Fig. 2.5 for the other environments). The matrix contains 6×6 major cells and 14×14 minor cells. The area of each minor cell is proportional to its weight in calculating the mean payoff in the major cell it belongs to (plotted in (**B**)). For each type of the utility-maximizing agents, the λs are $-1/0/1$ both from left to right and from top to bottom. The plotted values are the payoffs for the "Self" agents (i.e., agents on the rows). (**B**) The pairwise mean payoffs averaged over different λs within each type of agent, for all the environments. The red dotted squares are the diagonal cells corresponding to the evolutionarily stable strategies (ESS) in each environment. Some squares are larger than a single cell in the matrix because some strategies (e.g., NUM, FBM and ToM in OS-PF) are equivalent and merged into a single strategy in the ESS analysis (see Section 2.2.3). (**C**) The population flow in OS-PF with only TfT, RL, and NUM (equivalent to FBM and ToM) in the environment. The arrows only represent the flow directions, not the flow speed. The colored dots are the fixed points (red: attractor; green: repeller; blue: saddle point).

To simulate the evolutionary process, we need to first define the evolutionary dynamics, i.e., how the proportions of agents change as a function of their payoffs. The classic population dynamics in evolutionary game theory is the replicator dynamics (Taylor & Jonker, 1978), which captures the intuition that the relative speed the population of an agent grows (or shrinks) is proportional to its average payoff when it interacts with all the agents in the environment with equal probability. Let n be the number of types of agents (6 in our case); v_{ij} be agent i's payoff when playing against agent j, as specified in the matrices of pairwise mean payoffs; x_i be the proportion of agent i in the population; and $\boldsymbol{x} = (x_1, \ldots, x_n)$ be the distribution of agents. Then the replicator dynamics are defined in terms of the following ordinary differential equation:

$$\dot{x}_i = x_i\big(f_i(\boldsymbol{x}) - \phi_i(\boldsymbol{x})\big), \tag{2.2}$$

$$f_i(\boldsymbol{x}) = \sum_{j=1}^{n} x_j v_{ij}, \tag{2.3}$$

$$\phi_i(\boldsymbol{x}) = \sum_{j=1}^{n} x_j f_j(\boldsymbol{x}),$$

where $\dot{x}_i$ is the derivative of x_i with respect to time, $f_i(\boldsymbol{x})$ is the fitness of agent i, and $\phi_i(\boldsymbol{x})$ is the weighted average population fitness. This definition guarantees that $\sum_{i=1}^{n} x_i$ is a constant (e.g., 1). We used the replicator dynamics to simulate the evolutionary process, starting from equal proportions of the 6 types of agents in the environment so that the agents are on an equal footing. We used a discrete-time simulation with a time-step of 0.5 [3], and the results are shown in Fig. 2.3. In Section 2.2.3 below we will generalize the results to arbitrary initial conditions.

First we see that the ToM agent does not have a unique advantage in the variable-opponent (OV) environments and its behavior and performance are identical to the fixed-

[3] Here the time scale follows the definition of Eq. (2.2), but it can be arbitrarily rescaled by multiplying the right-hand side of Eq. (2.2) by a constant or rescaling the payoffs. Such a rescaling does not affect the relative proportions of the agents.

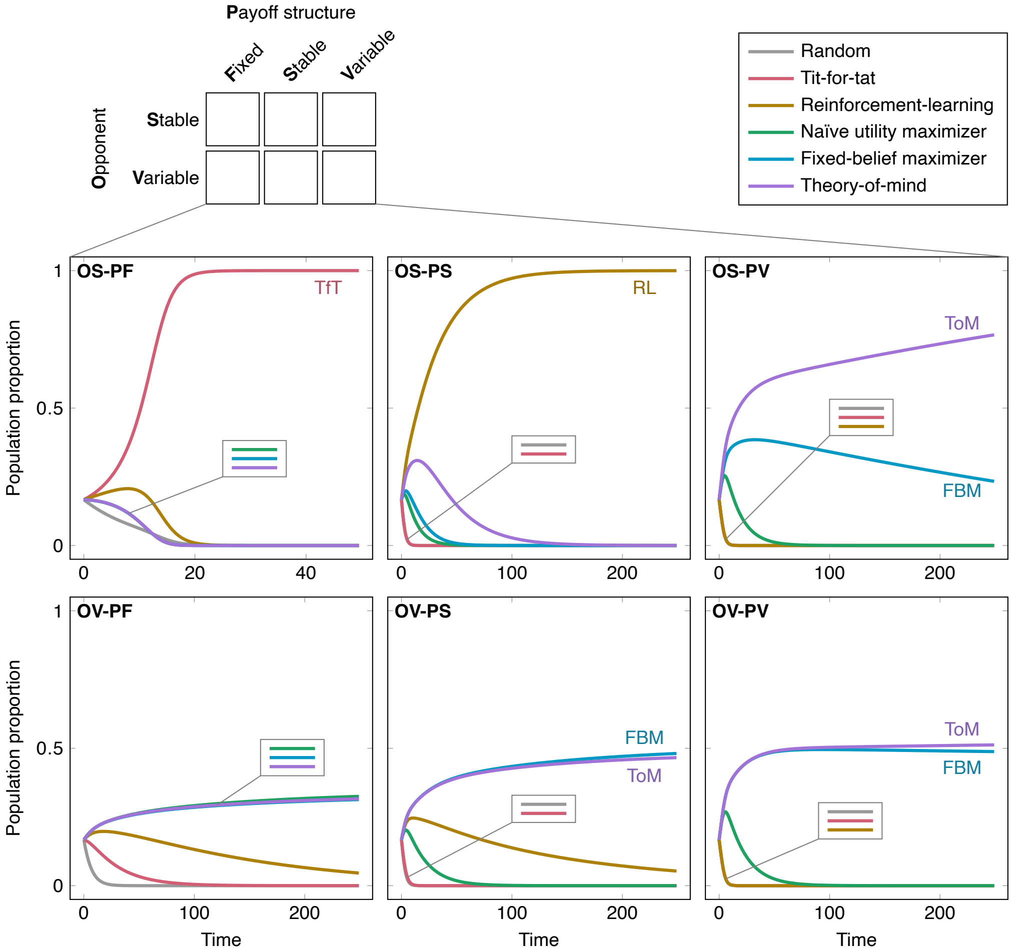


Figure 2.3. Simulation of evolution for the 2×3 grid of stability. The ToM agent boasts a unique advantage in the environment with stable opponents and variable payoff structures (OS-PV). The evolutionary process is based on the pairwise mean payoffs for all the agents in each environment (Fig. 2.2B), starts with equal proportions of all the agents, and follows the replicator dynamics (Taylor & Jonker, 1978). Note that the scale of the x-axis in OS-PF is different from the other environments because OS-PF converges much faster. The small gaps between FBM and ToM in OV-PS and OV-PV are due to Monte Carlo errors in the simulation of the tournaments that provide payoffs for the evolutionary simulation; FBM and ToM are *by definition* equivalent in the OV environments.

belief maximizer. This is because when the opponent changes at every round of the game, nothing learned about the opponent is useful in the future, and the agent might as well assume a fixed distribution of the opponent's λ. However, more interesting patterns arise when opponents are stable.

When opponents are stable and the payoff structure is fixed to a Prisoner's Dilemma (OS-PF), the TfT agent easily dominates the environment, which replicates previous results in IPD (Axelrod, 1984). All the utility-maximizing agents perform worse because TfT is specifically adapted to IPD (other fixed repeated games likely have other simple action-level strategies adapted to them), and hardly any deviation from it can improve performance.

When opponents and the payoff structure are stable (OS-PS), the TfT agent is no different from the random agent because the payoff structures vary and option A is no more likely to mean "cooperation" than option B. In contrast, the utility-maximizing agents who can adapt to different payoff structures perform better than TfT. Nonetheless, the RL agent performs best because when both the opponent and the payoff structure are stable, it has the opportunity to learn the best action-level strategy for specific opponent–payoff structure combinations. Although the ToM agent can learn about the opponent and adapt to different payoff structures, its hypothesis space about the opponent is quite restricted (only naïve utility maximizers) and cannot predict the behavior of other types of agents very accurately, while the RL agent uses a model-free algorithm and does not have such a restriction, and can exploit whatever regularities arise in the specific combination of opponent and payoff structure.

The ToM agent performs best only when opponents are stable, but the payoff structure is variable (OS-PV). With stable opponents but variable payoffs, the RL agent cannot learn the best action for each new payoff structure and its performance deteriorates to that of the random agent. Utility-maximizing agents without a ToM can adapt to new payoff structures instantly, but they fail to adjust to their opponents. The ToM agent performs best in a stable-opponent, variable-payoff environment because it can adapt to opponents with different degrees of prosociality and can use a decision model for the opponent to generalize and predict their choices across different payoff structures.

Stability

The previous simulation only captured the evolutionary process from a single initial condition for a finite amount of time. One question that remains is whether the converged state of the population is stable: will the converged population distribution resist invasion from otherwise eliminated strategies? This question can be answered through a formal analysis of the pairwise mean payoffs (Fig. 2.2B).

The classic stability concept in evolutionary game theory is the evolutionarily stable strategy (ESS)—a strategy that is impermeable by other strategies when adopted by the full population in an environment (Maynard Smith & Price, 1973). We can apply this concept to our evolutionary environment by abstracting one level to consider each type of agent as a pure strategy in a standard normal-form game. At this level of abstraction, the mean outcome for agent i playing agent j (across many simulated game rounds, v_{ij}, Fig. 2.2B) corresponds to the payoff for the row player in a symmetric normal-form game describing the selection of agent type. In some environments there are equivalent strategies that exhibit exactly the same behavior (e.g., in OS-PF, NUM, FBM and ToM are equivalent). Since playing any mixture of these strategies is equivalent, we collapse them into a single pure strategy. The formal definition of a (pure-strategy) ESS is a pure strategy i such that $\forall j \neq i$, either (a) $v_{ii} > v_{ji}$, or (b) $v_{ii} = v_{ji}$ and $v_{ij} > v_{jj}$ (Maynard Smith & Price, 1973).

First we use this formal criterion to identify the pure ESSes in the payoff matrices in each of our environments. Fig. 2.2B shows via red dotted squares the only pure ESSes in each environment. For all of these pure ESSes, all paired strategies satisfy condition (a) above, meaning that when playing against a population comprised entirely of the pure ESS i, all other strategies have lower mean payoffs than i playing against itself. We also evaluated whether any of our environments have any further mixed-strategy (rather than pure-strategy) ESSes, and found none (see Appendix 2.B). Thus the pure strategy ESSes shown in Fig. 2.2B are the only ESSes in these environments.

We see that there is only one ESS in each environment except for OS-PF, and they coincide with the best-performing agents in Fig. 2.3: payoff-aware agents (NUM, FBM and ToM) in all variable-opponent environments (OV-P*), the RL agent when opponents and payoffs are stable (OS-PS), and the ToM agent when opponents are stable but payoffs are variable (OS-PV). In OS-PF, the two ESSes roughly correspond to the two known equilibria in IPD: tit-for-tat (TfT) and always-defect (AllD), respectively[4] (Axelrod, 1984). These results indicate that the agents that won in the particular simulations in the previous section are indeed evolutionarily stable for their respective environments.

Robustness

The previous simulation and evolutionary stability results leave open the question of robustness: will the population converge to the ESSes regardless of the initial conditions? It is possible for a strategy to be evolutionarily stable, but not to serve as an attractor for the full range of population distributions in an evolutionary process (Zeeman, 1980). Thus, here we aim to characterize the relationship between the initial condition and the end state of the evolutionary process for our environments. We only consider initial conditions comprised of *nonzero* proportions of all the agents, since the replicator dynamics is non-innovative (i.e., if the proportion of an agent is exactly 0, it will always stay at 0; if the proportion of an agent is nonzero, it will always be nonzero) (Taylor & Jonker, 1978).

First, we can iteratively eliminate strictly dominated strategies because their pop-

[4] In fact, the NUM, FBM and ToM agents always defect when they have $\lambda = -1$ or $\lambda = 0$, and always cooperates when they have $\lambda = 1$. The RL agent has the same behavior except (a) in the first several rounds, when it has not learned the game, and (b) when playing against TfT, in which case the RL agent with $\lambda = 0$ slowly learns to cooperate with TfT.

ulation proportions will always converge to 0 (see Samuelson and Zhang (1992) for a proof). A strategy is strictly dominated if it performs worse than another strategy against any opponent[5]. When we apply this elimination process, in all the environments except OS-PF, the unique ESSes are the only strategies left, which indicates that in these environments the population will converge to the ESSes regardless of the initial condition.

In OS-PF, the situation is trickier because none of the four strategies (Random, TfT, RL, NUM (equivalent to FBM or ToM)) is strictly dominated. If we have three or fewer strategies, we can use the classification scheme of Zeeman (1980) to characterize the population flow on the full simplex $(x_1, \dots, x_n)$. To our knowledge there is no general classification of the population flow for four or more strategies[6]. To make the analysis feasible, we treat Random as "practically" dominated and eliminate it because its population proportion strictly decreases whenever the total proportion of the other strategies is greater than ≈ 1%. The remaining three strategies correspond to the -4_1 stable class in the classification scheme of Zeeman (1980), which guarantees that the population will always converge to either of the two attractor strategies (TfT and NUM in our case). This is confirmed by the population flow shown in Fig. 2.2C. The population converges to either TfT or NUM regardless of the initial condition, and for the vast majority of the initial conditions the population converges to TfT. When the population is comprised of only TfT and NUM (the bottom edge of the simplex in Fig. 2.2C), it converges to TfT whenever the proportion of TfT is greater than 11.6%.

These results indicate that all (*most* in OS-PF) initial conditions of the population converge to the best-performing agent identified in the previous simulation. To summarize, it is an evolutionarily robust result that the ToM agent performs best only in an environment with stable opponents and variable payoff structures.

[5] Formally, strategy i is strictly dominated iff $\exists j \neq i$ such that $\forall k$, $v_{ik} < v_{jk}$.

[6] In fact, chaos can arise with as few as four strategies (Skyrms, 1992).

2.3 Behavioral experiments

From the evolutionary results, we see that the only environment that uniquely favors ToM is one in which the opponent is stable and the payoff structure is variable. Although the simulation is a much simplified version of the real evolutionary environment, it captures the essence of the stability of other agents and social situations that an agent encounters, and supports the speculation that early humans evolved a ToM on others' WTRs toward themselves because they were once placed in a social environment with stable opponents and variable payoff structures. This speculation can be further supported by demonstrating that people do use a ToM on WTRs in such an environment. However, as mentioned above, ToM is a complex capacity with many components, one of which is being able to predict the opponent's action in a new payoff structure instantly. Because previous experiments in behavioral game theory have only involved one-shot games or fixed repeated games (Camerer, 2011), we do not know whether people can predict, and adapt to, the behavior of opponents under variable payoffs. Therefore, we did two behavioral experiments to test people's ToM on WTRs incrementally[7].

2.3.1 Experiment 1

First, can people play games with variable payoff structures? To answer this, in Experiment 1, we tested whether people have the capacity of a fixed-belief maximizer, i.e., whether they can predict the optimal move for a utility-maximizing opponent and choose their own best move accordingly—a prediction consistent with the naïve utility calculus hypothesis (Jara-Ettinger et al., 2016).

Design

The experiment was implemented as a web page and can be viewed at https://experiments.evullab.org/var-games-8/. The participant plays a sequence of 2×2

[7] The raw data is available at https://osf.io/54d6m/.

normal-form games with a computer agent. The computer agent is a selfish ($\lambda = 0$) naïve utility maximizer, and we explicitly tell the participants that the computer only cares about its own payoffs and does not adapt to the participant's choices. Each trial (each round of the game) is presented as a 2×2 matrix where the computer chooses a row and the participant predicts a row that the computer chooses and chooses a column for herself. The 8 payoff values are integers within $[1, 10]$ and they are size-, shape- and color-coded. Other elements in the interface are shape- and color-coded. After the participant confirms her prediction and choice, the computer's choice is revealed, the payoffs in the resulting cell are added to the participant's and computer's total payoffs respectively via an animation, and a feedback is given as to whether the participant's prediction is correct. The computer's total payoff is hidden because we want the participant to focus on her own payoffs. The experiment ends after (a) the participant's total payoff has reached 120 (i.e., we encourage the participant to have a utility function with $\lambda = 0$), and (b) the participant has correctly predicted the computer's choices 20 times. At the start of each trial (except for the first trial), the participant's total payoff decreases by 2, which serves as another incentive for the participant to maximize her payoffs.

Given the parametrization in Fig. 1A where the participant is player Y, each payoff structure in the experiment satisfies the constraint $w_2 + w_6 = w_4 + w_8$ so that the participant's rational choice is always contingent on the prediction of the computer's choice (otherwise she would be indifferent between the two options). Then the payoff structures are selected such that $w_1 + w_3 - w_5 - w_7$ and $w_2 - w_4$ (which is equal to $w_8 - w_6$) vary over a large range of combinations. There are some catch trials interspersed among the main trials to ensure that the participant is paying attention. In the catch trials, the constraint above is replaced with $(w_2 - w_4)(w_6 - w_8) > 0$ so that the participant's rational choice is *not* contingent on the prediction of the computer's choice. The game sequence is fixed across participants.

Before the main trials, the participant goes through an interactive tutorial to be-

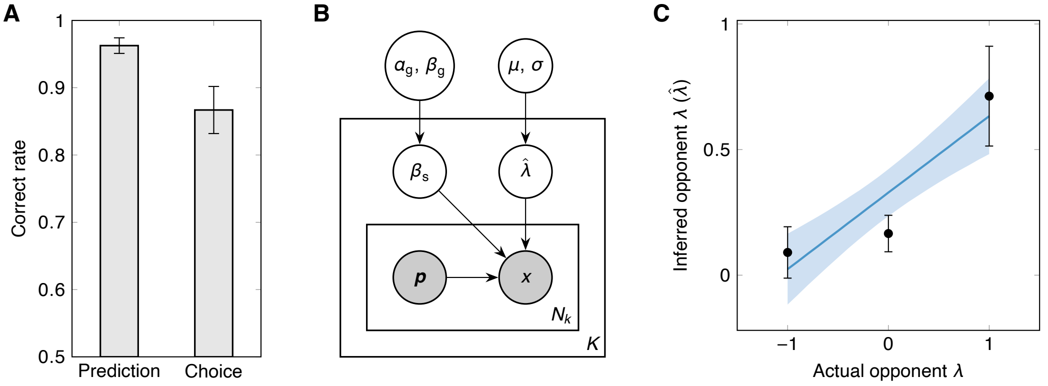


Figure 2.4. Results of the behavioral experiments. (**A**) Correct prediction rates and correct choice rates in Experiment 1. Error bars are standard error of the means (SEMs). (**B**) The hierarchical Bayesian model for Experiment 2, using plate notation. The shaded nodes are observed variables and the unshaded nodes are latent variables. (**C**) Opponent λ inferred by the participants ($\hat{\lambda}$) in Experiment 2, as reflected in their predictions, for different conditions of actual opponent λ. Each point-range indicates the mean and standard deviation of the posterior over μ, and the blue line is an OLS regression fit to the mean of the posterior over each $\hat{\lambda}$.

come familiar with the game and the goal. The payoff structures in the tutorial are set in the same way as the catch trials. After the main trials, the participant completes a questionnaire about her strategy and her interpretation of the goal of the experiment.

Participants

Nineteen participants were recruited on the UC San Diego SONA System and completed the experiment online for course credits. The number of trials they went through ranged from 21 to 34, with 15 participants going through ≤ 23 trials.

Results

The proportion of trials in which participants made the correct prediction or choice is shown in Fig. 2.4A. The correct choice is defined as the option with the higher payoff for the participant given the actual choice of the computer. The participants had almost perfect performance in predictions and very good performance in choices, which indicates that people have the capacity of a fixed-belief maximizer.

2.3.2 Experiment 2

So people can play games with variable payoff structures, but can they infer the opponent's degree of prosociality towards themselves? To answer this, in Experiment 2, we tested whether people behave differently against opponents with different WTRs and whether their behavior reveals an ability to infer the opponent's WTR.

Design

Experiment 2 is similar to Experiment 1 and can be viewed at https://experiments.evullab.org/var-games-9/. The main differences in the design are (a) the model of the computer agent, and (b) how the payoff structures are generated.

The computer agent is still a naïve utility maximizer, but this time the computer's λ varies across three between-subjects conditions: $\lambda = -1/0/1$. We tell the participants that "the computer is simple-minded" and "has a fixed goal", but does not tell them what the goal is.

Each payoff structure in the experiment still satisfies the constraint $w_2 + w_6 = w_4 + w_8$. As described in Appendix 2.A, the only information that the computer's choice offers the participant is whether the computer's λ is greater than or less than the critical λ associated with the payoff structure. So we want the critical λ for the computer to vary over a wide range in order to facilitate the participant's inference of the computer's λ. We split all the trials into blocks of 6 trials (which is opaque to the participant). Within each block, the critical λs of the 6 trials are in the ranges $[-1.3, -1.2]$, $[-0.8, -0.7]$, $[-0.3, -0.2]$, $[0.2, 0.3]$, $[0.7, 0.8]$, and $[1.2, 1.3]$, respectively. The order of the trials within each block and the specific payoff values are randomized. In the tutorial, the critical λ can only be within $[-1.3, -1.2]$ or $[1.2, 1.3]$ so that the computer's choices are the same across conditions.

The game sequence is still fixed across participants. Since it is easier to gain higher payoffs when playing against an agent with a higher λ, the participant's goal for the total payoffs and the payoff deducted at the start of each trial are different across conditions (70 and 2 for $\lambda = -1$, 80 and 3 for $\lambda = 0$, 90 and 4 for $\lambda = 1$).

Participants

Seventy-two participants were recruited on the UC San Diego SONA System and completed the experiment online for course credits. Each participant is randomly assigned to one of the three conditions ($n = 27/22/23$ for $\lambda = -1/0/1$, respectively). The number of trials they went through ranged from 21 to 48, with 56 participants going through ≤ 32 trials.

Data analysis

Since the participant's inference of the opponent's WTR is not directly reflected in their behavior, but indirectly reflected in their predictions and choices, we need to assume a generative model that gives rise to the participant's behavior, with the inferred opponent WTR ($\hat{\lambda}$) as a parameter in the model, and "invert" the generative model to produce an estimate of $\hat{\lambda}$ in each condition. Two assumptions we make in the generative model are (a) that the participant assumes that the computer is a naïve utility maximizer, and (b) that the participant infers a single fixed WTR of the computer ($\hat{\lambda}$). These two assumptions cannot be entirely correct, so the participant's predictions would appear noisy. We make an additional assumption that the participant's prediction in a trial is sampled according to a softmax function over the computer's utilities for the two choices:

$$\begin{aligned} P(A) &= \frac{\exp\left(\beta_{\mathrm{s}} \cdot u(A)\right)}{\exp\left(\beta_{\mathrm{s}} \cdot u(A)\right) + \exp\left(\beta_{\mathrm{s}} \cdot u(B)\right)} \\ P(B) &= 1 - P(A) \end{aligned}$$

where $\beta_{\mathrm{s}} \geq 0$ ("s" stands for "softmax") is the softmax parameter. It would be undesirable to set a common β_{s} for all the participants because (a) what value to choose is rather arbitrary, and (b) some participants are noisier than others, and we want to give them less weight when estimating the mean $\hat{\lambda}$ for a condition.

Therefore, we fit a hierarchical Bayesian model (HBM) to the prediction data to infer each participant's $\hat{\lambda}$ and β_{s} and also the mean $\hat{\lambda}$ for each condition (Fig. 2.4B). We only use the prediction data because given the way we set the payoff structures ($w_2 + w_6 = w_4 + w_8$), predictions represent a more direct inference about the opponent's cooperative stance than choices, the latter of which might reflect participants' idiosyncratic preferences. Suppose there are K participants and participant k plays N_k trials. We assume that each participant's softmax parameter β_{s} is independently sampled from a gamma distribution with shape parameter α_{g} and rate parameter β_{g} ("g" stands for "gamma"), and that each participant's $\hat{\lambda}$ is independently sampled from a normal distribution with mean μ and standard deviation σ. The participant's prediction x in a trial is determined by β_{s}, $\hat{\lambda}$, and the payoff structure $\boldsymbol{p}$ in that trial. We set the priors over α_{g}, β_{g} and σ all to be an exponential distribution with a rate parameter of 1, and use an (improper) uniform prior for μ.

We implemented the HBM in Stan (Stan Development Team, 2021) and fit the model to all the data in each condition separately. We used the default Markov chain Monte Carlo parameters in CmdStan (e.g., No-U-Turn sampler, 4 chains, 1000 warmup iterations per chain, 2000 total iterations per chain).

Results

The inferred $\hat{\lambda}$s for different conditions are illustrated in Fig. 2.4C. The OLS regression on $\hat{\lambda}$ with the actual opponent λ as the predictor reveals a highly significant positive trend ($p < 0.001$), indicating that people can adapt to opponents with different λs, even when such a difference is only reflected in the opponent's choices in the games. This suggests that people do use a ToM on WTRs in an environment with stable opponents and variable payoff structures.

2.4 Discussion

We showed that a minimal form of ToM that infers the opponent's prosociality towards oneself boasts a unique evolutionary advantage in an environment with stable opponents and variable payoff structures. These results highlight a key feature of the social environment that humans find themselves in—that the specific conditions of social interactions change at a much faster rate than the agents with whom one interacts. This feature of an environment is necessary to create the selective pressures to evolve a ToM capacity on WTRs, despite its computational complexity. Furthermore, we show in two behavioral experiments that people are capable of rapid adjustment to payoffs in such variable games, and can adapt to their opponent's prosociality in this manner. Together, these results show that people seem adapted to a game environment largely unexplored in the literature, and that this environment is critical for studying human social reasoning and behavior.

The RL agent as we define it cannot "see" the current payoff structure like the NUM, FBM, and ToM agents. In principle, the RL agent can also be defined in terms of the current payoff structure and potentially have similar capability as and be more flexible than the ToM agent. However, we do not implement this possibility in this work for two reasons. First, the RL agent is meant to be a ceiling of action-level strategies, to illustrate that to succeed in environments with variable payoff structures and/or variable opponents, strategies must be aware of payoff structures, and no matter how adaptive, action-level strategies do not suffice. In settings where both payoff structures and opponents are stable or fixed, action-level strategies and learning seem to be more adaptive than ToM-based strategies. Second, if the RL agent were defined in terms of the current payoff structure, in order for it to accurately predict the opponent's actions and adapt to new payoff structures instantly, it would learn a behavioral policy that maps payoff structures onto actions that is effectively indistinguishable from the ToM agent's policy.

The only difference would be that the RL agent would need some time to *learn* that policy, while the ToM agent derives this policy from its assumptions about the opponent's decision rule. We are agnostic about how people's ToM capacity developed: it could be an innate capacity sculpted by evolution, or could be acquired through an RL-like general learning process, or any mixture of both. In this sense, a sufficiently flexible payoff structure–aware RL agent achieves the same end inference state as the ToM agent, but with a different learning trajectory, and they are, for our purposes, indistinguishable.

We explicitly do not consider reciprocity among agents—agents have fixed λs towards one another regardless of the behavior they observe. This is, of course, an unrealistic description of human social interactions (Fehr & Schmidt, 2006). With reciprocity, stable opponents will remain critical (because inferring an opponent's λ will require sufficient experience with that opponent). However, the mechanism by which ToM helps in the context of variable games might differ slightly. Without reciprocity, ToM is useful in variable games insofar as it allows the agent to better predict how their opponent would behave, to effectively optimize their own choices in light of that forecast. This mechanism of action only applies to non-decomposable games, wherein the best action is contingent on the choice of the opponent (Messick & McClintock, 1968). With reciprocity, ToM would offer a second advantage in variable games: allowing for unexploitable cooperation in environments where behavior-level strategies like tit-for-tat fail due to variable payoffs. Thus, we believe that our simplifying assumption of non-reciprocating agents is sufficient to show the key environment that makes ToM advantageous; however, the advantage of ToM in a variable-payoff, stable-opponent scenario would be even greater in the presence of reciprocity.

Another caveat about our results pertains to the behavioral experiments. While we observe in Experiment 2 that people adapt to their opponent's λ towards themselves, the range of λs they seem to infer appears to be quite limited, as though people have a strong prior toward opponents being either neutral, or positively disposed, toward themselves. This is likely a well-calibrated prior, as few environments provide incentives compatible with sustained spite (i.e., persistently negative λs toward another person). One such setting is costly punishment (Güth et al., 1982), where people are willing to pay a price to harm a transgressor. It would be fruitful to explore in future work whether the restricted range of inferred λs expands in settings where costly punishment may be expected.

Altogether, we show that human-like ToM capacities can evolve in environments with repeated interactions with the same person, but in different payoff structures. This result shows that it may be possible to use the tools of evolutionary game theory and go beyond its traditional constraints such as fixed payoff structures. We believe that with this innovation, future research might explore how other environmental features engender richer ToM capacities and other complex abilities that constitute human social cognition.

Data availability

All code and data are available at https://github.com/jameswhqi/evo-tom.

Acknowledgements

This work was supported by UCSD Academic Senate Grant RG095178.

Chapter 2, in full, is a reprint of the material as it appears in "Qi, W., & Vul, E. (2022). The evolution of theory of mind on welfare tradeoff ratios. *Evolution and Human Behavior*, *43*(5), 381–393." The dissertation author was the primary investigator and author of this paper.

Appendices

2.A Detailed description of the agents

RL agent

The RL agent uses the Q-learning algorithm (Watkins, 1989) to learn the best action in each round given the outcomes in the previous rounds. Each outcome is the conjunction of the agent's choice and the opponent's choice in a given round. The RL agent learns a function that maps from the outcomes in the previous k rounds and the current action to the expected (time-discounted) total utility in the future, so that given the history of previous k rounds, it can choose the current action that corresponds to a higher total utility in the future.

Formally, a Q-learning agent's goal is to choose an action a_t at time t to maximize the expected total reward in the future R_t, with a discount factor $\gamma \in (0,1]$:

$$\begin{aligned} a_t &= \arg\max_{a \in \mathscr{A}} R_t \\ &= \arg\max_{a \in \mathscr{A}} \sum_{t'=t}^{+\infty} \gamma^{t'-t} r_{t'}, \end{aligned}$$

where $\mathscr{A}$ is the set of possible actions and $r_{t'}$ is the reward at time t'. R_t is assumed to be a function of the current state s_t and action a, so the Q-learning algorithm learns a Q-table that records an estimate of R (called the Q-value) for each state–action combination:

$$Q : \mathscr{S} \times \mathscr{A} \to \mathbb{R},$$

where $\mathscr{S}$ is the set of possible states. For our agent, let $a_t, b_t \in \mathscr{A} = \{A, B\}$ be the agent's choice and the opponent's choice in round t, respectively, and $o_t = (a_t, b_t)$ be the outcome of round t. The state is the conjunction of the outcomes in the previous k rounds:

$$s_t = (o_{t-k}, o_{t-k+1}, \dots, o_{t-1}) .$$

The reward r_t is the agent's utility (not just objective payoff, since the RL agent also has a WTR; Eq. (2.1)) in round t. There are 4 possible outcomes in each round, so there are 4^k possible states. Thus a Q-table has 4^k (states) $\times 2$ (actions) entries. If $k = 0$, there is only one state, and values of the possible actions are independent of the history of the game. With a larger k, the agent will learn more slowly (because there are more entries to learn in the Q-table; see below) but be able to adapt to more complex opponent strategies. To make the agent learn quickly in simple scenarios while maintaining flexibility, we let the agent learn 4 Q-tables with $k = 0, 1, 2, 3$ simultaneously.

In round t, given the state s_t, the agent deterministically chooses the action with a higher estimated future reward (the Q-value) averaged across the 4 Q-tables (denoted by $Q_0, \dots, Q_3$):

$$a_t = \arg\max_{a \in \mathscr{A}} \sum_{k=0}^{\min\{3,t-1\}} Q_k(s_t, a) ,$$

where t is one-based, and $\min\{3, t-1\}$ indicates that when $t < 4$, the length of the history is too short to be applicable to some of the Q-tables, which are ignored in calculating the average Q-value. These Q-tables are also ignored in updating the entries according to Eq. (2.4) below. We do not add explicit exploration to the agent, which would entail a non-deterministic decision policy, because (a) the environment is quite simple, and (b) the counterfactual updating mentioned below amounts to partial exploration of the state space. If the Q-values for the two actions are exactly the same (such as in the first round), the RL agent chooses one of the two actions with equal probability.

The learning proceeds by updating one entry in each Q-table in each round after the outcome is observed. In round t, where the state (the previous outcomes) is s_t, after taking an action a_t and receiving a reward r_t, the agent updates the entry in the Q-table corresponding to s_t and a_t as

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left(r_t + \gamma \cdot \max_{a \in \mathscr{A}} Q(s_{t+1}, a) - Q(s_t, a_t) \right), \tag{2.4}$$

where $\alpha \in (0, 1]$ is the learning rate, and s_{t+1} is the next state (i.e., removing the oldest outcome o_{t-k} from s_t and appending the current outcome o_t). We use a discount factor of $\gamma = 0.8$ and learning rates of $\alpha = 0.05 + 0.25k$ for different ks; these choices correspond to hand-tuning to yield effective learning, but qualitative results of the evolutionary simulation do not depend on these parameters.

Usually in the Q-learning algorithm, only one entry (s_t, a_t) in the Q-table is updated after an action a_t is taken, because the rewards for the untaken actions are unknown. However, in our case, we can also update the entry (s_t, a'_t) corresponding to the unchosen action a'_t (the counterfactual choice), whose reward (or utility) is given by the counterfactual outcome $o'_t = (a'_t, b_t)$. Therefore, we update the entries for both the actual choice and the counterfactual choice in each Q-table to accelerate learning.

Critical λ

To formally describe how the FBM and ToM agents make choices and inferences in light of the payoffs and (for ToM) the opponent's choice in a given 2×2 normal-form game, it is useful to introduce the notion of the "critical λ" of such a game. Consider the parameterization of the game shown in Fig. 2.1A. If player X is a NUM and its λ toward Y is λ_X, it will choose A if $(w_1 + w_3) + \lambda_X(w_2 + w_4) > (w_5 + w_7) + \lambda_X(w_6 + w_8)$ and otherwise choose B. Let $\lambda_X^* = \frac{w_1 + w_3 - w_5 - w_7}{w_6 + w_8 - w_2 - w_4}$. Assuming $w_6 + w_8 - w_2 - w_4 > 0$, player X will choose A if $\lambda_X < \lambda_X^*$ and choose B if $\lambda_X > \lambda_X^*$ (the directions of the inequalities are reversed if $w_6 + w_8 - w_2 - w_4 < 0$). We call λ_X^* the critical λ for player X because it is the threshold of λ_X at which X will switch its choice, and λ_X^* is solely determined by the payoff structure. Thus, anyone trying to infer λ_X from observed behavior in a given game gains only limited information: if X is a NUM agent, then the only information that player X's choice in a round provides is whether λ_X is greater than or less than λ_X^* (assuming a perfect decision rule). This constraint applies to any observer, whether it be a ToM player Y adapting to their opponent (as in our simulations), or a researcher studying the behavior of a person (as in our behavioral experiments).

Fixed-belief maximizer

The FBM assumes that the opponent is a NUM whose λ is sampled from a normal distribution with $\mu = 0$ and $\sigma = 0.75$ [8]. This normal distribution roughly matches the actual distribution of λ in the population ($\lambda = -1/0/1$ with probabilities 0.25/0.5/0.25; see Section 2.2.2). To see this, suppose $\lambda \sim N(0, 0.75^2)$. Then $P(\lambda < -0.5) = 0.252 \approx 0.25$ and $E(\lambda \mid \lambda < -0.5) = -0.949 \approx -1$. In each round the FBM calculates the probability that the opponent chooses either action by calculating the probability that the opponent's λ is greater than or less than its critical λ for the payoff structure of that round. Formally, suppose the FBM is player Y and player X's choice in round t is x_t. Then

$$P(x_t = A \mid \mu, \sigma) = \begin{cases} \Phi_{\mu,\sigma}(\lambda_X^*) & w_6 + w_8 - w_2 - w_4 > 0, \\ 1 - \Phi_{\mu,\sigma}(\lambda_X^*) & w_6 + w_8 - w_2 - w_4 < 0, \end{cases} \tag{2.5}$$

$$P(x_t = B \mid \mu, \sigma) = 1 - P(x_t = A \mid \mu, \sigma), \tag{2.6}$$

where $\Phi_{\mu,\sigma}$ is the cumulative distribution function of the normal distribution with parameters μ and σ. After calculating these probabilities, the FBM, like the NUM, deterministically chooses the action with the higher expected utility for itself.

[8] For a FBM, it is equivalent to assume that λ is sampled from the normal distribution and fixed for each opponent, versus that λ is independently *re*sampled from the normal distribution *in each round*, because it does not affect the FBM's prediction of the opponent's choices. However, for a ToM agent, these two assumptions are different, and we will use the second assumption in order to accommodate the non-NUM opponents, as described in the next section.

ToM agent

Similar to the FBM, a ToM agent assumes that the opponent is a NUM, but instead of assuming a fixed distribution of the opponent's λ, it tries to infer λ from the opponent's choices in the games, using iterated Bayesian inference. However, if the ToM agent's model of the opponent is *strictly* a NUM, who has a fixed λ and makes noiseless decisions, the ToM agent is unable to adapt to non-NUM opponents because their decisions in different rounds will likely be inconsistent for any fixed λ (i.e., the likelihood of any λ is 0). In order to behave sensibly against non-NUM opponents, the ToM agent needs to either assume that the opponent's λ is not fixed, or assume that the opponent's decision policy is noisy, resulting in two possible models of the opponent (and a third possible model that assumes both, which we do not consider here).

The first model assumes that λ is not fixed, but re-sampled from a (normal) distribution *in each round*, and the ToM agent infers *simultaneously* the mean μ and standard deviation σ of such a distribution. The learned distribution of λ will have a small standard deviation for a true NUM opponent and a large standard deviation for a non-NUM opponent. Once a λ is sampled in a given round, the opponent deterministically chooses the action with the higher expected utility for itself. We call this model the "μ–σ model".

The second model assumes that λ is fixed, but the decision rule is a noisy (softmax) function over the calculated expected utilities, and the ToM agent infers *simultaneously* the fixed λ and the softmax parameter β. The learned β will be large (less noisy) for a true NUM opponent and small (noisier) for a non-NUM opponent. We call this model the "λ–β model".

These two models are conceptually similar and yield very similar results, so it barely matters which of them we choose. The μ–σ model has the slight advantages that (a) the λ–β model is sensitive to scaling of the payoffs (because of the softmax function), while the μ–σ model is not, and (b) it is easier to enforce the equivalence of FBM and

ToM in the variable-opponent (OV) environments under the μ–σ model since the FBM assumes a noiseless decision rule of the opponent. Therefore, we use the μ–σ model when presenting the results and describe its details here.

The ToM agent assumes that the opponent is a NUM whose λ *in each round* is independently sampled from a normal distribution $N(\mu, \sigma^2)$, and does iterated Bayesian inference on the joint distribution of μ and σ in light of the opponent's behavior in past rounds. In each round of the game, the ToM agent uses the current maximum *a posteriori* (MAP) estimate of $(\mu, \log\sigma)$ (which corresponds to a single normal distribution over λ) to make a decision, in the same way a FBM does. In our simulation, the prior over μ is $N(0, 1)$ and the prior over $\log_2\sigma$ is $N(\log_2 0.75, 1.5^2)$, and they are independent, so that initially the mode of $(\mu, \log_2\sigma)$ is $(0, \log_2 0.75)$, matching the distribution assumed by the FBM. After observing the opponent's choice in round t, x_t, it updates the posterior distribution as

$$p(\mu, \sigma \,|\, x_t, x_{t-1}, \dots, x_1) \propto P(x_t \,|\, \mu, \sigma) p(\mu, \sigma \,|\, x_{t-1}, \dots, x_1),$$

where $P(x_t \,|\, \mu, \sigma)$ is the likelihood function given by Equations (2.5) and (2.6), and $p(\mu, \sigma \,|\, x_{t-1}, \dots, x_1)$ is the posterior from the previous round. We use a two-dimensional grid approximation over μ ($\{-2, -1.8, \dots, 2\}$) and $\log_2\sigma$ ($\log_2 0.75 + \{-3, -2, \dots, 3\}$) for the Bayesian inference.

2.B Mixed-strategy ESS

To look for mixed ESSes in the stability analysis of Section 2.2.3, we first eliminated all strategies that cannot be in a mixed ESS, and check if any possibility of a mixed ESS remains. The Bishop–Cannings theorem (Bishop & Cannings, 1978) tells us that no pure ESS can be part of a mixed ESS, so we can first eliminate the pure ESSes we already found. A mixed strategy containing a strictly dominated strategy can always be invaded by the strategy that dominates it. Formally, if strategy i is strictly dominated by j, then we always have $f_i(\boldsymbol{x}) < f_j(\boldsymbol{x})$ (as defined in Eq. (2.3)). Therefore, according to Eq. (2.2) we always have $\frac{\dot{x}_i}{x_i} < \frac{\dot{x}_j}{x_j}$ as long as $x_i, x_j > 0$, and the population cannot be stable. Thus, we can also iteratively eliminate the strictly dominated strategies. For example, this strategy elimination process in OS-PS proceeds as follows: we first eliminate RL because it is a pure ESS; among the remaining 5 agents, we eliminate Random, TfT and FBM because they are strictly dominated by ToM; among the remaining 2 agents, we eliminate NUM because it is strictly dominated by ToM, leaving only ToM. After applying these two rounds of elimination to each of our payoff matrices, there is always only one strategy left, which means that there is no mixed ESS in any of the environments.

2.C Raw pairwise mean payoffs for all the environments

Fig. 2.5 shows the raw pairwise mean payoffs for all the environments resulting from the simulated tournaments described in Section 2.2.2.

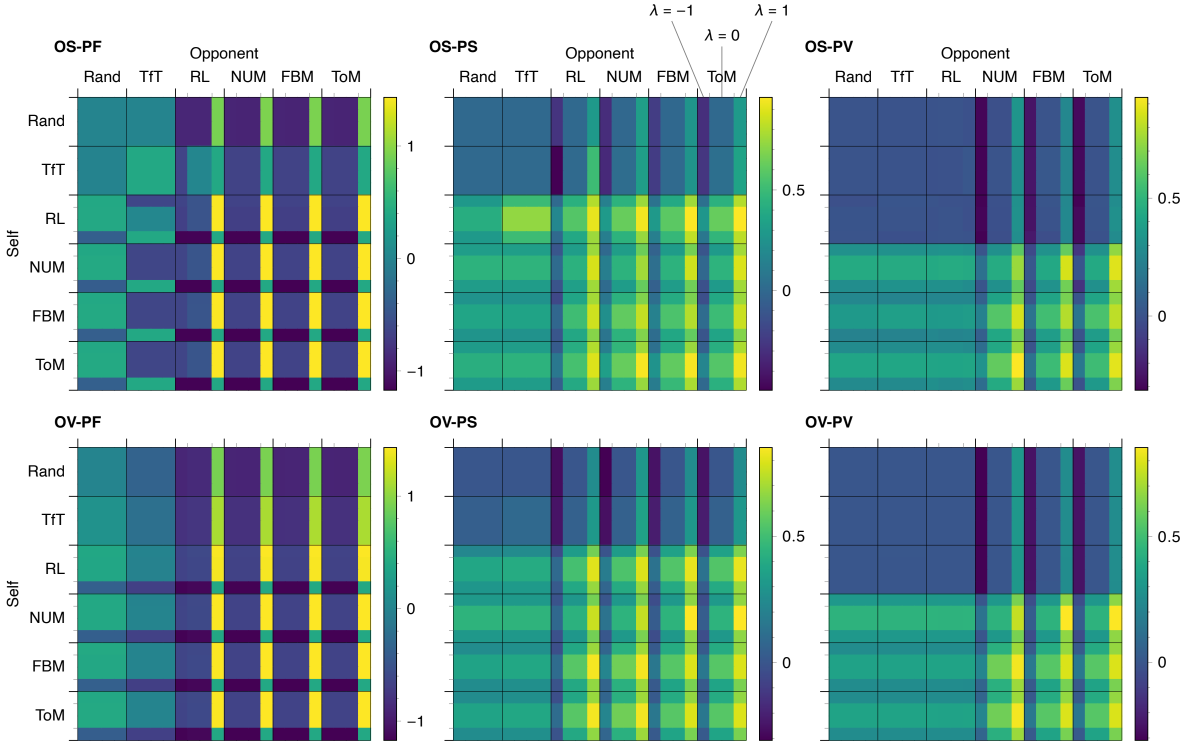


Figure 2.5. The raw pairwise mean payoffs for all the environments.

Chapter 3

The joint evolution of theory of mind and reciprocity in noisy games

We tackle two related evolutionary puzzles: how cooperation among genetically unrelated individuals can evolve across varying social interactions, and how humans evolved the ability to infer others' interpersonal values, a component of "theory of mind". Modeling social interactions as repeated Prisoner's Dilemma games with a fixed payoff structure has enabled the identification of direct reciprocity as a mechanism driving the evolution of cooperation. This work does not address, however, either the range of interactions in which reciprocity can be successful or the mechanism by which it can be achieved when payoffs vary. Human reciprocity relies on not just responding to another person's actions, but also inferring from their actions how much weight that person places on one's own welfare. Such a weight can be formalized as the welfare tradeoff ratio (WTR), a continuous variable. In this work, we examine the evolutionary success of agents that infer and reciprocate graded WTRs in repeated games with variable payoff structures. Across two experiments we also vary whether all agents perceive exactly the same payoffs or instead face some noise in their understanding of one another's perceived payoffs, mimicking the challenges involved in real-life theory of mind. We find that WTR-based agents succeed in both environments but only have a unique evolutionary advantage over simpler agents with a binary conception of cooperation and defection in the noisy environment. These results suggest that both variable payoff structures and uncertainty about how others perceive those changing payoffs play a role in the evolution of graded WTR inference and reciprocity.

3.1 Introduction

Sayings like “one good turn deserves another” and “an eye for an eye and a tooth for a tooth” underscore the importance of direct reciprocity in human society. Both personal relationships and market-based exchange rely on people’s tendency to respond in kind to one another (Buunk & Schaufeli, 1999; Fehr et al., 1998). This reciprocity is motivated by norms as well as by emotions, such as gratitude, anger, and guilt (Fessler & Haley, 2003; Gouldner, 1960; Wubben et al., 2009). Direct reciprocity has been shown to promote cooperation in evolutionary models (Nowak, 2006b; Trivers, 1971), a classic example being Axelrod’s tournament in the iterated Prisoner’s Dilemma game (Axelrod & Hamilton, 1981), where direct reciprocity amounts to returning cooperation with cooperation and defection with defection.

Existing work leaves open questions about the nature and extent of human reciprocity, however. One question regards why humans take a mentalistic approach to reciprocity. Successful cooperation in the Prisoner’s Dilemma can be achieved through the simple tit-for-tat behavioral strategy described above. And yet, people track not just others’ specific actions but also what others’ actions suggest about their underlying social values (Eisenbruch & Krasnow, 2022; Lim, 2012; Quillien et al., 2023). Why engage in this computationally costly form of theory of mind? One answer is that this approach can confer an advantage in interdependent interactions, in which participants make decisions simultaneously and the payoffs of each person’s choice depend on what the other decides to do (Qi & Vul, 2022). Being able to anticipate an opponent’s choice in this type of interaction allows the player to maximize their own gain. However, social life is not dominated by interactions of this sort. Instead, reciprocity often occurs in alternation rather than simultaneously, with relationship partners making independent decisions about what benefits to offer one another. When Alice is deciding whether to give her umbrella to Bob, Bob is not simultaneously making a choice whose outcome would affect the costs and benefits of Alice's decision.

Here we ask: are there conditions under which theory of mind supports reciprocity in repeated interactions without strong interdependence? We address this question using the tools of evolutionary game theory, where social interactions are modeled with games and different strategies compete for evolutionary success. We focus on repeated games with variable payoff structures, an ecologically valid context identified by Qi and Vul (2022) as an environment where WTR inference might be useful.

3.1.1 Welfare tradeoff ratio inference and adjustment

When one person's actions benefit another person's welfare, fully understanding the generosity of those actions involves knowing how they affected the actor. If the actor also benefited from the action, then they may not have been motivated by the recipient's welfare at all. In contrast, if the action involved a cost to the actor, then it was likely motivated by a desire to improve the welfare of the recipient.

How much one person values another's welfare relative to their own can be quantified as a "welfare tradeoff ratio (WTR)" (Delton & Robertson, 2016; Tooby & Cosmides, 2008). Suppose Alice is interacting with Bob. Alice's utility, which she tries to maximize in each decision, can be written as

$$u = w_{\mathrm{s}} + \lambda w_{\mathrm{t}}, \tag{3.1}$$

where w_{s} is the welfare of Alice (the self), w_{t} is the welfare of Bob (the "target" person), and λ is Alice's WTR toward Bob (Delton & Robertson, 2016; Delton et al., 2023; Tooby & Cosmides, 2008)[1]. In the rest of this paper, we mostly use WTR in the text and λ

[1] In this work, "welfare" or "payoff" refers to the objective welfare of an organism that directly affects its reproductive fitness, as is standard in the WTR and evolutionary game theory literature. This is slightly different from classic game theory, where the payoffs in games are often conceptualized as von Neumann–Morgenstern utilities, so it is unnecessary to consider any preferences other than pure self-interest (Fudenberg & Tirole, 1991). Given the utility functions of the players, it is possible to rewrite the payoffs in a game from denoting objective welfare to denoting subjective utility, but that would confound the objective

in equations, but they have the same meaning. The WTR is closely related to another construct called the social value orientation that is also meant to capture how people consider others' welfare in their decision making (SVO; Messick & McClintock, 1968; Murphy et al., 2011; Van Lange et al., 1997).

Alice's WTR toward Bob at a given moment is influenced by many factors, including Alice's basic generosity toward everyone else (Piff et al., 2010; Van Lange et al., 1997), Alice and Bob's social distance and relationship (are they friends, strangers, or foes? Jones & Rachlin, 2006; Qi et al., under review), and their recent history of interactions (did Bob just do something bad to Alice, thus revealing a low WTR on Bob's part toward Alice? Ackermann et al., 2016; Lim, 2012). In this paper, we are mostly concerned with the last-mentioned source of influence on WTRs: the adjustment of WTRs in a reciprocal fashion, based on the ability to infer another person's WTR toward oneself.

People can infer others' WTRs from observing their decisions. Observers can infer others' specific WTR toward themselves, or a broader WTR toward social partners in general (i.e., trait-level generosity or "warmth") (Lim, 2012; Quillien et al., 2023; Sell et al., 2017). Observers use these inferences to predict future behavior and place particular importance on others' WTRs in partner choice and impression formation (Hackel et al., 2015; Quillien et al., 2023; Wojciszke et al., 1998). People rationally seek out information diagnostic of others' WTRs toward them, and when choosing partners for future interactions they weight WTRs over both competence and raw rewards received from different partners in the past (Eisenbruch & Krasnow, 2022; Quillien, 2023).

People's responses to specific decisions others make are also consistent with WTR inference, rather than a simple strategy of reacting positively when a partner acts to benefit them and negatively when a partner forgoes benefits or imposes costs. For example, people are more grateful for cooperation if it indicates their partner's high WTR toward

payoffs and the parameters in the utility function (WTR in our case). Since we are interested in variations in both the objective payoffs and WTRs, we keep them separate.

themselves, independent of the overall magnitude of the benefit they receive. Conversely, they are more upset about defection when it indicates a very low WTR held by the defector toward themselves, independent of the magnitude of costs imposed (Lim, 2012; Sell et al., 2017; Smith et al., 2017). People also adjust their own WTRs based on such inference in a reciprocal fashion—when they infer that their partner has a higher WTR toward themselves than their expectation, they tend to increase their WTR toward the partner, and vice versa (Ackermann et al., 2016; Lim, 2012).

3.1.2 The evolution of WTR inference

Why has evolution selected such sophisticated and expensive cognitive mechanisms for reciprocity and partner choice? It is plausible that WTR inference and adjustment support the maintenance of cooperation through reciprocal altruism (Axelrod & Hamilton, 1981; Trivers, 1971) across varying circumstances. But despite the widespread interest in the evolution of cooperative behavior, there has been little formal work on the evolution of the capacity for WTR inference (except, e.g., Eisenbruch & Krasnow, 2022; Qi & Vul, 2022).

Most existing formal models of reciprocal altruism are based on repeated games with fixed payoff structures, like the iterated Prisoner's Dilemma (Axelrod & Hamilton, 1981). When payoffs in these games are fixed and cooperation and defection can be defined as specific, raw actions, then tit-for-tat or win–stay, lose–shift strategies defined in terms of those raw actions are successful against a wide variety of other strategies (Axelrod, 1984; Nowak & Sigmund, 1993). Being able to infer the opponent's WTR adds computational costs but does not make an agent perform better (Qi & Vul, 2022), because in such simple games, the agent's behavior can almost always be reduced to a strategy defined in terms of the raw actions, at least approximately.

Qi and Vul (2022) expanded the space of game environments beyond fixed repeated games by manipulating the variability of the social partners ("opponents" in game

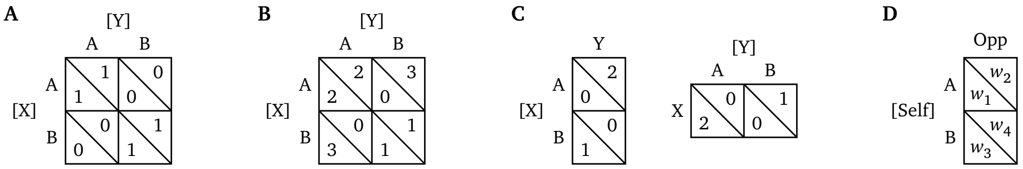


Figure 3.1. Decomposition of 2 × 2 games. A and B are possible actions, X and Y (or Self and Opp) are the two players, and the numbers (or variables) are payoffs to the two players given their actions. Players in square brackets are acting; otherwise they are only receiving payoffs. (**A**) A non-decomposable game, which is also a coordination game. (**B**) A decomposable game, which is also a Prisoner's Dilemma, with A and B corresponding to cooperation and defection, respectively. (**C**) One possible decomposition of the game in (**B**) into the sum of two one-player games. (**D**) The parameterization of a one-player game.

theoretical terms) and the payoff structures. Out of the game environments they considered, WTR inference offers an unique advantage only in environments with stable opponents (two agents interact with each other many times over their lifetime, i.e., they play repeated games) and variable payoff structures (the payoffs involved are different for each social interaction). In such game environments, WTR inference allows players to have a more accurate model of their opponent, so that they can better predict their opponent's actions in games with novel payoff structures, while agents who do not make decisions based on actual payoffs or agents with a fixed model for every opponent perform worse. Better predictions lead to better actions for the players themselves, so that they can maximize their utility more effectively.

However, such an advantage depends on the strong interdependence between the two players' actions. In 2 × 2 normal form games (the class of games studied in Qi and Vul, 2022), where two players act simultaneously, one player's best action for herself often depends on the other player's action. For instance, in the game of Fig. 3.1A, player X's best action depends on what Y chooses: assuming X is entirely selfish (her WTR toward Y is 0), if Y chooses action A, X would prefer A to B, and vice versa. In such situations with strong interdependence, better predictions of the opponent's actions lead to higher utilities for oneself.

But there are also games without strong interdependence, such as the game of

Fig. 3.1B. This particular Prisoner's Dilemma[2] can be decomposed into the sum of two one-player decisions, such as those depicted in Fig. 3.1C, called the decomposed game (of the corresponding matrix game in Fig. 3.1B) (Messick & McClintock, 1968; Pruitt, 1967). The matrix game and the decomposed game are equivalent as long as the two players act simultaneously, and either player's best action is independent of the other player's action as long as they are maximizing a utility function in the form of Eq. (3.1). Either player only needs to consider the 4 payoffs in her part of the decomposed game to make a decision, which will be the same decision as considering the 8 payoffs in the matrix game based on any prediction of the other player's action. The two players are essentially playing two independent one-player games.

Games with strong interdependence are widely studied in the game theory literature, partly because they often cannot be simply solved by iterated elimination of dominated strategies (Fudenberg & Tirole, 1991). However, strong interdependence seems rare in real-world social interactions. Most of our decisions that affect someone else's welfare are not influenced by a simultaneous decision by that person (although they might be influenced by our predictions of how they will influence that person's future decisions). Even when the ambiguity about the other person's simultaneous action arises, we often use language or other means to resolve that ambiguity—we are rarely stuck in a "prison" situation. Therefore, while the same behavior might have different consequences for oneself and others in different situations, in each situation it is often clear what the consequences are, and we are effectively playing one-player games.

In decomposable games such as Fig. 3.1B, better predictions of the other player's actions do not lead to better actions and higher utilities for oneself in the current game, so WTR inference as studied in Qi and Vul (2022) would not be advantageous. One way WTR inference might contribute to evolutionary success, even if all the games are decomposable or one-player, is through direct reciprocity (Axelrod & Hamilton, 1981;

[2] Not all Prisoner's Dilemmas are decomposable.

Nowak, 2006b; Trivers, 1971), which is not possible in Qi and Vul (2022) because the agents are assumed to have fixed WTRs. In other words, although there is no strong interdependence between the two players' actions within each one-shot game, the ability of inferring and adjusting WTRs creates interdependence between the two players' actions over time and repeated interactions.

3.1.3 The present research

In this work, we examine the conditions under which WTR inference creates an evolutionary advantage in an environment without strong interdependence but with repeated interactions. We model the interactions as alternating games (detailed below) instead of repeated simultaneous games, while keeping the payoff structures variable (Qi & Vul, 2022). We include different types of agents with varying complexity, and pit them against each other in a simulated tournament, resulting in a pairwise mean payoff matrix. We then use the pairwise mean payoff matrix to derive the relative evolutionary viability of different agents. We also randomly perturb the hyperparameters to test the robustness of the conclusions.

Two experiments take different approaches to noise in the repeated games. We first consider a noiseless environment in Experiment 1. Agents in these games implement their own strategies and perceive other players' options and choices without error. However, real-world environments are imbued with noise. One common type of noise relevant to theory of mind involves uncertainty about other people's perception of the world (Jara-Ettinger et al., 2016). In the context of cooperation, this sort of noise may take the form of uncertainty about the payoffs another person associates with each action or inaction. When I give my umbrella to someone, how much exactly does it decrease my welfare and increase that person's welfare? The two individuals involved can perceive different payoffs—while I think my cost is low when I give out my umbrella since I have another one available, the other person might not know this and might think my cost is high. Such asymmetry in the noise can lead to misperception of intentions; e.g., an observer perceiving an action as more or less generous than the actor considered it to be. Reciprocity based on graded WTR inferences may be less impacted by these mistakes than reciprocity based on binary "cooperate" or "defect" classifications. In Experiment 2, we add noise to the recipient's perception of the choosing player's payoffs to test if an agent that makes graded, probabilistic inferences about others' WTRs has a particular advantage in the context of this ecologically valid type of uncertainty.

3.2 Methods

3.2.1 Variable alternating games

Since there is no strong interdependence between the two players' actions, we could model the interactions as repeated (decomposable) simultaneous games, strictly alternating games (two players take turns playing one-player games), or randomly alternating games (in each round one player is randomly chosen to play a one-player game) (Nowak & Sigmund, 1994). Alternating games are usually more realistic than simultaneous games as discussed above, and the randomly alternating games are usually more realistic than the strictly alternating ones since one player's decision is not necessarily followed by the other player's. We choose to model the interactions as strictly alternating games, because there is already plenty of randomness in our setup due to the variable payoff structures: in some games the player's decision might have great significance (being cooperative and being selfish lead to different choices and very different outcomes), while in others games the player's decision might have little significance (the player would choose the same action whether she is cooperative or selfish). This has a similar effect as the randomly alternating Prisoner's Dilemma (Nowak & Sigmund, 1994).

Two players take turns making decisions in a series of one-player games. In each one-player game (Fig. 3.1D), called a stage game, one player is the actor and the other player is the observer (Self and Opp in Fig. 3.1D, respectively). Which player serves as the actor first is determined randomly. The actor makes a binary decision between two options, each of which leads to one payoff for the actor and another payoff for the observer. The 4 possible payoff values ($w_{\{1,2,3,4\}}$) are i.i.d. sampled from the standard normal distribution. The overall repeated game, called a supergame[3], encompasses 200 rounds of stage games (i.e., either player makes 100 decisions).

A useful property of each stage game is the "critical WTR" of the game, denoted by $\hat{\lambda}$, which is the threshold of the actor's WTR at which she switches her choice given a noiseless decision rule:

$$\hat{\lambda} = \frac{w_2 - w_1}{w_3 - w_4}.$$

For instance, in Fig. 3.1C, the one-player game for either player has $\hat{\lambda} = 0.5$—if X's WTR toward Y is greater (/less) than 0.5, she will choose option A (/B); if X's WTR toward Y is exactly 0.5, she will be indifferent between the two options. From the actor's choice, the observer can gain information about whether the actor's WTR is above or below the critical WTR. Since $w_2 - w_1$ and $w_3 - w_4$ are independently distributed according to the same normal distribution with zero mean, $\hat{\lambda}$ follows the standard Cauchy distribution.

It is also useful to have a general notion of cooperation and defection in this environment. The core conflict in a social dilemma like the Prisoner's Dilemma is the conflict between maximizing one's own payoff and maximizing the total payoff of the group. Therefore, we define full cooperation as having $\lambda = 1$ (maximizing the sum of the two players' payoffs), and full defection as having $\lambda = 0$ (maximizing the actor's own payoff). Evolutionarily, full cooperation between agents with the same strategy maximizes the payoffs they receive, which is important for cooperation to establish and maintain dominance in a population. On the other hand, if strategies are non-reciprocal, full defection performs at least as well as any other strategy given any fixed opponent, and is thus able

[3] "Supergame" in game theory is often a synonym of "repeated game", but here it refers to the specific repeated game we use.

to invade any dominant strategy (Weibull, 1997). Having a WTR between 0 and 1 might strike a balance between these two objectives, while having $\lambda > 1$ or $\lambda < 0$ is almost never beneficial, unless it can induce the opponent to be more cooperative or change future payoff structures (which are not properties of any agents or games we consider). In our setting, we define a social dilemma to be a game whose critical WTR ($\hat{\lambda}$) is between 0 and 1, where full cooperation and full defection dictate different decisions. Given the distribution of $\hat{\lambda}$, on average 25% of games are social dilemmas.

3.2.2 Agents

We include nine types of agents of varying complexity in the environment. In classic game theory, strategies are usually defined in terms of the raw actions in a game, such as cooperation and defection in a Prisoner's Dilemma, with no reference to the payoffs, because the payoffs are often fixed (Axelrod, 1984; Fudenberg & Tirole, 1991). We do not consider such action-level strategies because they cannot be efficiently defined when payoffs vary from game to game (Qi & Vul, 2022). Instead, we assume that each agent maximizes a utility function in the form of Eq. (3.1) in each round of the game, and different agents have different patterns of WTRs toward their opponents.

The first three types of agents have fixed WTRs. Always Defect (AllD) has $\lambda = 0$, Always Cooperate (AllC) has $\lambda = 1$, and Half Cooperate (HalfC) has $\lambda = 0.5$.

The fourth type is called tit-for-tat (TfT). It implements a form of direct reciprocity, and is a generalization of the tit-for-tat strategy in the iterated Prisoner's Dilemma (Axelrod & Hamilton, 1981) based on the definitions of cooperation and defection given above. Critically, TfT has a binary, not graded, notion of cooperation and defection, in contrast to the Bayesian agent described below. TfT starts with $\lambda = 1$ toward its opponent. It assumes that its opponent is in one of two states at any instance, either $\lambda = 0$ or $\lambda = 1$. It can distinguish between these two states only after its opponent has made a decision in a social dilemma (i.e., a game that pits the pair of payoffs with the higher sum against a pair of payoffs with greater value for the deciding agent), at which point the TfT agent will update its own WTR to match its opponent's. If the game its opponent played was not a social dilemma, and thus provides no information about the opponent's WTR within the range of 0 to 1, the TfT agent's WTR will stay the same. Therefore, unlike tit-for-tat in the iterated Prisoner's Dilemma which copies its opponent's action in every round, here TfT only updates its WTR occasionally, about 25% of the time.

Although the description of TfT here makes reference to mental processes about inferring the opponent's WTR, this is not the type of WTR inference we are really interested in, because people have both graded WTRs (not just 0 and 1) (Jones & Rachlin, 2006; Murphy et al., 2011; Qi et al., under review) and graded representations of other people's WTRs (Lim, 2012; Quillien et al., 2023). In addition, TfT can be implemented as a heuristic rule that does not involve explicit inference of WTRs: if my opponent's choice led to a higher payoff for himself but a lower total payoff than the alternative, set my WTR to 0; if the inverse is true, set my WTR to 1; otherwise keep my previous WTR. It may seem more natural to us to express this rule as WTR inference, but computationally it is much simpler than, e.g., the Bayesian agent below. Therefore, it seems plausible that natural selection would prefer this rule to an equivalent strategy based on WTR inference.

The fifth type of agent is called the Naïve TfT, which behaves similarly to TfT but has a simpler and inaccurate conception of cooperation and defection. In every game where Naïve TfT is the observer, if its opponent chooses the option with the higher payoff for Naïve TfT, it treats it as a cooperation; otherwise it treats it as a defection. This is essentially a self-centered view of others' behavior, which only considers the benefits others generate for oneself but not the costs they incur. Like TfT, when the Naïve TfT thinks its opponent has cooperated (/defected), it sets its own WTR to 1 (/0). This means that Naïve TfT cooperates not by selecting the higher payoff for its opponent (i.e., its definition of cooperation when perceiving the opponent's decision), but by selecting the pair of payoffs with the higher total value. For the Naïve TfT, every game played by the opponent is informative, even if the game is not a social dilemma, because the payoffs w_2 and w_4 are different with probability 1.

The sixth and seventh types are generous versions of TfT and Naïve TfT. In the iterated Prisoner's Dilemma, when there is implementation noise (i.e., one agent makes a mistake in its selection, which certainly happens in the real world), tit-for-tat when playing against itself may be stuck in alternating cooperation and defection (in simultaneous games) or prolonged mutual defection (in simultaneous and alternating games) (Nowak, 1990; Nowak & Sigmund, 1994). Generous tit-for-tat is a strategy that, with some probability, forgives its opponent's defection and reverts back to cooperation, thus dealing with this problem effectively (Nowak, 1990; Nowak & Sigmund, 1994). Similarly, in our setting, the Generous TfT and the Generous Naïve TfT unconditionally cooperate (set $\lambda = 1$) with some probability in each round. In the iterated Prisoner's Dilemma with a fixed payoff structure, the optimal probability for unconditional cooperation can be derived from the payoff values. In our setting, the analysis is more difficult, so we set a default value of 0.1 and examine its robustness later.

In the noisy iterated Prisoner's Dilemma, a strategy that can perform even better than generous tit-for-tat is "win–stay, lose–shift" (Nowak & Sigmund, 1993), but it performs poorly in the alternating version of the game, unless it is redefined in a much more complex and unnatural way (Nowak & Sigmund, 1994). Therefore we do not include it in our experiments.

The eighth type of agent, called Bayesian, is the only agent able to perform graded WTR inference. The Bayesian agent assumes that its opponent is always maximizing a utility function in the form of Eq. (3.1), and performs approximate Bayesian inference on its opponent's WTR in the i-th decision, denoted by $\lambda_{\text{opp}}^{(i)}$. It assumes that λ_{opp} can change gradually over time (i) according to a Markov process, so its model of its opponent is a hidden Markov model where the hidden states are $\lambda_{\text{opp}}^{(i)}$, and the emission probabilities are the probability that its opponent chooses either action in game i given $\lambda_{\text{opp}}^{(i)}$. After observing the i-th decision by its opponent, the Bayesian agent calculates a posterior distribution over $\lambda_{\text{opp}}^{(i)}$ given all evidence so far, and sets its own WTR to match the posterior in a tit-for-tat-like fashion. Specifically, it sets its WTR to the 55th percentile of the posterior, but never greater than 1. This slight bias toward 1 prevents two Bayesian agents from drifting away from mutual full cooperation, while retaining their ability to almost exactly copy their opponent's WTR when enough evidence has been accumulated (i.e., the posterior has a small standard deviation). The initial distribution over λ_{opp} is a normal distribution with mean 1 and standard deviation 0.5, so that the Bayesian agent starts with full cooperation, like TfT. See Appendix 3.A for details of the Bayesian agent.

The last type of agent is called Slow TfT. It is inspired by the Bayesian agent's behavioral signature—we wonder whether an agent who behaves similarly to the Bayesian agent but does not have graded WTR inference can perform as well. One behavioral signature of the Bayesian agent is that when its opponent's decision reveals a very different WTR from its current belief, it tends to adjust its belief and its own WTR gradually, rather than suddenly, to match the opponent's, especially in a noisy environment (Experiment 2 below). For instance, suppose that the Bayesian agent is playing against AllD, and all of AllD's decisions have a critical WTR of 0.5. In a noisy environment, the median of the Bayesian agent's posterior over AllD's WTR might go through such a sequence: 1, 0.5, 0.3, 0.1, 0, ... This is due to the assumption that sudden large changes in the opponent's WTR have low probability (which is a necessary assumption for accumulating evidence across multiple decisions). Slow TfT mimics this kind of behavior. It still has a binary conception of cooperation and defection, but when it infers that its opponent's WTR is different from its own, it adjusts its own graded WTR only partially toward that. Let λ_{self} be its original WTR, and λ_{opp} be the opponent's WTR it infers (0 or 1). It sets its new

WTR to be

$$\lambda'_{\text{self}} = (1-a)\cdot\lambda_{\text{self}} + a\cdot\lambda_{\text{opp}},$$

where a is called the change rate, and $a = 0.55$ if $\lambda_{\text{opp}} = 1$ and $a = 0.45$ if $\lambda_{\text{opp}} = 0$, so that two Slow TfT agents have a tendency to converge back to $\lambda = 1$, like the Bayesian agent.

The nine types of agents can be grouped in terms of the level or complexity of "theory of mind" they have. AllD, AllC and HalfC have no theory of mind whatsoever. The TfT variants impute a binary state to their opponent—either cooperation or defection. Naïve TfT and Generous Naïve TfT have a weak theory of mind because they only consider their own payoffs when interpreting the opponent's actions as cooperation or defection. TfT, Generous TfT and Slow TfT have a stronger theory of mind because they consider both their own and their opponent's payoffs when interpreting the opponent's actions. The Bayesian agent has the strongest theory of mind because it imputes a continuous state to its opponent—the graded WTR.

3.2.3 Tournament

For each pair of agents, including the same type, we simulate a supergame between them, calculate their mean payoffs per round, and repeat this process 100,000 times so that the standard errors of the mean payoffs are negligible. This results in a 9×9 pairwise mean payoff matrix, like the one in Fig. 3.2A, which is used in the evolutionary analyses below.

3.2.4 Evolution

To assess the relative evolutionary success of different agents, we consider the evolutionary game dynamics in a finite well-mixed population of size n, modeled by the frequency-dependent Moran process (Moran, 1962; Nowak et al., 2004)[4]. We set $n = 100$ by default. At each time step, one agent in the population is randomly chosen to die. At the same time, with probability u (the mutation probability), a random mutant out of the 9 possible types is added to the population; with probability $1-u$, an existing agent in the population (which can be the one who is dying) is probabilistically chosen to reproduce (i.e., create a new agent of the same type). Agent i's probability to be chosen for reproduction is proportional to its fitness f_i, which is an exponential transformation[5] of its current mean payoff $\bar{w}_i$:

$$\begin{aligned} f_i &= \exp(\alpha \bar{w}_i)\,, \\ \bar{w}_i &= \frac{1}{n} \sum_{1 \le j \le n} w_{ij}\,, \end{aligned} \tag{3.2}$$

where $\alpha > 0$ reflects the selection strength, and w_{ij} is agent i's payoff when playing against agent j according to the pairwise mean payoff matrix. We set $\alpha = 5$ by default.

The relative success of different types of agents can be reflected in the stationary distribution of agent types under the low-mutation limit. Under the low-mutation limit (vanishingly small u), the population is almost always homogeneous and consists of only one type of agents. When a mutant is introduced, it either goes extinct or takes over the population before another mutant is introduced, and the probability that it takes over is called the fixation probability (Nowak, 2006a). Therefore, the evolutionary process can be described as a Markov process where the 9 possible types of agents correspond to the 9 possible states. We can analytically calculate the stationary distribution of the agents, which reflects their average abundance in the population over very long time periods.

[4] We use a discrete-time stochastic dynamics in a finite population instead of a continuous-time deterministic dynamics in an infinite population, such as the replicator dynamics (Taylor & Jonker, 1978), because (a) the former is usually more realistic, (b) the latter is sensitive to initial conditions, even when mutation is present, and (c) for the latter, it is more difficult to calculate long-run averages.

[5] We use an exponential transformation instead of the usual linear transformation (Nowak et al., 2004) because the former ensures that fitness is always positive, which is particularly relevant given our variable payoffs.

See Rand and Nowak (2012) for details of the calculation.

We also run an agent-based simulation to verify the results from the low-mutation limit. We set $u = 0.01$, initialize the population randomly, run the simulation for 10^7 time steps, and calculate the average distribution of the agent types over the later 90% of the time steps.

3.2.5 Hyperparameter perturbations

There are 14 hyperparameters we can tweak in the alternating games, the agent definitions, and the evolutionary dynamics, which might influence the results (see Appendix 3.B for the full list of hyperparameters). To test the robustness of the results, we perturb the 14 hyperparameters independently from their default values according to some random distribution (see Appendix 3.B for the details of the distributions). Each set of perturbations results in a new set of hyperparameters, which we use to simulate the tournament for 10,000 repetitions and compute the stationary distribution of agent types under the low-mutation limit. We do this for 200 sets of hyperparameters and examine the variability of the distribution.

3.2.6 Noise

In Experiment 1, there is no noise in the environment. All agents observe the payoffs and make decisions perfectly. In Experiment 2, we add noise to simulate people's uncertainty about one another's expected utilities, a factor that may increase the value of graded WTR inference. Specifically, we introduced noise to the observer's perceptions of the payoffs faced by the actor. In each game, the actor perceives the four actual payoffs, while the observer's perceived payoffs are the actual payoffs plus four error terms sampled i.i.d. from a normal distribution with mean 0 and standard deviation 0.1 (by default). Neither player knows what exact payoffs the other player perceives, but can have beliefs about their plausible values based on the error distribution. The only

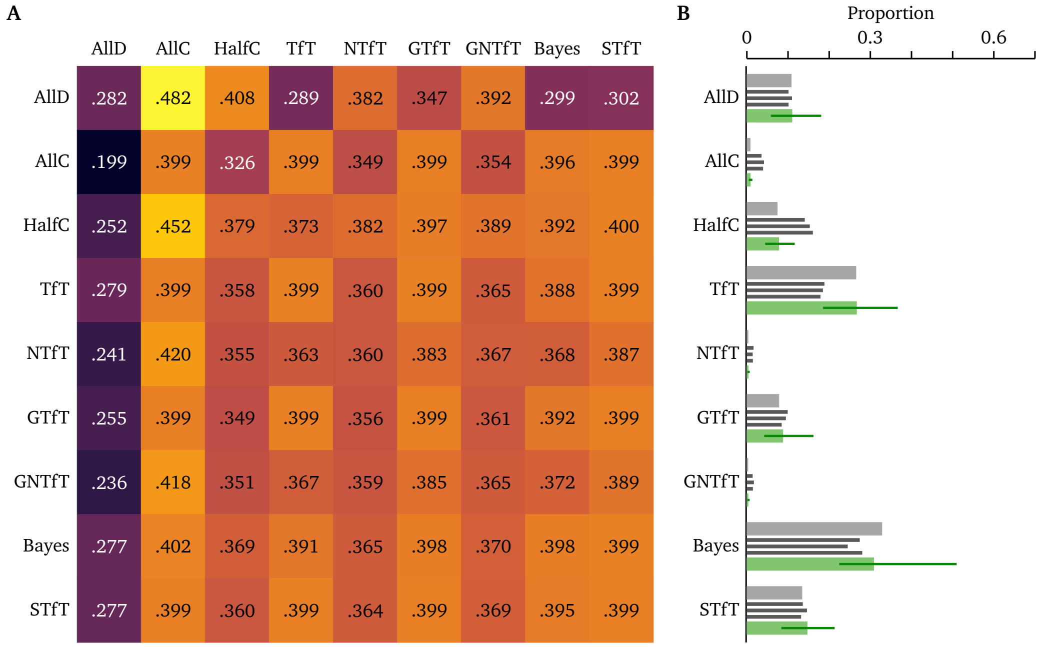


Figure 3.2. Main results in Experiment 1. NTfT: Naïve TfT; GTfT: Generous TfT; GNTfT: Generous Naïve TfT; Bayes: Bayesian; STfT: Slow TfT. (**A**) The pairwise mean payoff matrix in the tournament. Each cell is the payoff received by the row player when playing against the column player. When two agents of the same type play against each other (cells on the diagonal), the payoffs of one of them are used to calculate the mean payoff. (**B**) The evolutionary distributions of the agents. The thick light gray bars are the stationary distribution under the low-mutation limit. The thin dark gray bars are the average distributions from three replications of the agent-based simulation. The green bars and ranges show the variability of the stationary distribution under the low-mutation limit with respect to different sets of hyperparameters. They show the 10th percentile, median and 90th percentile of each agent's proportion in the stationary distribution.

agent that makes use of such beliefs is the Bayesian agent (see Appendix 3.A). The other reciprocating agents rely on the individual or summed payoffs they perceive to make binary judgments about whether the actor cooperated or defected.

3.3 Results

3.3.1 Experiment 1

The main results are shown in Fig. 3.2 and some notable behavioral traces are shown in Fig. 3.3.

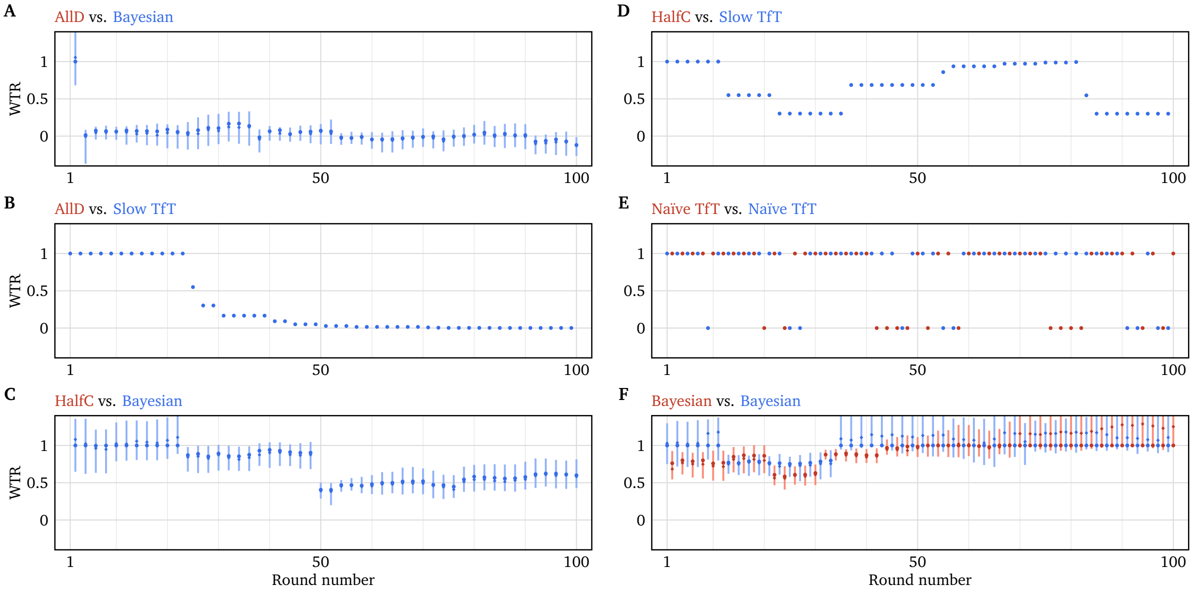


Figure 3.3. Notable behavioral traces with respect to WTR in Experiment 1. Each panel displays one supergame between two types of agents. Only the first 100 rounds are shown. For AllD, AllC and HalfC, no trace is shown because they have constant WTRs. For TfT variants, each circle dot is the agent's WTR when making a decision. For the Bayesian agent, in each round, the vertical line is the interquartile range of the agent's current posterior over the opponent's WTR, the rhombus is the median of the posterior, and the circle dot is the WTR that the agent actually uses when making the decision.

Payoff matrix

From the pairwise mean payoff matrix (Fig. 3.2A), we see that the two (strict) evolutionary stable strategies are AllD and the Bayesian agent, who cannot be invaded by a mutant strategy after establishing dominance in the population[6], because they receive a strictly higher payoff playing against themselves than against any other strategy (Maynard Smith, 1982). AllC, TfT, Generous TfT and Slow TfT cooperate perfectly with themselves. The neutrally stable strategies are TfT, Generous TfT and Slow TfT, who can be invaded by AllC, TfT, Generous TfT or Slow TfT through neutral drift, but the invaders cannot thrive by earning a higher payoff than the incumbent strategy (Maynard Smith, 1982). Of these cooperative agents, TfT is the best at resisting invasion by AllD (it gains the highest payoff while AllD gains the lowest). If AllC takes up a large enough proportion of the population, the mutually cooperative population composed of AllC and TfT variants can be invaded by AllD or HalfC (or even Naïve variants of TfT because their WTRs essentially switch between 0 and 1 randomly; see below).

The Naïve versions of TfT and Generous TfT cannot cooperate well with themselves because they do not have an accurate conception of cooperation and defection in varying circumstances and often mistake cooperation for defection and vice versa (Fig. 3.3E). Their WTRs effectively switch between 0 and 1 randomly, and they are strictly dominated by HalfC. Therefore, we mostly ignore them in the following discussions. This explains why, when people form a positive or negative impression of a social partner, they take into account not only the benefits the partner generates for themselves, but also the costs that the partner incurs (Eisenbruch & Krasnow, 2022; Lim, 2012).

The Bayesian agent is slightly worse at cooperating with itself compared to AllC, but only by a tiny amount. Two Bayesian agents occasionally do not cooperate perfectly

[6] This only applies to deterministic dynamics in an infinite population, like the replicator dynamics (Taylor & Jonker, 1978). For stochastic dynamics in a finite population, like the Moran process we use (Moran, 1962; Nowak et al., 2004), any mutant could in principle invade an evolutionary stable strategy, although with low probability.

because games with $\hat{\lambda} > 1$ can lead to a posterior whose 55th percentile is less than 1, even when the opponent's WTR is in fact always 1 (see, e.g., Fig. 3.3F, Round 2). Despite such deviances, two Bayesian agents are likely to converge back to mutual full cooperation due to the slight generosity in the 55th-percentile decision rule (Fig. 3.3F, Round 51). The Bayesian agent's ability to resist invasion by AllD is slightly worse than TfT and similar to Slow TfT because its average WTR is slightly higher than 0 when playing against AllD (Fig. 3.3A and B).

Slow TfT's payoff patterns are very similar to the Bayesian agent. The biggest differences are when the opponent is HalfC or TfT. When playing against HalfC, the Bayesian agent's payoff is higher than Slow TfT by 0.009. Slow TfT adjusts its WTR upward about 59% of time (when $0.5 < \hat{\lambda} < 1$) and downward about 41% of time (when $0 < \hat{\lambda} < 0.5$) and adjusts upward more quickly than downward, so its average WTR is significantly higher than 0.5 (Fig. 3.3D). The Bayesian agent has a better estimate of HalfC's WTR, although its average WTR is still slightly higher than 0.5 (Fig. 3.3C), because (a) its estimate of HalfC's WTR is positively biased, likely due to the fact that there are more games with $\hat{\lambda} < 0.5$ than with $\hat{\lambda} > 0.5$, and (b) the Bayesian agent is slightly generous due to the 55th-percentile rule.

When playing against TfT, Slow TfT's payoff is higher than the Bayesian agent by 0.008. This is because occasionally the Bayesian agent's WTR is slightly lower than 1, and very rarely this induces TfT to start defecting.

Low-mutation limit

The relative evolutionary viability of different agents is reflected in the average distribution of the agent types over long time periods (Fig. 3.2B, light gray bars). Under the low-mutation limit, the Bayesian agent is the best-performing agent, taking up 33% of the population, closely followed by TfT (27%). Even though the Bayesian agent does not cooperate perfectly with itself, it is less invadable by AllC than TfT is because its average WTR toward AllC is slightly less than 1, which makes it slightly more stable than TfT. Nevertheless, the Bayesian agent's advantage is not decisive, and TfT's simplicity suggests that graded WTR inference is not very likely to evolve in this environment.

Apart from these two agents, AllD (11%), HalfC (7%), Generous TfT (8%) and Slow TfT (13%) perform relatively well. Like in the iterated Prisoner's Dilemma, AllD takes up a significant share of the population whenever there are unconditional cooperators (AllC and HalfC in our case), who can invade reciprocators through neutral drift and are then quickly replaced by AllD. HalfC's relatively good performance compared to AllC is due to its stronger ability both to invade Generous TfT and Slow TfT (and, to a lesser extent, the Bayesian agent) and to resist invasion by AllD. In turn, vulnerability to invasion by HalfC makes Generous TfT and Slow TfT perform worse than TfT and the Bayesian agent.

Effects of higher mutation rates

The results of three independent runs of the agent-based simulation based on a higher mutation rate of $u = 0.01$ are shown in Fig. 3.2B (dark gray bars). The general pattern of the distributions matches the stationary distribution under the low-mutation limit. The biggest differences are HalfC, whose proportion is significantly increased, and TfT, whose proportion is significantly decreased.

To explain this, let us consider the invasion cycle AllD → TfT → Slow TfT → HalfC → AllD (Generous TfT is similar to Slow TfT). When the mutation rate is higher, it is more likely for a third type of agent, C, to be present in the population when type B is trying to invade type A. The transition probabilities of TfT → Slow TfT and Slow TfT → HalfC remain similar to the low-mutation limit, because the relative payoffs between the mutant and the incumbent do not change much with or without a third type of agent. For instance, when Slow TfT is trying to invade TfT, the addition of an AllD agent only increases the difference between TfT's (total) payoff and Slow TfT's payoff by 0.002

(0.279 vs. 0.277). However, the transition probabilities of AllD → TfT and HalfC → AllD become much smaller than under the low-mutation limit. For instance, when AllD is trying to invade HalfC, the addition of a TfT agent increases the difference between HalfC's payoff and AllD's payoff by 0.084 (0.373 vs. 0.289). This results in a higher HalfC proportion and a lower TfT proportion.

The Bayesian agent's proportion is also decreased for a similar reason as TfT, except that it is more likely to be directly invaded by HalfC than first by Slow TfT.

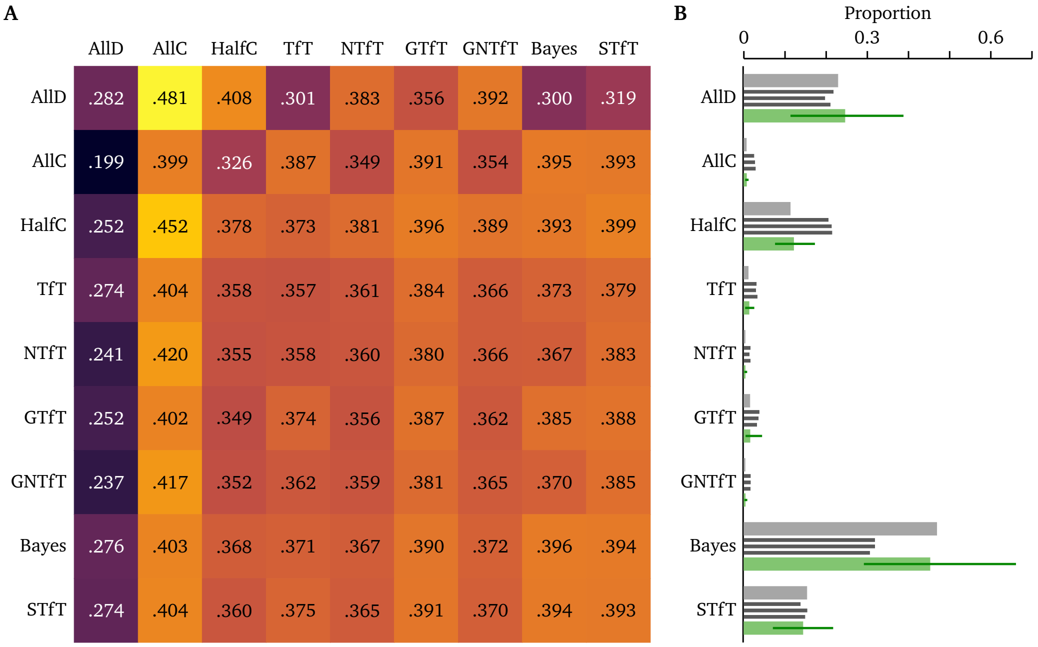
Robustness in hyperparameters

Each agent's proportion in the stationary distribution varies within the 200 sets of hyperparameters we sample. The 10th percentile, median and 90th percentile of each agent's proportion are shown in Fig. 3.2B (green bars and ranges). Out of the 200 distributions, the Bayesian agent and TfT rank first in 65% and 36% of them, respectively, suggesting that the results are robust to hyperparameter variations.

3.3.2 Experiment 2

In Experiment 1, where there is no noise in the game environment, we find that the Bayesian agent—an agent with graded WTR inference and reciprocity—performs the best, but TfT, which can be implemented as a heuristic rule, also performs well. Therefore, it is unlikely that the costly computation of graded WTR inference can evolve in such an environment.

As discussed above, in Experiment 2 we add noise to the payoffs perceived by the observing player, to simulate realistic uncertainty regarding what other people expect to be costly or valuable. We add noise only to the observer's perception of the payoffs based on the assumption that the actor usually has more accurate information about their own expected payoffs than the observer. The Bayesian agent incorporates beliefs about this error in perceiving the payoffs when inferring the actor's WTR, in effect recognizing its



	AllD	AllC	HalfC	TfT	NTfT	GTfT	GNTfT	Bayes	STfT
AllD	.282	.481	.408	.301	.383	.356	.392	.300	.319
AllC	.199	.399	.326	.387	.349	.391	.354	.395	.393
HalfC	.252	.452	.378	.373	.381	.396	.389	.393	.399
TfT	.274	.404	.358	.357	.361	.384	.366	.373	.379
NTfT	.241	.420	.355	.358	.360	.380	.366	.367	.383
GTfT	.252	.402	.349	.374	.356	.387	.362	.385	.388
GNTfT	.237	.417	.352	.362	.359	.381	.365	.370	.385
Bayes	.276	.403	.368	.371	.367	.390	.372	.396	.394
STfT	.274	.404	.360	.375	.365	.391	.370	.394	.393

Figure 3.4. Main results in Experiment 2. (**A**) The pairwise mean payoff matrix in the tournament. Each cell is the payoff received by the row player when playing against the column player. (**B**) The evolutionary distributions of the agents. The thick light gray bars are the stationary distribution under the low-mutation limit. The thin dark gray bars are the average distributions from three replications of the agent-based simulation. The green bars and ranges show the variability (10th percentile, median and 90th percentile) of the stationary distribution under the low-mutation limit with respect to different sets of hyperparameters.

own uncertainty about tradeoffs faced by the actor.

After implementing this noise in perceived payoffs, we repeated the same pairwise tournaments and evolutionary analyses employed in Experiment 1. The main results are shown in Fig. 3.4 and some notable behavioral traces are shown in Fig. 3.5.

Payoff matrix

From the pairwise mean payoff matrix (Fig. 3.2B), we see that AllD and the Bayesian agent are still the only (strict) evolutionary stable strategies. AllC still cooperates perfectly with itself, but this is no longer the case for TfT, Generous TfT and Slow TfT. TfT suffers from prolonged mutual defection due to misperception (Fig. 3.5B). Generous TfT recovers from mistakes more quickly (Fig. 3.5C), but performs much worse

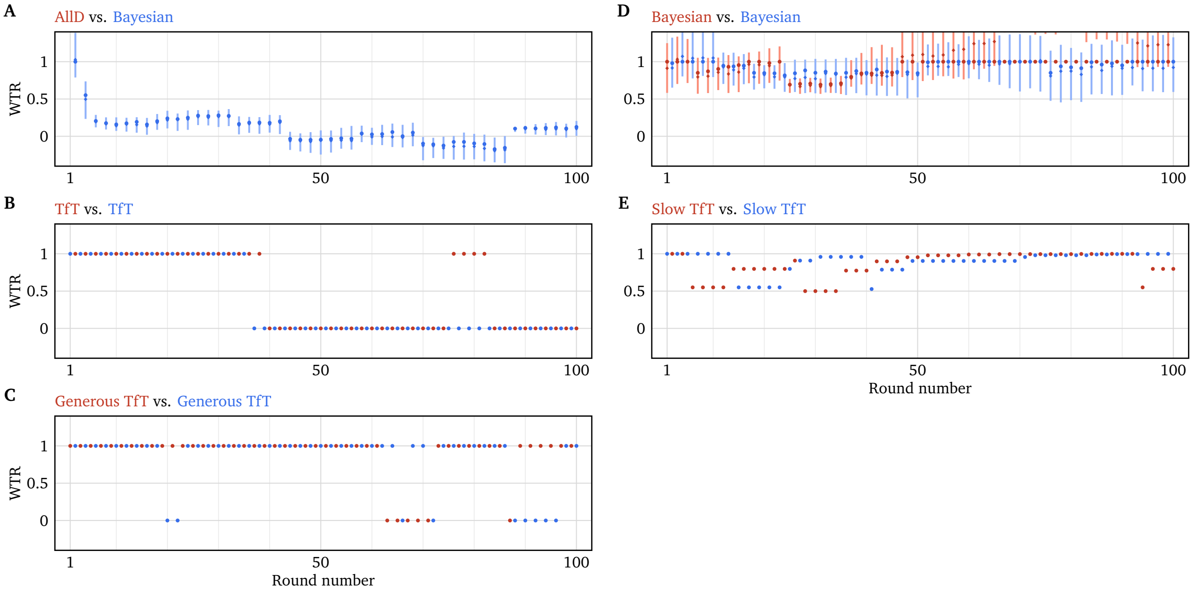


Figure 3.5. Notable behavioral traces with respect to WTR in Experiment 2. For AllD, AllC and HalfC, no trace is shown because they have constant WTRs. For TfT variants, each circle dot is the agent's WTR when making a decision. For the Bayesian agent, in each round, the vertical line is the interquartile range of the agent's current posterior over the opponent's WTR, the rhombus is the median of the posterior, and the circle dot is the WTR that the agent actually uses when making the decision.

against AllD compared to TfT or the Bayesian agent. Slow TfT is the best at cooperating with itself out of all the TfT variants because when it mistakes an act of cooperation as a defection, it does not adjust its WTR all the way to 0, which both reduces the costs of mutual defection and leaves the door open for quick recovery (Fig. 3.5E). It is also more resistant to invasion by AllD than Generous TfT with a behavioral trace similar to Fig. 3.3B. There is no neutrally stable agent in this environment except AllD and the Bayesian agent.

The Bayesian agent is both the second best (second to AllC) at cooperating with itself (Fig. 3.5D) and the best at resisting invasion by AllD (Fig. 3.5A), even slightly better than TfT, suggesting its ability to effectively deal with noise in the environment. Slow TfT's payoffs are very similar to the Bayesian agent. The biggest difference is when the opponent is HalfC—the Bayesian agent's average payoff is higher than Slow TfT by 0.007—for the same reason as in Experiment 1 (Fig. 3.3C and D).

Interestingly, while Generous TfT worked well in the iterated Prisoner's Dilemma (Nowak, 1990; Nowak & Sigmund, 1994), here it is almost entirely dominated by HalfC, the Bayesian agent and Slow TfT, the three agents who can have WTRs between 0 and 1. This highlights the value of having intermediate WTRs in a noisy environment.

Low-mutation limit

In the stationary distribution under the low-mutation limit (Fig. 3.4B, light gray bars), the Bayesian agent (47%) is the best-performing agent, and substantially outperforms the second place agent, AllD (23%). AllD performs better than it did in Experiment 1 because (a) the Bayesian agent is slightly more easily invaded by AllC and HalfC, which then give way to AllD, and (b) there is one fewer agent (TfT) who can both invade AllD and remain relatively stable against AllC and HalfC.

Compared to Experiment 1, where there was no noise, TfT's performance degrades considerably, while HalfC remains strong. Generous TfT also performs badly because its binary conception of cooperation and defection makes it sometimes too generous and sometimes too selfish.

Effects of higher mutation rates

The results of three independent runs of the agent-based simulation based on a higher mutation rate of $u = 0.01$ are shown in Fig. 3.4B (dark gray bars). The general pattern of the distributions matches the stationary distribution under the low-mutation limit. The biggest differences are HalfC, whose proportion is significantly increased, and the Bayesian agent, whose proportion is significantly decreased, with an explanation similar to Experiment 1.

Robustness in hyperparameters

Each agent's proportion in the stationary distribution varies within the 200 sets of hyperparameters we sample. The 10th percentile, median and 90th percentile of each agent's proportion are shown in Fig. 3.4B (green bars and ranges). Out of the 200 distributions, the Bayesian agent and AllD rank first in 80% and 19% of them, respectively, suggesting that the results are robust to hyperparameter variations.

3.4 Discussion

In this work we studied the evolutionary viability of agents that take different approaches to reciprocity in the context of alternating, independent decisions with changing payoffs. One type of agent based reciprocity on whether its opponent had previously selected the best option for the agent. Another type of agent based reciprocity on a binary classification of whether its opponent cooperated (i.e., made a decision that maximized their joint payoff) or defected (i.e., took a higher payoff for itself over the best option for the pair). A third type of agent based reciprocity on an inference of its opponent's graded WTR toward itself, adjusting to adopt a similar graded WTR in return. This final strategy was inspired by evidence that humans employ graded WTRs in their own social inferences and decisions (Delton & Robertson, 2016; Lim, 2012; Qi et al., under review; Quillien et al., 2023).

We showed that agents who base reciprocity on graded WTR inference (the Bayesian agent) are evolutionarily successful in game environments with or without noise. In a noiseless environment, agents that base reciprocity on simpler, binary conception of cooperation and defection (TfT) perform about as well, suggesting that costly WTR inference is unlikely to evolve in this environment. In a more realistic noisy environment where the actor and observer perceive slightly different payoffs, it is more important to have graded representations of WTRs and the Bayesian agent has a decisive advantage. This suggests that particular kinds of noise (not just "trembling hand" noise, for instance) in the environment might have driven the evolution of graded representations and adjustments of WTRs, and, more generally, more sophisticated mental representations and inference mechanisms.

3.4.1 Strategic concerns and "true WTRs"

We have assumed that the agents maximize their immediate utility in each decision based on a WTR value. This is the probably the most "proximate" conceptualization of WTRs. Since WTRs can change rapidly, the utility function also can change rapidly. We believe that such a conceptualization is necessary in order to explain how people make decisions and interpret others' decisions in varying circumstances, and it is also supported by empirical evidence that people do represent proximate WTRs (Delton et al., 2023; Lim, 2012).

Alternatively, WTRs could refer to a deeper concern about other people's welfare, which might be called "true WTRs". True WTRs are probably more stable, and it is even possible to assume that deep down everyone is entirely selfish, or except toward close relatives with shared genes (Dawkins, 1976; Hamilton, 1964). If we define a person's utility in terms of her true WTR, she might have strategic concerns when interacting repeatedly with another person. For instance, she might punish a defector in the hope of inducing him to be more cooperative in the future, thus increasing her overall material payoff. Although true WTRs are conceptually intuitive, it is hard to pin down what exactly they refer to and what role they play in determining people's behavior. In evolutionary game theory, it is not immediately clear whether true WTRs are useful variables to consider because strategies can be defined without reference to an overall utility that the strategies serve to maximize, although they might be useful when more complex environments and more sophisticated agents are under consideration.

3.4.2 Good performance for fixed, moderate cooperation

One of our agents, HalfC, had a fixed WTR of 0.5, meaning that it cooperated when the benefit to its opponent was substantially higher than its own cost but not when the benefit to the opponent was only a little higher than its own cost. This type of agent, with a fixed WTR between 0 and 1, cannot be defined in the iterated Prisoner's Dilemma. In the iterated Prisoner's Dilemma, we can only define a mixed strategy who randomly cooperates half of the time, but its behavior is closer to Naïve TfT than HalfC in our variable games. Interestingly, HalfC performs relatively well in both the noiseless and noisy environments, especially when the mutation rate is non-negligible. This further highlights the limitation of fixed games in studying interesting and plausible strategies.

HalfC's good performance depends on the existence of reciprocal agents who can invade a fixed defector (AllD). Along with the relatively good performance of AllD, our results help explain why different people have different levels of overall generosity toward others and also different tendency to reciprocate (Lim, 2012; Qi et al., under review; Van Lange et al., 1997). Future work can explore a larger variety of agents who vary continuously in these two aspects.

3.4.3 Complexity and optimality of the Bayesian agent

The Bayesian agent is much more complex computationally than other agents we consider, and uses an optimal Bayesian inference scheme with correct assumptions about the distribution of payoffs and noises in the environment (although it does so approximately). If the Bayesian agent suffers a reduction in payoffs from the computational costs or the non-optimality of the computations, its performance likely deteriorates. Therefore, although this work constitutes a first step toward identifying the evolutionary pressures toward more sophisticated mental representations and processes, more work is needed to show that graded WTR inference and reciprocity are beneficial in a wider variety of situations among a wider variety of simpler alternative agents.

3.4.4 Reducing uncertainty through emotion and language

Like other forms of uncertainty, uncertainty about whether social partners share one's own perception of potential payoffs makes it harder for agents to establish cooperative relationships while punishing defectors (compare, e.g., TfT, Generous TfT and the Bayesian agent in Experiment 2). Emotion and language are powerful ways to reduce such uncertainty. For instance, Alice's expression of anger toward Bob signals her perception of Bob's low WTR toward her in his decision (Sell et al., 2017), which can be assuaged through a verbal explanation from Bob that clarifies his own, differing perception of the payoffs involved in that decision. Thus verbal and non-verbal communication can provide other routes, beyond graded and probabilistic WTR inference, to handle the difficulty of social inference. However, emotion can be faked and language can be lies, so the ability to represent uncertainty in payoffs is still advantageous, even when accompanied by these channels of communication. Given the variety of emotions people express and the variety of functions that language serves, it is likely that their evolution is also driven by other factors.

Acknowledgements

Chapter 3, in full, is currently being prepared for submission for publication of the material. "Qi, W., Wang, B., & Powell, L. J. (in preparation). The joint evolution of theory of mind and reciprocity in noisy games." The dissertation author was the primary investigator and author of this paper.

Appendices

3.A Details of the Bayesian agent

The Bayesian agent's model of its opponent is a hidden Markov model in which the states are $\lambda_{\text{opp}}^{(i)}$, the transition probability density is $p\big(\lambda_{\text{opp}}^{(i+1)} \mid \lambda_{\text{opp}}^{(i)}\big)$, and the emission probability mass is $P\big(x^{(i)} \mid \lambda_{\text{opp}}^{(i)}, \boldsymbol{w}^{(i)}\big)$, where $x^{(i)}$ is the opponent's choice in his i-th decision and $\boldsymbol{w}^{(i)}$ is the four payoff values that the Bayesian agent observes. For simplicity of notation, we will omit the superscript (i) and subscript "opp" when there is no ambiguity.

We set the prior distribution $p(\lambda^{(0)})$ to be $N(1, 0.5^2)$.

We set the transition probability to be a mixture of two normal distributions centered on the previous state:

$$\lambda^{(i+1)} \mid \lambda^{(i)} \sim \begin{cases} N\big(\lambda^{(i)}, \sigma_{\text{small}}\big) & l = 0 \\ N\big(\lambda^{(i)}, \sigma_{\text{large}}\big) & l = 1 \end{cases}, \tag{3.3}$$

$$l \sim \text{Bernoulli}(p_{\text{large}}), \tag{3.4}$$

where p_{large} is the probability that the sample comes from the normal distribution with the larger standard deviation σ_{large}. This allows the Bayesian agent to both accommodate rapid changes in the opponent's WTR and effectively accumulate evidence about a fixed

WTR across games. We set $\sigma_{\text{small}} = 0.05$, $\sigma_{\text{large}} = 0.5$, and $p_{\text{large}} = 0.05$ by default.

In a noiseless environment (Experiment 1), the emission probability is either 0 or 1, depending on whether λ is less or greater than $\hat{\lambda}$, the critical WTR of that game. Formally, using the parameterization in Fig. 3.1D,

$$P(x = \mathrm{A} \mid \lambda, \boldsymbol{w}) = \begin{cases} 1 & w_1 + \lambda \cdot w_2 > w_3 + \lambda \cdot w_4 \\ 0 & \text{otherwise} \end{cases},$$

$$P(x = \mathrm{B} \mid \lambda, \boldsymbol{w}) = 1 - P(x = \mathrm{A} \mid \lambda).$$

In a noisy environment (Experiment 2), the emission probability is influenced by noise in the payoffs. Let $\boldsymbol{e} = e_{\{1,2,3,4\}}$ be the noise values (i.i.d. sampled from $N(0, \sigma^2)$) added to the payoffs that the opponent perceives (the actual payoffs of the game, i.i.d. sampled from $N(0, 1)$). We have

$$\begin{aligned} P(x = \mathrm{A} \mid \lambda, \boldsymbol{w}) &= \int_{\boldsymbol{e}} p(x = \mathrm{A}, \boldsymbol{e} \mid \lambda, \boldsymbol{w}) \mathrm{d}\boldsymbol{e} \\ &= \int_{\boldsymbol{e}} P(x = \mathrm{A} \mid \lambda, \boldsymbol{w}, \boldsymbol{e}) p(\boldsymbol{e} \mid \boldsymbol{w}) \mathrm{d}\boldsymbol{e} \\ &= \int_{\boldsymbol{e}} P\big(w_1 - e_1 + \lambda(w_2 - e_2) > w_3 - e_3 + \lambda(w_4 - e_4)\big) p(\boldsymbol{e} \mid \boldsymbol{w}) \mathrm{d}\boldsymbol{e} \\ &= \int_{\boldsymbol{e}} P\big(e_1 - e_3 + \lambda(e_2 - e_4) < w_1 - w_3 + \lambda(w_2 - w_4)\big) p(\boldsymbol{e} \mid \boldsymbol{w}) \mathrm{d}\boldsymbol{e}. \end{aligned}$$

To calculate this probability, we need the conditional distribution $p\big(e_1 - e_3 + \lambda(e_2 - e_4) \mid \boldsymbol{w}\big)$, which can be derived by considering the random vector $\boldsymbol{u} = \big(e_1 - e_3 + \lambda(e_2 - e_4), w_1, w_2, w_3, w_4\big)$, which has a multivariate normal distribution with mean $\mathbf{0}$ and covariance matrix

$$\begin{bmatrix} 2(1+\lambda^2)\sigma^2 & \sigma^2 & \lambda\sigma^2 & -\sigma^2 & -\lambda\sigma^2 \\ \sigma^2 & 1+\sigma^2 & 0 & 0 & 0 \\ \lambda\sigma^2 & 0 & 1+\sigma^2 & 0 & 0 \\ -\sigma^2 & 0 & 0 & 1+\sigma^2 & 0 \\ -\lambda\sigma^2 & 0 & 0 & 0 & 1+\sigma^2 \end{bmatrix}.$$

By partitioning $\boldsymbol{u}$ into $\boldsymbol{u}_1 = \big(e_1 - e_3 + \lambda(e_2 - e_4)\big)$ and $\boldsymbol{u}_2 = (w_1, w_2, w_3, w_4)$, we can derive the conditional distribution $p(\boldsymbol{u}_1 \mid \boldsymbol{u}_2)$ as (Eaton, 1983)

$$e_1 - e_3 + \lambda(e_2 - e_4) \mid \boldsymbol{w} \sim N\left(\frac{\big(w_1 - w_3 + \lambda(w_2 - w_4)\big)\sigma^2}{1+\sigma^2}, \frac{2(1+\lambda^2)\sigma^2}{1+\sigma^2}\right).$$

Then we have

$$\begin{aligned} P\big(x = \mathrm{A} \mid \lambda, \boldsymbol{w}\big) &= \Phi\left(\frac{w_1 - w_3 + \lambda(w_2 - w_4) - \frac{(w_1 - w_3 + \lambda(w_2 - w_4))\sigma^2}{1+\sigma^2}}{\sqrt{\frac{2(1+\lambda^2)\sigma^2}{1+\sigma^2}}}\right) \\ &= \Phi\left(\frac{w_1 - w_3 + \lambda(w_2 - w_4)}{\sqrt{2(1+\lambda^2)\sigma^2(1+\sigma^2)}}\right), \end{aligned}$$

where Φ is the cumulative distribution function of the standard normal distribution.

We use a standard particle filtering algorithm to approximate the posterior distribution after each observation $p\big(\lambda^{(i+1)} \mid x^{(1:i)}\big)$ (Arulampalam et al., 2002). We use the transition prior $p\big(\lambda^{(i+1)} \mid \lambda^{(i)}\big)$ as the proposal distribution. We resample before sampling from the proposal distribution if the effective sample size is below $N/2$, where N is the number of particles, which we set to 100 by default.

3.B Hyperparameter perturbations

Table 3.1 lists the 14 hyperparameters in the simulations. We perturb each hyperparameter independently according to one of three types of distributions—Bernoulli, lognormal, or beta—depending on the type and range of the hyperparameter.

The Bernoulli distribution only applies to IncludeAgents, a boolean vector of length 9 that specifies which of the 9 types of agents are considered as possible mutants. In the main results, all 9 types of agents are included. When perturbing this hyperparameter, the 9 elements of the vector are i.i.d. sampled from Bernoulli(0.9), meaning that each agent has a 90% probability of being included.

The lognormal distribution applies to parameters with a lower bound of 0 and without an upper bound. If the default value of the parameter is x^*, the distribution is $\text{lognormal}(\log x^*, 0.3)$, parameterized by the mean and standard deviation of the normal distribution on the log scale, so the median of the distribution is x^*. When $x^* = 1$, the standard deviation of the distribution is about 0.32. If the parameter is an integer, it is rounded from the real number sampled from the distribution (same for beta below).

The beta distribution applies to parameters with both a lower bound $x_{\min}$ and an upper bound $x_{\max}$. We use $\text{beta}\langle x_{\min}, x_{\max}\rangle$ to denote such a distribution. If the default value of parameter x is x^*, the distribution is

$$
\begin{aligned}
&x = x_{\min} + x'(x_{\max} - x_{\min}),\\
&x' \sim \text{beta}(\alpha, \beta),\\
&\alpha = 50\frac{x^* - x_{\min}}{x_{\max} - x_{\min}},\\
&\beta = 50\frac{x_{\max} - x^*}{x_{\max} - x_{\min}},
\end{aligned}
$$

so the mean of the distribution is x^*. When $x^* = (x_{\min} + x_{\max})/2$, the standard deviation of the distribution is about $0.07(x_{\max} - x_{\min})$.

Table 3.1. The hyperparameters in the simulation, their default values, and the distributions from which their perturbations are sampled.

Name	Description	Type	Default value	Distribution
IncludeAgents	Whether to consider each agent as a possible mutant	boolean vector	NA	Bernoulli
GameNRounds	Number of rounds in the repeated game	integer	200	lognormal
NoiseSD	Standard deviation of noise values added to payoffs in Experiment 2	real	0.1	lognormal
GTfTGenerousProb	Generous TfT's probability to unconditionally cooperate in each game	real	0.1	$\text{beta}\langle 0, 1\rangle$
BayesianPriorSD	Standard deviation of the Bayesian agent's prior distribution over λ_{opp}	real	0.5	lognormal
BayesianNParticles	Number of particles in the Bayesian agent's particle filter algorithm	integer	100	lognormal
BayesianSmallSD	σ_{small} in Eq. (3.3)	real	0.05	lognormal
BayesianLargeSD	σ_{large} in Eq. (3.3)	real	0.5	lognormal
BayesianLargeProb	p_{large} in Eq. (3.4)	real	0.05	$\text{beta}\langle 0, 0.5\rangle$
BayesianQuantile	The quantile of the posterior distribution over λ_{opp} that the Bayesian agent uses for making a decision	real	0.55	$\text{beta}\langle 0.5, 1\rangle$
STfTChangeRateMean	Slow TfT's average change rate between upwards and downwards	real	0.5	$\text{beta}\langle 0, 1\rangle$
STfTChangeRateDiff	Slow TfT's difference in change rates between upwards and downwards	real	0.1	$\text{beta}\langle 0, 0.5\rangle$
PopSize	Population size in the evolutionary dynamics	integer	100	lognormal
Softmax	Selection strength in the evolutionary dynamics (α in Eq. (3.2))	real	5	$\text{beta}\langle 0, 10\rangle$

Chapter 4

Future directions

In the three chapters of this dissertation, I have created a measure of WTRs that supports fine-grained investigation of the dynamics of WTRs, and built evolutionary models based on variable games to explain the evolution of the capacity of WTR inference and adjustment. They raise some important questions to be addressed by future research.

The Lambda Slider developed in Chapter 1 makes it easier to study the dynamics of people's WTRs over time or space (i.e., different social partners). One interesting area of investigation is the fine-grained patterns of people's reciprocity in terms of WTRs over multiple rounds of interactions. Are some participants better or faster at converging on the same WTR they observe in their partner? Beyond responding to others' WTRs, people may also be able to perceive the degree of reciprocal adjustment displayed by their partners. If so, do people behave differently when interacting with a partner who has a fixed WTR and does not reciprocate, versus a partner who does reciprocate? In the former situation, do people realize the non-reciprocity of the partner and choose to be selfish even when the partner is generous? In the latter situation, what kind and strength of the partner's reciprocity induce the highest level of mutual cooperation? To answer these questions, we can let participants and a computer agent (who might pretend to be a human) take turns making decisions on the Lambda Slider and manipulate the strategy of computer agent. As mentioned in the Discussion of Chapter 1, informing participants of the computer agent's WTRs might be better achieved through a 2D presentation of the Lambda Slider.

Apart from WTRs, other social motivations such as inequity aversion (Fehr & Schmidt, 1999) and social norms (Fehr & Fischbacher, 2004) play important roles in people's social decisions. The influence of inequity aversion has been preliminarily explored in Experiment 3 of Chapter 1, but a fuller characterization of how different social motivations influence people's decisions in different contexts is necessary. An interesting challenge is building models that can simultaneously explain people's decisions in different versions of simple behavioral games, such as the dictator game, the ultimatum game, and the public goods game (Camerer, 2011). Such models might need to take into account the nonlinearity effects of subjective utilities (Kahneman & Tversky, 1979), and should be tested on novel predictions to combat overfitting. This line of research is likely to benefit from better measures that can tease apart different motivations and are theoretically rigorous, to which the Lambda Slider is only a starting point. It is also ideally accompanied by evolutionary models that take these motivations into account and demonstrate under what conditions they can evolve.

The data from Chapter 1 shows, consistent with social discounting theory (Jones & Rachlin, 2006), that people have higher WTRs toward those who are closer to them in terms of social distance and lower WTRs toward those who are more distant. This results in a negative gradient in WTRs as a function of social distance. Evolutionarily, this might be due to a higher probability of future interactions with people who are closer to you. This hypothesis can be tested by formal evolutionary modeling, probably using a network social structure (Lieberman et al., 2005). In addition, from unreported data analyses in Chapter 1, there seems to be substantial variation in this gradient across people; i.e., some people are more "parochial" than others. This phenomenon can be studied in more detail, probably using the Lambda Slider and independent variables other than the social distance ranking, such as interaction frequency, genetic relatedness, utilitarian cooperation and competition, etc. Evolutionary models can include agents with different levels of parochialism and test if they can coevolve.

Chapters 2 and 3 in this dissertation present a first step toward using more complex and realistic environments to explain the evolution of more sophisticated mental representations and processes, both in social and non-social domains. In non-social settings, where an organism's fitness does not depend on other organisms in the environment, such models are less interesting because evolution has a clear direction toward resource rationality (Lieder & Griffiths, 2020), or maximizing fitness given particular computational requirements of the environment. As long as a cognitive ability improves the rewards of an organism after subtracting the computational costs, it can naturally evolve. In social settings, where an organism's fitness depends on other organisms in the environment and also the social structure, detailed modeling of the environment and the cognitive abilities can lead to unexpected findings, such as the finding in Chapter 3 that graded WTR inference is not very useful in a noiseless environment even when the payoffs are variable. An interesting future direction is understanding what game environments and social structures allow other kinds of social motivations, like inequity aversion and social norms, to evolve.

In Chapter 3, I mentioned that having a WTR greater than 1 or less than 0 is almost never beneficial evolutionarily. However, previous research (e.g., Duntley & Buss, 2011; Hrdy, 2009) and data in Chapter 1 show that people often exhibit WTRs greater than 1 or less than 0. How to resolve this discrepancy? There are multiple possible explanations that apply to different relationships (kin, mates, etc.), but here I will focus on one that applies to genetically unrelated individuals. In the environments studied in Chapters 2 and 3, the players' actions do not change future payoff structures, but in the real world this often happens. For instance, if I spend a lot of resources to teach another person a useful skill, although it does not improve their welfare immediately, it might significantly increase their ability to create future benefits, both for themselves and for me. As another extreme example, if two animals are competing for a finite resource, killing the competitor through an extremely low WTR will significantly increase one's future payoffs. These situations might be reduced to a WTR between 0 and 1 over long time horizons (in the killing example, one could say that the ultimate goal is completely selfish), but when manifested through WTRs that affect each decision, extreme WTRs are psychologically reasonable. Such dependencies between current actions and future payoffs have been explored using stochastic games (Hilbe et al., 2018), but more work is needed to establish what ecologically valid dependency structures allow the evolution of extreme WTR values.

References

Ackermann, K. A., Fleiß, J., & Murphy, R. O. (2016). Reciprocity as an individual difference. *Journal of Conflict Resolution*, *60*(2), 340–367. https://doi.org/10.1177/0022002714541854

Almlund, M., Duckworth, A. L., Heckman, J., & Kautz, T. (2011). Personality psychology and economics. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (pp. 1–181, Vol. 4). Elsevier. https://doi.org/10.1016/B978-0-444-53444-6.00001-8

Arulampalam, M., Maskell, S., Gordon, N., & Clapp, T. (2002). A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, *50*(2), 174–188. https://doi.org/10.1109/78.978374

Axelrod, R., & Hamilton, W. D. (1981). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. https://doi.org/10.1126/science.7466396

Axelrod, R. (1984). *The evolution of cooperation*. Basic Books.

Baron-Cohen, S. (2000). The evolution of a theory of mind. In M. Corballis & S. E. G. Lea (Eds.), *The descent of mind: Psychological perspectives on hominid evolution* (pp. 261–277). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780192632593.003.0013

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic child have a "theory of mind"? *Cognition*, *21*(1), 37–46. https://doi.org/10.1016/0010-0277(85)90022-8

Baumeister, R. F., & Masicampo, E. J. (2010). Conscious thought is for facilitating social and cultural interactions: How mental simulations serve the animal–culture interface. *Psychological Review*, *117*(3), 945–971. https://doi.org/10.1037/a0019393

Bishop, D. T., & Cannings, C. (1978). A generalized war of attrition. *Journal of Theoretical Biology*, *70*(1), 85–124. https://doi.org/10.1016/0022-5193(78)90304-1

Bostyn, D. H., Sevenhant, S., & Roets, A. (2018). Of mice, men, and trolleys: Hypothetical judgment versus real-life behavior in trolley-style moral dilemmas. *Psychological Science*, *29*(7), 1084–1093. https://doi.org/10.1177/0956797617752640

Brüne, M., & Brüne-Cohrs, U. (2006). Theory of mind—evolution, ontogeny, brain mechanisms and psychopathology. *Neuroscience & Biobehavioral Reviews*, *30*(4), 437–455. https://doi.org/10.1016/j.neubiorev.2005.08.001

Bürkner, P.-C. (2017). Brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01

Bürkner, P.-C., & Charpentier, E. (2020). Modelling monotonic effects of ordinal predictors in Bayesian regression models. *British Journal of Mathematical and Statistical Psychology*, *73*(3), 420–451. https://doi.org/10.1111/bmsp.12195

Buunk, B. P., & Schaufeli, W. B. (1999). Reciprocity in interpersonal relationships: An evolutionary perspective on its importance for health and well-being. *European Review of Social Psychology*, *10*(1), 259–291. https://doi.org/10.1080/14792779943000080

Call, J., & Tomasello, M. (2008). Does the chimpanzee have a theory of mind? 30 years later. *Trends in Cognitive Sciences*, *12*(5), 187–192. https://doi.org/10.1016/j.tics.2008.02.010

Camerer, C. F. (2011). *Behavioral game theory: Experiments in strategic interaction*. Princeton University Press.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Delton, A. W., Jaeggi, A. V., Lim, J., Sznycer, D., Gurven, M., Robertson, T. E., Sugiyama, L. S., Cosmides, L., & Tooby, J. (2023). Cognitive foundations for helping and harming others: Making welfare tradeoffs in industrialized and small-scale societies. *Evolution and Human Behavior*, *44*(5), 485–501. https://doi.org/10.1016/j.evolhumbehav.2023.01.013

Delton, A. W., & Robertson, T. E. (2012). The social cognition of social foraging: Partner selection by underlying valuation. *Evolution and Human Behavior*, *33*(6), 715–725. https://doi.org/10.1016/j.evolhumbehav.2012.05.007

Delton, A. W., & Robertson, T. E. (2016). How the mind makes welfare tradeoffs: Evolution, computation, and emotion. *Current Opinion in Psychology*, *7*, 12–16. https://doi.org/10.1016/j.copsyc.2015.06.006

Duntley, J. D., & Buss, D. M. (2011). Homicide adaptations. *Aggression and Violent Behavior*, *16*(5), 399–410. https://doi.org/10.1016/j.avb.2011.04.016

Eaton, M. L. (1983). *Multivariate statistics: A vector space approach*. John Wiley & Sons, Inc.

Eisenbruch, A. B., & Krasnow, M. M. (2022). Why warmth matters more than competence: A new evolutionary approach. *Perspectives on Psychological Science*, *17*(6), 1604–1623. https://doi.org/10.1177/17456916211071087

Fehr, E., & Fischbacher, U. (2004). Social norms and human cooperation. *Trends in Cognitive Sciences*, *8*(4), 185–190. https://doi.org/10.1016/j.tics.2004.02.007

Fehr, E., Kirchsteiger, G., & Riedl, A. (1998). Gift exchange and reciprocity in competitive experimental markets. *European Economic Review*, *42*(1), 1–34. https://doi.org/10.1016/S0014-2921(96)00051-7

Fehr, E., & Schmidt, K. M. (1999). A theory of fairness, competition, and cooperation. *The Quarterly Journal of Economics*, *114*(3), 817–868. https://doi.org/10.1162/003355399556151

Fehr, E., & Schmidt, K. M. (2006). The economics of fairness, reciprocity and altruism – Experimental evidence and new theories. In S.-C. Kolm & J. M. Ythier (Eds.), *Handbook of the economics of giving, altruism and reciprocity* (pp. 615–691, Vol. 1). Elsevier. https://doi.org/10.1016/S1574-0714(06)01008-6

FeldmanHall, O., Mobbs, D., Evans, D., Hiscox, L., Navrady, L., & Dalgleish, T. (2012). What we say and what we do: The relationship between real and hypothetical moral choices. *Cognition*, *123*(3), 434–441. https://doi.org/10.1016/j.cognition.2012.02.001

Fessler, D. M. T., & Haley, K. J. (2003). The strategy of affect: Emotions in human cooperation. In P. Hammerstein (Ed.), *Genetic and cultural evolution of cooperation* (pp. 7–36). MIT Press.

Fudenberg, D., & Tirole, J. (1991). *Game theory*. MIT Press.

Gouldner, A. W. (1960). The norm of reciprocity: A preliminary statement. *American Sociological Review*, *25*(2), 161–178. https://doi.org/10.2307/2092623

Griesinger, D. W., & Livingston Jr., J. W. (1973). Toward a model of interpersonal motivation in experimental games. *Behavioral Science*, *18*(3), 173–188. https://doi.org/10.1002/bs.3830180305

Gronau, Q. F., Singmann, H., & Wagenmakers, E.-J. (2020). Bridgesampling: An R package for estimating normalizing constants. *Journal of Statistical Software*, *92*, 1–29. https://doi.org/10.18637/jss.v092.i10

Güth, W., Schmittberger, R., & Schwarze, B. (1982). An experimental analysis of ultimatum bargaining. *Journal of Economic Behavior & Organization*, *3*(4), 367–388. https://doi.org/10.1016/0167-2681(82)90011-7

Hackel, L. M., Doll, B. B., & Amodio, D. M. (2015). Instrumental learning of traits versus rewards: Dissociable neural correlates and effects on choice. *Nature Neuroscience*, *18*(9), 1233–1235. https://doi.org/10.1038/nn.4080

Hall, J., Kahn, D. T., Skoog, E., & Öberg, M. (2021). War exposure, altruism and the recalibration of welfare tradeoffs towards threatening social categories. *Journal of Experimental Social Psychology*, *94*, 104101. https://doi.org/10.1016/j.jesp.2021.104101

Hamilton, W. D. (1964). The genetical evolution of social behaviour. *Journal of Theoretical Biology*, *7*(1), 1–52. https://doi.org/10.1016/0022-5193(64)90038-4

Hauert, C., Holmes, M., & Doebeli, M. (2006). Evolutionary games and population dynamics: Maintenance of cooperation in public goods games. *Proceedings of the Royal Society B: Biological Sciences*, *273*(1600), 2565–2571. https://doi.org/10.1098/rspb.2006.3600

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., & McElreath, R. (2001). In search of homo economicus: Behavioral experiments in 15 small-scale societies. *The American Economic Review*, *91*(2), 73–78. https://doi.org/10.1257/aer.91.2.73

Hilbe, C., Šimsa, Š., Chatterjee, K., & Nowak, M. A. (2018). Evolution of cooperation in stochastic games. *Nature*, *559*(7713), 246–249. https://doi.org/10.1038/s41586-018-0277-x

Höglinger, M., & Wehrli, S. (2017). Measuring social preferences on Amazon Mechanical Turk. In B. Jann & W. Przepiorka (Eds.), *Social dilemmas, institutions, and the evolution of cooperation*. De Gruyter Oldenbourg. https://doi.org/10.1515/9783110472974-025

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

Hurwicz, L., & Reiter, S. (2006). *Designing economic mechanisms*. Cambridge University Press. https://doi.org/10.1017/CBO9780511754258

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. https://doi.org/10.1016/j.tics.2016.05.011

Johnson, M. W., & Bickel, W. K. (2002). Within-subject comparison of real and hypothetical money rewards in delay discounting. *Journal of the Experimental Analysis of Behavior*, *77*(2), 129–146. https://doi.org/10.1901/jeab.2002.77-129

Jones, B., & Rachlin, H. (2006). Social discounting. *Psychological Science*, *17*(4), 283–286. https://doi.org/10.1111/j.1467-9280.2006.01699.x

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–292. https://doi.org/10.2307/1914185

Kirkpatrick, M., Delton, A. W., Robertson, T. E., & de Wit, H. (2015). Prosocial effects of MDMA: A measure of generosity. *Journal of Psychopharmacology*, *29*(6), 661–668. https://doi.org/10.1177/0269881115573806

Kleiman-Weiner, M. (2018). *Computational foundations of human social intelligence* [Doctoral dissertation, Massachusetts Institute of Technology].

Krasnow, M. M., Delton, A. W., Cosmides, L., & Tooby, J. (2016). Looking under the hood of third-party punishment reveals design for personal benefit. *Psychological Science*, *27*(3), 405–418. https://doi.org/10.1177/0956797615624469

Krupenye, C., Kano, F., Hirata, S., Call, J., & Tomasello, M. (2016). Great apes anticipate that other individuals will act according to false beliefs. *Science*, *354*(6308), 110–114. https://doi.org/10.1126/science.aaf8110

Leider, S., Möbius, M. M., Rosenblat, T., & Do, Q.-A. (2009). Directed altruism and enforced reciprocity in social networks. *The Quarterly Journal of Economics*, *124*(4), 1815–1851. https://doi.org/10.1162/qjec.2009.124.4.1815

Lieberman, E., Hauert, C., & Nowak, M. A. (2005). Evolutionary dynamics on graphs. *Nature*, *433*(7023), 312–316. https://doi.org/10.1038/nature03204

Liebrand, W. B. G. (1984). The effect of social motives, communication and group size on behaviour in an N-person multi-stage mixed-motive game. *European Journal of Social Psychology*, *14*(3), 239–264. https://doi.org/10.1002/ejsp.2420140302

Liebrand, W. B. G., & McClintock, C. G. (1988). The ring measure of social values: A computerized procedure for assessing individual differences in information processing and social value orientation. *European Journal of Personality*, *2*(3), 217–230. https://doi.org/10.1002/per.2410020304

Lieder, F., & Griffiths, T. L. (2020). Resource-rational analysis: Understanding human cognition as the optimal use of limited computational resources. *Behavioral and Brain Sciences*, *43*, e1. https://doi.org/10.1017/S0140525X1900061X

Lim, J. (2012). *Welfare tradeoff ratios and emotions: Psychological foundations of human reciprocity* [Doctoral dissertation, University of California, Santa Barbara].

Liu, D., Wellman, H. M., Tardif, T., & Sabbagh, M. A. (2008). Theory of mind development in Chinese children: A meta-analysis of false-belief understanding across cultures and languages. *Developmental Psychology*, *44*(2), 523–531. https://doi.org/10.1037/0012-1649.44.2.523

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards in self-control and social discounting. *Judgment and Decision Making*, *6*(6), 552–564. https://doi.org/10.1017/S1930297500002515

Lord, F. M., & Novick, M. R. (1968). *Statistical theories of mental test scores*. Addison-Wesley Publishing Company.

Maki, J. E., Thorngate, W. B., & McClintock, C. G. (1979). Prediction and perception of social motives. *Journal of Personality and Social Psychology*, *37*(2), 203–220. https://doi.org/10.1037/0022-3514.37.2.203

Makowski, D., Ben-Shachar, M. S., & Lüdecke, D. (2019). bayestestR: Describing effects and their uncertainty, existence and significance within the Bayesian framework. *Journal of Open Source Software*, *4*(40), 1541. https://doi.org/10.21105/joss.01541

Maynard Smith, J., & Price, G. R. (1973). The logic of animal conflict. *Nature*, *246*(5427), 15–18. https://doi.org/10.1038/246015a0

Maynard Smith, J. (1982). *Evolution and the theory of games*. Cambridge University Press.

Messick, D. M., & McClintock, C. G. (1968). Motivational bases of choice in experimental games. *Journal of Experimental Social Psychology*, *4*(1), 1–25. https://doi.org/10.1016/0022-1031(68)90046-2

Monroe, A. (2020). Moral elevation: Indications of functional integration with welfare trade-off calibration and estimation mechanisms. *Evolution and Human Behavior*, *41*(4), 293–302. https://doi.org/10.1016/j.evolhumbehav.2020.05.002

Moran, P. A. P. (1962). *The statistical processes of evolutionary theory*. Clarendon Press.

Murphy, R. O., & Ackermann, K. A. (2014). Social value orientation: Theoretical and measurement issues in the study of social preferences. *Personality and Social Psychology Review*, *18*(1), 13–41. https://doi.org/10.1177/1088868313501745

Murphy, R. O., Ackermann, K. A., & Handgraaf, M. J. J. (2011). Measuring social value orientation. *Judgment and Decision Making*, *6*(8), 771–781. https://doi.org/10.1017/S1930297500004204

Nowak, M. A. (1990). Stochastic strategies in the Prisoner's Dilemma. *Theoretical Population Biology*, *38*(1), 93–112. https://doi.org/10.1016/0040-5809(90)90005-G

Nowak, M. A. (2006a). *Evolutionary dynamics: Exploring the equations of life*. Harvard University Press.

Nowak, M. A. (2006b). Five rules for the evolution of cooperation. *Science*, *314*(5805), 1560–1563. https://doi.org/10.1126/science.1133755

Nowak, M. A., Sasaki, A., Taylor, C., & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature*, *428*(6983), 646–650. https://doi.org/10.1038/nature02414

Nowak, M. A., & Sigmund, K. (1993). A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature*, *364*(6432), 56–58. https://doi.org/10.1038/364056a0

Nowak, M. A., & Sigmund, K. (1994). The alternating Prisoner's Dilemma. *Journal of Theoretical Biology*, *168*(2), 219–226. https://doi.org/10.1006/jtbi.1994.1101

Penn, D. C., & Povinelli, D. J. (2007). On the lack of evidence that non-human animals possess anything remotely resembling a 'theory of mind'. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *362*(1480), 731–744. https://doi.org/10.1098/rstb.2006.2023

Piff, P. K., Kraus, M. W., Côté, S., Cheng, B. H., & Keltner, D. (2010). Having less, giving more: The influence of social class on prosocial behavior. *Journal of Personality and Social Psychology*, *99*(5), 771–784. https://doi.org/10.1037/a0020092

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences*, *1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Pruitt, D. G. (1967). Reward structure and cooperation: The decomposed Prisoner's Dilemma game. *Journal of Personality and Social Psychology*, *7*(1), 21–27. https://doi.org/10.1037/h0024914

Qi, W., & Vul, E. (2020). Adaptive behavior in variable games requires theory of mind. In S. Denison, M. Mack, Y. Xu, & B. C. Armstrong (Eds.), *Proceedings of the 42nd Annual Conference of the Cognitive Science Society* (pp. 1369–1375). Cognitive Science Society. https://doi.org/10.31234/osf.io/7kw4z

Qi, W., & Vul, E. (2022). The evolution of theory of mind on welfare tradeoff ratios. *Evolution and Human Behavior*, *43*(5), 381–393. https://doi.org/10.1016/j.evolhumbehav.2022.06.003

Qi, W., Vul, E., & Powell, L. J. (under review). An accurate and efficient measure of welfare tradeoff ratios. *PLOS One*. https://doi.org/10.31234/osf.io/yn8e5

Qi, W., Wang, B., & Powell, L. J. (in preparation). The joint evolution of theory of mind and reciprocity in noisy games.

Quillien, T. (2023). Rational information search in welfare-tradeoff cognition. *Cognition*, *231*, 105317. https://doi.org/10.1016/j.cognition.2022.105317

Quillien, T., Tooby, J., & Cosmides, L. (2023). Rational inferences about social valuation. *Cognition*, *239*, 105566. https://doi.org/10.1016/j.cognition.2023.105566

Rand, D. G., & Nowak, M. A. (2012). Evolutionary dynamics in finite populations can explain the full range of cooperative behaviors observed in the centipede game. *Journal of Theoretical Biology*, *300*, 212–221. https://doi.org/10.1016/j.jtbi.2012.01.011

Rapoport, A., Chammah, A. M., & Orwant, C. J. (1965). *Prisoner's Dilemma: A study in conflict and cooperation*. University of Michigan Press.

Samuelson, L., & Zhang, J. (1992). Evolutionary stability in asymmetric games. *Journal of Economic Theory*, *57*(2), 363–391. https://doi.org/10.1016/0022-0531(92)90041-F

Sell, A., Sznycer, D., Al-Shawaf, L., Lim, J., Krauss, A., Feldman, A., Rascanu, R., Sugiyama, L., Cosmides, L., & Tooby, J. (2017). The grammar of anger: Mapping the computational architecture of a recalibrational emotion. *Cognition*, *168*, 110–128. https://doi.org/10.1016/j.cognition.2017.06.002

Shapley, L. S. (1953). Stochastic games. *Proceedings of the National Academy of Sciences*, *39*(10), 1095–1100. https://doi.org/10.1073/pnas.39.10.1095

Skyrms, B. (1992). Chaos in game dynamics. *Journal of Logic, Language and Information*, *1*(2), 111–130. https://doi.org/10.1007/BF00171693

Smith, A., Pedersen, E. J., Forster, D. E., McCullough, M. E., & Lieberman, D. (2017). Cooperation: The roles of interpersonal value and gratitude. *Evolution and Human Behavior*, *38*(6), 695–703. https://doi.org/10.1016/j.evolhumbehav.2017.08.003

Sonnemans, J., Dijk, F. van, & Winden, F. van. (2006). On the dynamics of social ties structures in groups. *Journal of Economic Psychology*, *27*(2), 187–204. https://doi.org/10.1016/j.joep.2005.08.004

Stan Development Team. (2021). Stan modeling language users guide and reference manual [Version 2.26]. https://mc-stan.org/

Stan Development Team. (2023). Stan modeling language users guide and reference manual [Version 2.32.2]. https://mc-stan.org/

Stan Development Team. (2024). RStan: The R interface to Stan [R package version 2.32.6]. https://mc-stan.org/

Taylor, P. D., & Jonker, L. B. (1978). Evolutionary stable strategies and game dynamics. *Mathematical Biosciences*, *40*(1), 145–156. https://doi.org/10.1016/0025-5564(78)90077-9

Thielmann, I., Spadaro, G., & Balliet, D. (2020). Personality and prosocial behavior: A theoretical framework and meta-analysis. *Psychological bulletin*, *146*(1), 30–90. https://doi.org/10.1037/bul0000217

Thomas, M., & Kyung, E. J. (2019). Slider scale or text box: How response format shapes responses. *Journal of Consumer Research*, *45*(6), 1274–1293. https://doi.org/10.1093/jcr/ucy057

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*(5), 675–691. https://doi.org/10.1017/S0140525X05000129

Tooby, J., & Cosmides, L. (2008). The evolutionary psychology of the emotions and their relationship to internal regulatory variables. In M. Lewis, J. M. Haviland-Jones, & L. F. Barrett (Eds.), *Handbook of emotions* (3rd ed., pp. 114–137). Guilford Press.

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly Review of Biology*, *46*(1), 35–57. https://doi.org/10.1086/406755

Van Lange, P. A. M., De Bruin, E. M. N., Otten, W., & Joireman, J. A. (1997). Development of prosocial, individualistic, and competitive orientations: Theory and preliminary evidence. *Journal of Personality and Social Psychology*, *73*(4), 733–746. https://doi.org/10.1037/0022-3514.73.4.733

Watkins, C. J. C. H. (1989). *Learning from delayed rewards* [Doctoral dissertation, King's College, University of Cambridge].

Weibull, J. W. (1997). *Evolutionary game theory*. MIT Press.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128. https://doi.org/10.1016/0010-0277(83)90004-5

Wiseman, D. B., & Levin, I. P. (1996). Comparing risky decision making under conditions of real and hypothetical consequences. *Organizational Behavior and Human Decision Processes*, *66*(3), 241–250. https://doi.org/10.1006/obhd.1996.0053

Wojciszke, B., Bazinska, R., & Jaworski, M. (1998). On the dominance of moral categories in impression formation. *Personality and Social Psychology Bulletin*, *24*(12), 1251–1263. https://doi.org/10.1177/01461672982412001

Wubben, M. J. J., Cremer, D. D., & Dijk, E. van. (2009). How emotion communication guides reciprocity: Establishing cooperation through disappointment and anger. *Journal of Experimental Social Psychology*, *45*(4), 987–990. https://doi.org/10.1016/j.jesp.2009.04.010

Zeeman, E. C. (1980). Population dynamics from game theory. In Z. Nitecki & C. Robinson (Eds.), *Global theory of dynamical systems* (pp. 471–497). Springer.