



ELSEVIER

Contents lists available at ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb



Conventional contracts, intentional behavior and logit choice: Equality without symmetry [☆]

Sung-Ha Hwang ^a, Wooyoung Lim ^b, Philip Neary ^c, Jonathan Newton ^{d,*}^a Korea Advanced Institute of Science and Technology, Seoul, Republic of Korea^b The Hong Kong University of Science and Technology, Hong Kong^c Royal Holloway, University of London, England, United Kingdom^d Institute of Economic Research, Kyoto University, Japan

ARTICLE INFO

Article history:

Received 29 August 2017

Available online 23 May 2018

JEL classification:

C73

C78

Keywords:

Evolution
Nash program
Logit choice
Egalitarianism

ABSTRACT

When coordination games are played under the logit choice rule and there is intentional bias in agents' non-best response behavior, we show that the Egalitarian bargaining solution emerges as the long run social norm. Without intentional bias, a new solution, the Logit bargaining solution emerges as the long run norm. These results contrast with results under non-payoff dependent deviations from best response behavior, where it has previously been shown that the Kalai–Smorodinsky and Nash bargaining solutions emerge as long run norms. We complement the theory with experiments on human subjects, results of which suggest that non-best response play is payoff dependent and displays intentional bias. This suggests the Egalitarian solution as the most likely candidate for a long run bargaining norm.

© 2018 Elsevier Inc. All rights reserved.

1. Introduction

Consider a two player coordination game with zero-payoffs for miscoordination (*contract games*, Young, 1998a) and with payoffs on the main diagonal that correspond to points on the efficient frontier of a convex bargaining set. A *bargaining solution* is a function that maps any given bargaining set to a payoff vector within the set (see Nash, 1950; Kalai and Smorodinsky, 1975).¹ In his paper, 'Conventional Contracts', Young (1998a) showed that if populations of agents play such a game, usually updating their strategies according to a best response rule, but occasionally making a *deviation*² and playing something other than a best response, then the long run social norm that emerges approximates the Kalai and Smorodinsky (1975) bargaining solution. Subsequently, Naidu et al. (2010) showed that if deviations are *intentional* so that an agent who deviates always demands more than the best response, never less, then the Nash (1950) bargaining solution emerges as

[☆] S.-H. Hwang was supported by the Ministry of Education of the Republic of Korea and the National Research Foundation of Korea (NRF-2016S1A3A2924944). W. Lim was supported by a grant from the Research Grants Council of Hong Kong (Grant No. ECS-699613). Sincere thanks are given to Michihiro Kandori, Heinrich Nax, Bary Pradelski and Peyton Young for comments and suggestions, as well as to the advisory editor and three referees. Thanks to Xiaotong Sun for help with diagrams.

* Corresponding author.

E-mail addresses: sungha@kaist.ac.kr (S.-H. Hwang), wooyoung@ust.hk (W. Lim), Philip.Neary@rhul.ac.uk (P. Neary), newton@kier.kyoto-u.ac.jp (J. Newton).

¹ In the words of Kalai and Smorodinsky (1975), 'A solution to the bargaining problem is a function $f : U \rightarrow \mathbb{R}^2$ such that $f(a, S) \in S$.' In their notation, U is the set of pairs (a, S) , in which S is a bargaining set and a is a disagreement point.

² Deviations are also referred to in the literature as 'errors', 'mistakes', 'mutations' and 'idiosyncratic shocks'.

Table 1

Long run bargaining norms by deviation process. Each bargaining solution emerges as the long run norm under the corresponding behavioral rule and without reference to any appealing ex-post properties that the solution might have.

	Unintentional	Intentional
Uniform	Kalai–Smorodinsky <i>Young (1998a)</i>	Nash bargaining <i>Naidu et al. (2010)</i>
Logit	Logit bargaining <i>This paper</i>	Egalitarian <i>This paper</i>

the long run social norm. Newton (2012a) showed that the Nash bargaining solution also emerges when the choice rule includes some degree of collective agency – joint decision making by pairs of agents. The deviations in the cited works are *uniform* – all possible deviations are equally likely. However, there is another commonly used model of perturbed best response, the *logit* choice rule.³ Under logit choice, deviations which incur a higher payoff loss for the agent making them are less likely to be made. The recent approximation results of Hwang and Newton (2017) allow us to solve the problem of conventional contracts under logit choice.

It is shown that if the logit choice rule is used with intentional deviations, then the Egalitarian bargaining solution (Kalai, 1977) emerges as the long run social norm. Justifications of Egalitarianism have usually assumed some symmetry in the problem faced (Alexander and Skyrms, 1999) or invoked ex-ante symmetry of players with respect to their position in the game (Binmore, 1998, 2005). One contribution of the current paper is to give a model of adaptive behavior that leads to the Egalitarian solution without any symmetry assumptions beyond those on population size and uniformity in matching.

Furthermore, we introduce a new bargaining solution, the Logit bargaining solution, that maximizes an adjusted Nash product, but, in the spirit of the Kalai–Smorodinsky solution, is influenced by the best possible payoffs for the players. Unlike the other solutions, the Logit bargaining solution is not designed to satisfy any particular set of appealing properties, but is instead the solution that emerges when agents in populations follow a given behavioral rule, the logit choice rule. This highlights an important difference between the traditional approach to bargaining solutions and the evolutionary approach. The traditional approach seeks to construct bargaining solutions with appealing properties and treats these properties as axiomatic. The evolutionary approach takes the behavioral process as axiomatic and sees what bargaining solutions emerge as long run norms of such processes. Given this, it is remarkable that three of the processes in Table 1 lead to solutions already known to the literature. In this sense, perhaps the fourth sibling of this family, the Logit bargaining solution, provides a cautionary tale, for although it emerges from one of the simplest and most common choice rules in the social sciences, it displays a quirky nonmonotonicity in comparison to the other solutions. This nonmonotonicity can be clearly and intuitively explained with reference to the underlying behavioral process, a good example of how a complex social norm can be generated by simple behavior.

Of course, the importance of the implications of any behavioral rule rests to some extent on its empirical validity. To begin to address such questions we report the results of experiments conducted to test deviation behavior in the context of the model of the paper. We find evidence in favor of intentional bias and payoff dependence in non-best response play. While the constraints of time mean that we cannot test the long run behavior of the empirical process, these results, together with the theoretical results summarized in Table 1, suggest the Egalitarian solution as the most likely of our four candidates for a long run bargaining norm. Importantly, our design gives subjects no information about the payoffs that can be attained by their potential opponents, thereby ensuring that neither pre-existing norms of surplus division nor other-regarding preferences can play a role in strategy choice.

1.1. Related literature

This study is part of a literature that studies connections between evolutionary game theory and cooperative game theory. This literature has been called the *Evolutionary Nash Program* (see Newton, 2018, Section 6, for a survey), emphasizing the parallel with the original Nash Program that studies connections between noncooperative game theory and cooperative game theory (see, e.g. Nash, 1953). In addition to the contract games considered in the current paper, another well-developed strand of this literature concerns long run norms in Nash demand games (Nash, 1953), where, in contrast to the coordination games considered here, players can obtain positive payoffs from imperfect coordination.⁴

³ The logit choice rule derives from the Boltzman distribution from statistical mechanics. This was later adapted to regression analysis by Cox (1958) and given a random utility interpretation by Block and Marschak (1959). Blume (1993) then used it in modeling repeated interactions over the long run, as it is used in the current paper.

⁴ Interestingly, Nash, when describing the Nash demand game and relating it to the bargaining problem, uses language that suggests a problem in which coordination ('agreement') on some utility pair is necessary to avoid the disagreement payoffs. The demands are described as restrictions on the utility pairs on which coordination can occur, before the payoff specification abstracts from the coordination problem that remains amongst utility pairs that satisfy demands.

Compared to contract games, Nash demand games have a less severe penalty for miscoordination that involves asking for too little. This creates a bias towards transitions in which best responding players lower their demands, similar to the transitions under intentional deviations in the current paper (see Table 3).⁵ In line with this intuition, Young (1993b) shows that the Nash bargaining solution emerges under uniform-unintentional deviations. Hwang and Rey-Bellet (2017) show that this is also true for uniform-intentional deviations. To solve the problem for logit deviations, Hwang and Rey-Bellet (2017) extend the results of Hwang and Newton (2017) on most probable transition paths under logit choice. They use these results to show that, under logit-unintentional deviations, the Nash bargaining solution emerges, and that, under logit-intentional deviations, a new solution emerges that is more equal than the Nash bargaining solution. For *generalized Nash demand games* with more than two players and payoff structures given by characteristic functions similar to those of cooperative games, Agastya (1999) and Newton (2012b), the latter featuring collective agency, find selection for minmax and maxmin (Rawlsian) solutions respectively under uniform-unintentional deviations.

Another strand of the literature concerns selection in matching environments with either non-transferable utility (Newton and Sawa, 2015) or transferable utility (Nax and Pradelski, 2015; Klaus and Newton, 2016). Interestingly, although both Nax and Pradelski (2015) and Newton (2012b) find maxmin selection, these results arise in different ways. Nax and Pradelski (2015) use logit deviations. Selection then comes from (i) how hard it is for a player to make deviations. Newton (2012b) uses uniform deviations and sampling of opponents' behavior: selection comes from (ii) given an incumbent strategy (a 'convention'), the number of deviations required to induce some player to respond with a different strategy. In the current study, we have that under logit choice, when deviations are intentional, effect (ii) is dominated by effect (i). Consequently, the long run social norm is the convention at which it is least likely that any player deviates. Under logit choice, this corresponds to the payoff of the poorest player being as high as possible – the Egalitarian solution. In contrast, when deviations are unintentional, effects (i) and (ii) combine to create the interesting nonmonotonicities of the Logit bargaining solution.

Our experiments contribute to a small literature that considers non-best response behavior in laboratory data as analogous to deviations in best response dynamics. For two strategy coordination games, when interaction is determined by a network, Mäs and Nax (2016) find deviations to be payoff dependent. When interaction is uniform, Lim and Neary (2016) find likewise. Furthermore, the cited studies find that deviations are predominantly made by agents who do relatively badly at the current convention. This points towards deviations being intentional. The current study has more than two strategies and is therefore able to provide more conclusive evidence on this point, as any given agent has several possible deviations that he could make, some of which can be interpreted as intentional and others which cannot. We find that 83% of deviations in our experiments can be interpreted as intentional.

The paper is organized as follows. Section 2 defines the bargaining solutions and gives the evolutionary model. Section 3 classifies bargaining solutions by the behavioral rules which give rise to them. Section 4 discusses our experimental evidence. Section 5 concludes.

2. Model

2.1. Bargaining solutions

Consider two positions, α and β , and a closed, convex bargaining set $S \subset \mathbb{R}^2$ containing the origin. The set S gives feasible payoffs for players in the α and β positions respectively. Let the bargaining frontier, the efficient points of S , be given by a strictly decreasing, differentiable, and concave function, $f : \mathbb{R} \rightarrow \mathbb{R}$, such that $(t, f(t))$ is the efficient payoff pair in which α and β players receive t and $f(t)$, respectively. Normalizing the disagreement point to $(0, 0)$, the maximum payoff that players α and β can obtain are

$$\bar{s}_\alpha := \max \{t : f(t) \geq 0\} \quad \text{and} \quad \bar{s}_\beta := \max \{f(t) : t \geq 0\}.$$

which implies $f(\bar{s}_\alpha) = 0$ and $\bar{s}_\beta = f(0)$.

A bargaining solution maps bargaining sets to payoffs. The three bargaining solutions most commonly used in economics are the Nash bargaining solution (Nash, 1950), the Kalai–Smorodinsky bargaining solution (Kalai and Smorodinsky, 1975), and the Egalitarian bargaining solution (Kalai, 1977). These solutions uniquely satisfy distinct sets of intuitively appealing properties. The traditional approach is to treat these properties as axiomatic and to find bargaining solutions which have these properties. This is not the approach of the current paper. Rather, we focus on how solutions emerge as long run behavioral norms when agents follow simple behavioral rules when faced with coordination problems. That is, the behavior that gives rise to the solution is treated as axiomatic rather than the properties of the solution itself. Definitions of the bargaining solutions that feature in this paper are given in Table 2. The Logit bargaining solution, which is new, is analyzed further and compared to existing bargaining solutions in Section 3.1, but for now we move to define the perturbed best response rules that lead to these solutions emerging as norms.

⁵ See Binmore et al. (2003) for further discussion of differences and similarities between contract games and Nash demand games under uniform-unintentional deviations.

Table 2
Bargaining solutions for frontiers given by $f(\cdot)$. Our assumptions on $f(\cdot)$ guarantee that $t f(t) \phi(t)$ is strictly concave, so t^L is unique.

Bargaining solution	Notation	Definition
Kalai and Smorodinsky (1975)	t^{KS}	$\frac{t^{KS}}{s_\alpha} = \frac{f(t^{KS})}{s_\beta}$.
Nash (1950)	t^{NB}	$t^{NB} \in \arg \max_{0 \leq t \leq \bar{s}_\alpha} t f(t)$.
Egalitarian (1977)	t^E	$t^E = f(t^E)$.
Logit bargaining solution	t^L	$t^L \in \arg \max_{0 \leq t \leq \bar{s}_\alpha} t f(t) \phi(t)$, $\phi(t) = \min \left\{ \frac{1}{t+s_\alpha}, \frac{1}{f(t)+s_\beta} \right\}$.

2.2. Evolutionary contracting

Consider two populations of agents – α and β populations – of size N .⁶ Each period, all agents are matched uniformly at random in heterogeneous pairs of one α -agent and one β -agent to play a coordination game. The set of possible outcomes on which coordination is possible corresponds to a bargaining frontier as described in Section 2.1. Similarly to previous literature, we discretize the bargaining frontier as follows. Let $n \in \mathbb{N}_+$, $\delta = \delta_n = n^{-1} \bar{s}_\alpha$, and $\mathcal{I} := \{0, 1, 2, \dots, n\}$ and suppose that α and β -agents play strategies i_α and i_β from set \mathcal{I} .

We consider *contract games* (Young, 1998a), coordination games in which players who demand the same outcome receive their associated payoffs, and receive nothing otherwise. That is, the payoffs for a contract game are

$$(\pi_\alpha(i_\alpha, i_\beta), \pi_\beta(i_\beta, i_\alpha)) := \begin{cases} (i\delta, f(i\delta)) & \text{if } i_\alpha = i_\beta = i \\ (0, 0) & \text{otherwise} \end{cases}.$$

Thus, when an α -agent plays $i \in \mathcal{I}$ this can be interpreted as him demanding $i\delta$, and when a β -agent plays $i \in \mathcal{I}$ this can be interpreted as him demanding $f(i\delta)$.⁷

A *population state* is described by $x := (x_\alpha, x_\beta)$, where x_α and x_β are vectors giving the number of agents using each strategy. Thus, the state space Ξ is

$$\Xi := \left\{ (x_\alpha, x_\beta) \in \mathbb{N}_0^{n+1} \times \mathbb{N}_0^{n+1} : \sum_{l \in \mathcal{I}} x_\alpha(l) = N, \sum_{l \in \mathcal{I}} x_\beta(l) = N \right\}.$$

More explicitly, we have $(x_\alpha, x_\beta) = ((x_\alpha(0), x_\alpha(1), \dots, x_\alpha(n)), (x_\beta(0), x_\beta(1), \dots, x_\beta(n)))$, where $x_\beta(2)$, for example, denotes the number of β -agents playing strategy 2.

Agents from each population are matched uniformly at random to play the contract game and thus, the expected payoff of an α agent who plays strategy i_α is $\pi_\alpha(i_\alpha, x) := \sum_{l \in \mathcal{I}} \pi_\alpha(i_\alpha, l) x_\beta(l) / N$, given that the fraction of the β population using strategy l is $x_\beta(l) / N$. Similarly, the expected payoff of a β -agent who plays strategy i_β is $\pi_\beta(i_\beta, x) := \sum_{l \in \mathcal{I}} \pi_\beta(i_\beta, l) x_\alpha(l) / N$. Thus, the best response of an α -agent to state x is to choose i to maximize $\pi_\alpha(i, i) x_\beta(i)$, and the best response of a β -agent is to choose i to maximize $\pi_\beta(i, i) x_\alpha(i)$.

We consider the following discrete time strategy updating process. At the beginning of each period, any given agent is independently *activated* with probability $v \in (0, 1)$. Any agent who is not activated will remain playing the same strategy as he did in the previous period. When the current population state is x , an activated agent in population $\gamma \in \{\alpha, \beta\}$ will choose a strategy according to the distribution $p_\gamma^\eta(l|x)$, $l \in \mathcal{I}$. This distribution will be such that an activated agent will usually choose a best response to the profile of strategies played by the opposing population. However, from time to time, an agent will make a *deviation* and play something other than a best response. The parameter η parameterizes the probability of such deviations, with larger values of η corresponding to higher deviation probabilities. As η approaches zero, the probability of any deviation should approach zero at an exponential rate. deviations can be understood as occasional idiosyncratic experimentation, mistakes in play, or as atypical choices arising from random utility shocks. This paper considers processes with perturbations varying in two dimensions: the support of the deviation distribution (unintentional vs. intentional) and the payoff dependence or otherwise of deviations within this support (uniform vs. logit).

Uniform mistake rule (see e.g. Young, 1993a; Kandori et al., 1993).

When deviations are uniform, every deviation occurs with the same probability. That is, from state x , a strategy-revising agent from population $\gamma \in \{\alpha, \beta\}$ will choose l with probability

⁶ Exposition is simplified by the assumption that the populations are of the same size. This is always the case when the two populations represent roles played by different agents in the same population. That is, each agent could be considered to appear twice: he will play one strategy when he plays as an α -player, and another strategy when he plays as a β -player. This differs from one population models of coordination games with two types (e.g. Neary, 2012) as there is always an α -player and a β -player in any matched pair.

⁷ Note that the discretization is uniform for α , but not for β . This can be reversed without changing results.

$$p_\gamma^\eta(l|x) := \begin{cases} \frac{1}{|\arg \max_{\tilde{l}} \pi_\gamma(\tilde{l}, x)|} (1 - \varepsilon) + \frac{1}{n+1} \varepsilon & \text{if } l \in \arg \max_{\tilde{l}} \pi_\gamma(\tilde{l}, x) \\ \frac{1}{n+1} \varepsilon & \text{otherwise} \end{cases}$$

where $\varepsilon = \exp(-\eta^{-1})$. Note, that as required above, as $\eta \rightarrow 0$, the probability of a strategy-revising agent playing anything other than a best response approaches zero.

Logit choice rule (see e.g. Blume, 1993, 1996; Alós-Ferrer and Netzer, 2010).

Under the (generalized) logit choice rule, from state x , a strategy-revising agent from population $\gamma \in \{\alpha, \beta\}$ will choose strategy l with probability

$$p_\gamma^\eta(l|x) := \frac{q_l \exp(\eta^{-1} \pi_\gamma(l, x))}{\sum_{\tilde{l}} q_{\tilde{l}} \exp(\eta^{-1} \pi_\gamma(\tilde{l}, x))} \tag{1}$$

where $q_l, l \in \mathcal{I}$, are positive constants. Again note that as $\eta \rightarrow 0$, the probability of a strategy-revising agent playing anything other than a best response approaches zero.

Intentional & Unintentional deviations (see e.g. Bowles, 2005, 2006; Naidu et al., 2010).

Let $\Delta_\gamma(x)$ be the set of strategies for an agent of type $\gamma \in \{\alpha, \beta\}$ which involve demanding at least as much as the agent demands when best responding to the strategy distribution of the other population.

$$\Delta_\gamma(x) := \{l : \pi_\gamma(l, l) \geq \pi_\gamma(l', l') \text{ for some } l' \in \arg \max_{\tilde{l}} \pi_\gamma(\tilde{l}, x)\}.$$

Unintentional deviation processes retain the probabilities $p_\gamma^\eta(l|x)$ described above for logit and uniform deviations. Intentional deviations are when agents never demand less than their best response, but can demand more. This fits with an interpretation of the perturbations as idiosyncratic experimentation by agents to see if they can obtain a higher payoff. There exists recent experimental evidence which supports such deviations (Lim and Neary, 2016; Mäs and Nax, 2016). The choice probabilities for intentional processes are

$$\hat{p}_\gamma^\eta(l|x) := \begin{cases} \frac{p_\gamma^\eta(l|x)}{\sum_{\tilde{l} \in \Delta_\gamma(x)} p_\gamma^\eta(\tilde{l}|x)} & \text{if } l \in \Delta_\gamma(x) \\ 0 & \text{otherwise} \end{cases}$$

where $p_\gamma^\eta(l|x)$ denotes the choice probability for the corresponding unintentional process.⁸

2.3. Conventions and long run social norms

The process with $\eta = 0$, or $\varepsilon = 0$, is the *unperturbed process*. The recurrent classes of the unperturbed process are the absorbing states in which all α and β -agents coordinate on the same strategy, and each agent type receives nonzero payoff (Young, 1998a). We shall denote by $E_i, i \in \{0, \dots, n\}$, the state in which all agents play strategy $i, x_\alpha(i) = N, x_\beta(i) = N$. Hence, the absorbing states of the process are precisely those in the set $\{E_1, \dots, E_{n-1}\}$. Following Young (1993a), we refer to these states as *conventions*. Let $L := \{1, \dots, n-1\}$ index this set, $\{E_i\}_{i \in L}$.

We consider the long run behavior of our perturbed processes when deviations are unlikely, that is as $\eta \rightarrow 0$. For the current model, each process, uniform or logit, unintentional or intentional, for given $\eta > 0$, has a unique stationary distribution, which we denote μ_η (see Lemma 1 in Appendix A). By standard arguments (see Young, 1998b), the limit $\mu := \lim_{\eta \rightarrow 0} \mu_\eta$ exists, and for any $x \in \Xi, \mu(x) > 0$ implies that x is in a recurrent class of the process with $\eta = 0$. In our setting, this implies that x is a convention.

Definition 1. A state $x \in \Xi$ is stochastically stable if $\mu(x) > 0$.

For small deviation probabilities, in the long run, our processes will spend nearly all of the time at or close to stochastically stable states, hence the interpretation of such states as long run social norms. In the next section we link the stochastically stable states of our four processes to our four bargaining solutions.

⁸ With care, the idea of intentional deviations can be extended to games beyond the coordination games of the current and cited papers. For example, Hwang and Rey-Bellet (2017) consider intentional logit deviations in Nash demand games. Trial and error (e.g. Pradeliski and Young, 2012; Marden et al., 2014) search can also be thought of as a perturbation process with an intentional basis, but with deviation probabilities based on realized payoffs rather than on conjectured states.

Table 3

Anatomy of the easiest escape path from a given convention E_i , $\delta i = t$, by deviation process. *We write that deviations are made by *relatively poor* agents when there exists a threshold \hat{t} , such that from E_i , $\delta i < \hat{t}$, the easiest escape path involves deviations by α -agents, and from E_i , $\delta i > \hat{t}$, the easiest escape path involves deviations by β -agents. When these inequalities are reversed, we write that deviations are made by *relatively rich* agents.

On the easiest escape path from convention E_i , $\delta i = t$, under...	Uniform unintentional	Logit unintentional	Uniform intentional	Logit intentional
Who makes deviations, relatively rich agents or relatively poor agents?*	Rich	Poor	Poor	Poor
Assuming these are α -agents and receive t at the current convention...				
What do they 'demand' when they deviate?	Zero	Zero	$t + \delta$	$t + \delta$
The probability of such deviations decreases exponentially at rate...	N/A	t	N/A	t
As t increases, this quantity...	—	↗	—	↗
The number of deviations required, as a proportion of population size, to induce something other than $f(t)$ as the best response by β -agents.	$\frac{f(t)}{f(t)+\bar{s}_\beta}$	$\frac{f(t)}{f(t)+\bar{s}_\beta}$	$\frac{f(t)}{f(t)+f(t+\delta)}$	$\frac{f(t)}{f(t)+f(t+\delta)}$
As t increases, this quantity...	↘	↘	↗	↗
Net effect of an increase in t on the rareness of such escape paths	↘	↗↘	↗	↗
The convention that is hardest to leave can be approximated by...	t^{KS}	t^L	t^{NB}	t^E

3. Characterization

In this section we characterize the stochastically stable conventions. For a fine discretization (small δ) and large populations (large N), the stochastically stable conventions of our four processes correspond to our four bargaining solutions. The results for uniform deviations are known from Young (1998a) and Naidu et al. (2010). The results for logit deviations are new.

Theorem 1. For any $\zeta > 0$, there exists $\bar{\delta}$ such that for all $\delta < \bar{\delta}$, there exists $N_\delta \in \mathbb{N}$ such that for all $N \geq N_\delta$, $\mu(E_i) > 0 \implies |\delta i - t^*| < \zeta$, where

$$t^* = \begin{cases} t^{KS} & \text{if deviations are uniform-unintentional.} \\ t^L & \text{if deviations are logit-unintentional.} \\ t^{NB} & \text{if deviations are uniform-intentional.} \\ t^E & \text{if deviations are logit-intentional.} \end{cases}$$

The reasons for each process giving rise to its corresponding solution can, for the most part, be simply and intuitively explained. For **unintentional** deviations, any strategy can be played in deviation. From the perspective of a β -agent, the most attractive deviation that can be made by α -agents is for them to switch to strategy 0, as this opens up the possibility of β -agents obtaining their highest payoff of \bar{s}_β by coordinating with such an α -agent. From a convention E_i , $\delta i = t$, the number of such deviations required to make the best response for a β -agent differ from i depends on the ratio of the payoff from successful coordination on the current convention, $f(t)$, to the highest payoff \bar{s}_β (see expression in Table 3). If $f(t)$ is small relative to \bar{s}_β , then few deviations by α -agents will be required to escape the convention. Reprising this argument, we see that if t is small relative to \bar{s}_α , then few deviations by β -agents will be required to escape the convention. For **uniform** deviations, all possible deviations are equally likely, so the difficulty of escaping a convention depends only on the number of deviations required to change the best response. This number is maximized when the payoff ratios discussed above are equal: the Kalai-Smorodinsky solution.

When deviations are **intentional**, agents who make deviations will always ask for more than they receive at the current convention. From convention E_i , $\delta i = t$, from the perspective of a β -agent, the most attractive deviation that can be made by α -agents is for them to ask for $t + \delta$, just a little bit more than they currently receive. The number of deviations required to change the best response of β -agents then depends on the ratio $\frac{f(t+\delta)}{f(t)}$ (see expression in Table 3). This quantity ($\frac{f(t+\delta)}{f(t)}$) and the equivalent quantity for transitions driven by deviations of β -agents ($\frac{(t-\delta)}{t}$) are respectively increasing and decreasing in t . For uniform deviations, the most robust convention is thus where these quantities are equal: the Nash bargaining solution.

For **logit** deviations, to find the rareness of transitions the number of deviations must be weighted by the payoff losses incurred when the deviations are made. For intentional deviations, as the discretization becomes fine, $\frac{f(t+\delta)}{f(t)} \rightarrow 1$ for all strictly positive t , so the number of α -agents who must make deviations in order to alter the best response of β -agents will approach half. Likewise, the number of β -agents who must make deviations in order to alter the best response of α -agents

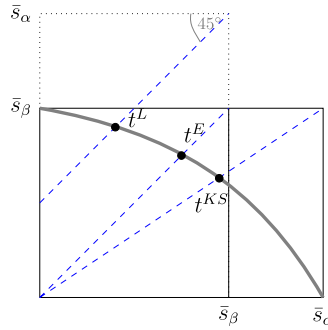


Fig. 1. The intermediate case (Case 2 in Appendix B) of the Logit bargaining solution, also illustrating Egalitarian and Kalai–Smorodinsky solutions for comparison. The Nash bargaining solution is omitted for clarity.

will also approach half. This means that the payoff loss effect dominates and the most robust conventions are those at which deviations by either population are as rare as possible. This occurs when the payoffs of α and β players are equal: the Egalitarian solution.

For logit-unintentional deviations, the effects linking the Kalai–Smorodinsky solution to uniform-unintentional deviations and linking the Egalitarian solution to logit-intentional deviations interact. They do this in a non-trivial way, giving rise to a piecewise solution, sometimes an adjusted Nash bargaining solution, sometimes a form of loss equalization. The logit effect means that deviations are more common when made by relatively poor agents (as defined in the caption to Table 3), but the unintentional effect means that fewer deviations by the other population are required to change the best response of relatively poor agents. Importantly, the logit effect dominates here so that the easiest escape paths are driven by the deviations of relatively poor agents, in contrast to the uniform-unintentional case. The Logit bargaining solution is examined in depth in Appendix B. For now, attention is restricted to one part of the solution as we examine the links between behavioral rules and the properties of the associated bargaining solutions.

3.1. The properties of the solutions as they relate to behavior

It is shown in Appendix B that the Logit bargaining solution reduces to a piecewise function, taking different forms dependent upon whether \bar{s}_α is small, medium or large relative to \bar{s}_β . Here we focus on the intermediate case, for which the solution is given by

$$t^L + \bar{s}_\alpha = f(t^L) + \bar{s}_\beta. \tag{2}$$

This is an *Equal-loss* solution (Chun, 1988) which equalizes players' losses with respect to the *ideal point* $(\bar{s}_\beta, \bar{s}_\alpha)$, effectively an Egalitarian solution with a notional disagreement point of $(\bar{s}_\beta, \bar{s}_\alpha)$ (see Fig. 1). The Equal-loss solution typically has an ideal point of $(\bar{s}_\alpha, \bar{s}_\beta)$, so that each player's ideal is his own highest possible payoff. In contrast, (2) gives each player's ideal as the highest possible payoff of his opponent. The intuition behind this reversal is explained later in this section. Note that, similarly to previous work that bounds bargaining solutions with respect to one another (see, e.g. Rachmilevitch, 2016, 2015; Cao, 1982), when the Logit solution is given by (2), the Egalitarian solution lies between the Logit solution and the Kalai–Smorodinsky solution.

It follows from (2) that the Logit solution, like the Egalitarian solution, does not scale with linear transformations of payoffs (the Invariance property – Table 4). This arises from logit choice. deviation rates are measured in probability and thus interpersonally comparable. The rate of deviations by any given agent under logit choice depends on his cardinal payoff loss from making that deviation. Therefore interpersonal comparability of payoffs follows from interpersonal comparability of deviation rates. However, interpersonal comparability of payoffs is precisely what is prohibited under the Invariance property, so the solutions that emerge from logit choice (Logit bargaining solution, Egalitarian solution) do not satisfy Invariance. In contrast, under uniform deviations, choice probabilities are unaffected by a linear transformation of an agent's payoffs, so the solutions that emerge (Kalai–Smorodinsky, Nash) satisfy Invariance.

The presence of \bar{s}_α and \bar{s}_β in expression (2) implies that the Logit bargaining solution does not have the Independence of Irrelevant Alternatives property (IIA – Table 4). This is noteworthy, as under the logit choice rule, the ratios of choice probabilities $\frac{p_{i'}^y(l|x)}{p_{i'}^y(l'|x)}$ are independent of the payoffs from any strategy $\tilde{l} \neq l, l'$. This shows that the IIA property at a micro level (choice behavior) does not translate into an IIA property at a macro level (long run social norm). The dependence of (2) on \bar{s}_α and \bar{s}_β , and hence the failure of IIA, arises from the fact that deviations in the underlying behavioral process are unintentional. When deviation behavior is unintentional, the least cost transition paths between conventions involve jumps to extreme states motivated by the prospect of extreme payoffs $(\bar{s}_\alpha$ and $\bar{s}_\beta)$, so the solutions that emerge (Kalai–Smorodinsky, Logit bargaining solution) do not satisfy IIA. In contrast, when deviations are intentional, least cost transition paths are between adjacent conventions and the solutions that emerge (Nash, Egalitarian) satisfy IIA.

Table 4

Definitions of properties and the bargaining solutions that satisfy them. In each definition, g , f are bargaining frontiers and t_g^* , t_f^* their associated solutions. Invariance implies Stretch Monotonicity, and Monotonicity implies Individual Monotonicity which implies Stretch Monotonicity. Note that t^L satisfies none of these properties.

Property	Definition	Satisfied by...
IIA	$g \geq f$, $g(t_g^*) = f(t_f^*) \implies t_g^* = t_f^*$.	t^{NB}, t^E
Invariance	$g(x) = f(ax)$, $a \in \mathbb{R} \implies t_g^* = \frac{1}{a} t_f^*$.	t^{NB}, t^{KS}
Monotonicity	$g \geq f \implies t_g^* \geq t_f^*$.	t^E
Individual Monotonicity	$g \geq f$, $g(0) = f(0) \implies t_g^* \geq t_f^*$.	t^{KS}, t^E
Stretch Monotonicity	$g(x) = f(ax)$, $a \in \mathbb{R}$, $a < 1 \implies t_g^* \geq t_f^*$.	t^{NB}, t^{KS}, t^E

There does not seem to be a similarly simple way to relate monotonicity properties of solutions to behavioral axioms. We can, however, fit the Logit bargaining solution into the hierarchy of solutions sorted by their monotonicity properties. It turns out that the Logit bargaining solution is the least monotonic of the four solutions. Specifically, we say that a solution is *Stretch Monotonic* if a stretch of the bargaining frontier parallel to the horizontal axis leads to the player in the α position obtaining a higher payoff (Stretch Monotonicity – Table 4). This property is implied by Invariance and Individual Monotonicity, and thus satisfied by the Egalitarian, Kalai-Smorodinsky and Nash solutions. However, we shall see that it is not satisfied by the Logit bargaining solution.

Consider a linear frontier given by $f(t) = \bar{s}_\beta - t \frac{\bar{s}_\beta}{\bar{s}_\alpha}$. A stretch parallel to the horizontal axis is equivalent to an increase in the value of \bar{s}_α . Substituting the expression for the frontier into (2), we obtain the solution $t^L = \frac{(2\bar{s}_\beta - \bar{s}_\alpha)\bar{s}_\alpha}{\bar{s}_\alpha + \bar{s}_\beta}$. This is decreasing in \bar{s}_α whenever \bar{s}_α and \bar{s}_β take similar values. That is, an increase in the best possible payoff for the α player leads to his obtaining a lower payoff under the Logit bargaining solution. This may seem puzzling at first, but makes sense when we consider that, under logit deviations, the easiest escape path from a convention involves deviations by relatively poor agents. Consider conventions for which $t > t^L$. These are the conventions at which α -agents obtain high payoffs and from which the most likely escape paths involve deviations by β -agents. Under unintentional deviations, the number of deviations by β -agents required to change the best response of α -agents is lower when the best possible payoff \bar{s}_α for α -agents is higher. That is, α -agents suffer from a high best possible payoff as this destabilizes the conventions where α -agents do well and pushes the value of t^L lower. See Appendix B for a complete illustrated solution for the linear bargaining frontier.

4. Experimental evidence for intentional and payoff dependent deviations

To study human agents' non-best response behavior in a context derived from the model of this paper, we conducted laboratory experiments on human subjects. Five sessions were conducted in English at the experimental laboratory at the Hong Kong University of Science and Technology. A fixed-role, between-subject design was used. In total, 100 participants, none of whom had any prior experience with this work, were recruited from the university graduate and undergraduate population. All sessions were conducted using z-Tree (Fischbacher, 2007). Each session lasted for approximately two hours, and the average amount earned per subject was HKD 132 (USD 17), including the HKD 30 show-up fee. The maximum payment was HKD 210 and the minimum payment was HKD 57. The standard deviation was 57.

Each session involved α and β populations of subjects (representing the agents in the model of Section 2) of size $N = 10$ playing for 200 periods. Each period, every subject, independently with probability ν ($= 0.9$ in Sessions 1, 2, 3; $= 0.5$ in Sessions 4, 5), was activated and got the opportunity to update his strategy. The information displayed to a subject was:

- (i) His *own* payoff when playing any given strategy and matched to a member of the other population also playing that strategy.
- (ii) For periods $\tau \geq 2$, the number of subjects in the other population who played each strategy in the preceding period.
- (iii) The strategy played and the payoff obtained by the subject in the preceding period.
- (iv) Whether or not the subject has the opportunity to update his strategy in the current period.

Note that subjects were given no information about the coordination payoffs of subjects in the other population and the laboratory was set up to ensure that no subjects would gain any such information during the session. This absence of information about others' coordination payoffs is similar to Mäs and Nax (2016).⁹ This lack of information about the payoffs of subjects in the other population is motivated by our interest in the emergence of conventions (i.e. convergence to E_1 , E_2 or E_3) as predicted by the theory. If such information were provided, it could be used to coordinate on a pre-existing norm from outside of the laboratory. For more on this, see the end of this section.

⁹ As our model is of perturbed best response dynamics, we provide aggregate information on the choices of subjects in the other population. This can be contrasted with a 'black box' approach (Nax et al., 2016) in which not even this information is provided.

		β -player				
		0	1	2	3	4
α -player	0	0, 100	0, 0	0, 0	0, 0	0, 0
	1	0, 0	80, 80	0, 0	0, 0	0, 0
	2	0, 0	0, 0	130, 65	0, 0	0, 0
	3	0, 0	0, 0	0, 0	180, 50	0, 0
	4	0, 0	0, 0	0, 0	0, 0	200, 0

Fig. 2. Entries give payoffs for the α and β players respectively.

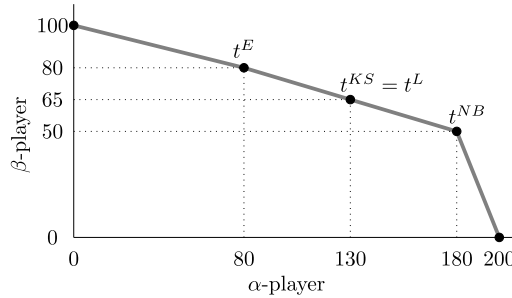


Fig. 3. Payoffs from pairwise interaction between subjects (see Fig. 2) depicted as points on the frontier of a (weakly) convex bargaining set.

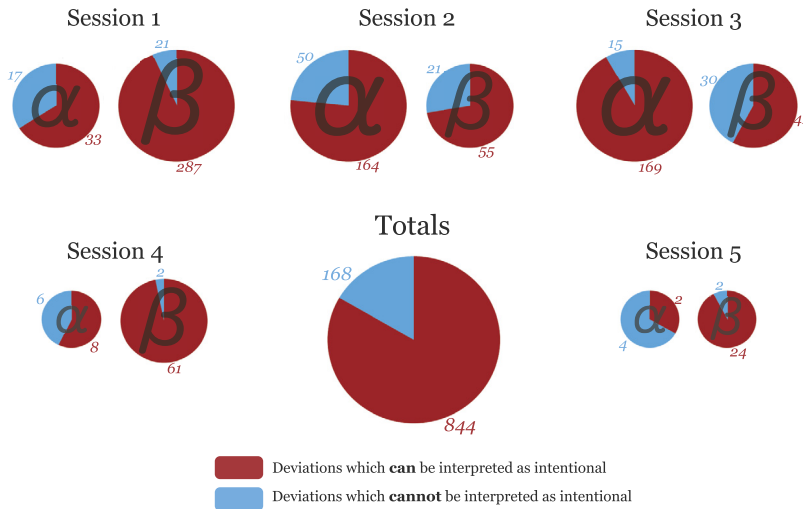


Fig. 4. Chart showing, for each session and position (α or β) in the game, the number of non-best responses that can be interpreted as intentional (those in $\Delta_V(x)$) and that cannot be interpreted as intentional (those not in $\Delta_V(x)$), as well as totals across all sessions and positions. The area of each pie chart is approximately proportional to the square root of the number of deviations it represents.

Subjects given the opportunity to change their strategy could choose any of strategies 1 to 5, which were labeled as A to E, the order of the labeling and the order of presentation of the strategies differing across sessions. If they were activated and failed to choose a strategy within a specified period, they remained playing the same strategy as in the previous period. Following strategy updating, α -subjects were paired with β -subjects and obtained payoffs corresponding to their chosen strategies played against one another in the game in Fig. 2 (depicted as points on a bargaining frontier in Fig. 3). The instructions given to subjects and images of the decision making interface are given in Appendix C.

In every session, best responses constituted a large majority (> 90%) of choices by the subjects and play converged to a convention. Analysis of non-best response play reveals that subjects rarely update to strategies that correspond to payoffs lower than the payoff associated with their best response strategy (see Fig. 4). In every session-population pair except one (Session 5, α -subjects, with only 6 data points) deviations which can be interpreted as intentional outnumber deviations which cannot be so interpreted. In total, 844 out of 1012 deviations (> 83%) can be interpreted as intentional. Methodologically, as the game has five strategies, rather than two as in recent similar studies (Lim and Neary, 2016; Mäs and Nax, 2016), for any given subject with a given best response there exist multiple non-best response strategies, and

Table 5

For given player position and best response strategy, aggregate rate of non-best response play and proportion of non-best responses that can be interpreted as intentional. An asterisk (*) indicates that a value differs significantly ($p < 0.001$) from the expected proportion that would obtain if all non-best responses were equally likely. Note that there were only 29 updating opportunities and 13 non-best responses when the best response was strategy 1 or strategy 5, hence these columns are omitted from the table.

For subjects in population...	α			β		
...with best response strategy...	2	3	4	2	3	4
Number of updating opportunities.	1720	1786	3851	1740	1773	3861
Number of non-best responses observed.	183	189	87	49	59	432
Rate of non-best response play.	0.11	0.11	0.02	0.03	0.03	0.11
Proportion of non-best responses that can be interpreted as intentional (those in $\Delta_\gamma(x)$).	0.94*	0.94*	0.30	0.55*	0.41	0.97*
Expected such proportion if all non-best responses were equally likely.	0.75	0.50	0.25	0.25	0.50	0.75

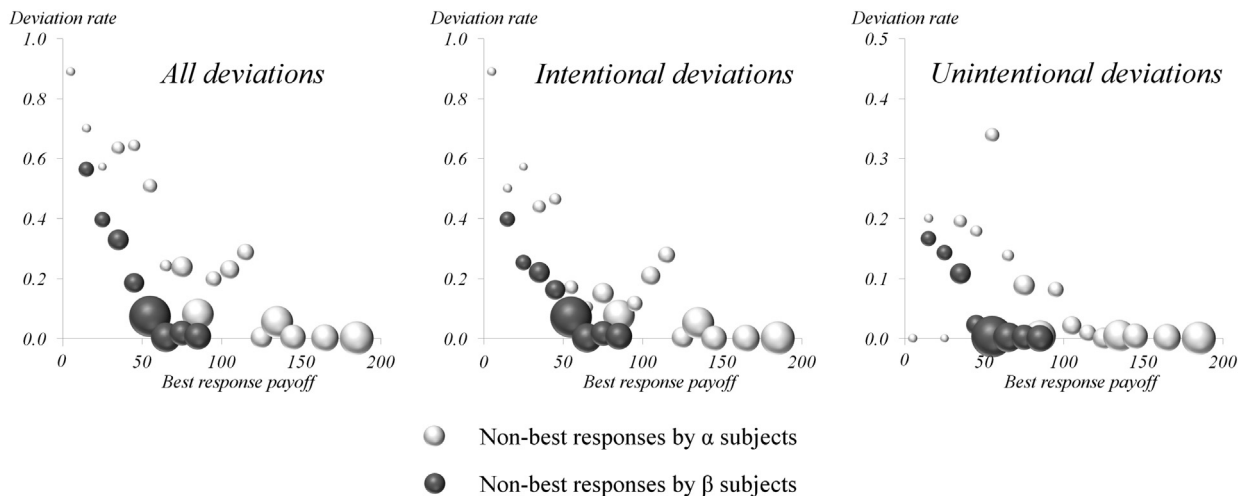


Fig. 5. Rates of non-best response play grouped by expected payoff from the best response strategy. Data are grouped into bins by best response payoff (0 to 10, 10 to 20, etc.). The area of each circle on the chart is proportional to the square root of the number of strategy updating opportunities it represents. Versions of this diagram that exclude early and late periods are given in Appendix D.

so we were able to classify observed deviations as intentional or not on an individual basis.¹⁰ Aggregating over subjects' best response strategies at observed states, we find that in four out of the six population-best response pairs for which we have enough data, the proportion of deviations that can be interpreted as intentional differs significantly (statistically, with $p < 0.001$) from that which we would expect to obtain if every non-best response strategy were equally likely (see Table 5). In summary, there is clear support for some level of intentional behavior in non-best response play.

Furthermore, we observe higher rates of non-best response play from subjects for whom the expected payoff from the best response strategy is lower (Fig. 5). That is, there is support for payoff dependence in non-best response play. From the Figure, it can be seen that this is true across both α and β populations and holds regardless of whether deviations can be interpreted as intentional. Note that although the charts appear similar, the vertical axis of the unintentional chart has a scale of half that of the other charts, reflecting the relative number of deviations of each type as discussed in the preceding paragraph. Further support for payoff dependence is found in Appendix D, where we conduct a logistic regression on payoff differences that accounts for the possibility of decreasing deviation rates as a function of time.

An important aspect of our design is that α -subjects do not know the coordination payoffs of β -subjects and vice versa. This brings two benefits. Firstly, the potential impact of other-regarding preferences is minimized, as any beliefs about the payoffs of the opposing population would have to be inferred from behavior. Secondly, and we believe more importantly, subjects' strategy choice cannot be influenced by pre-existing social norms. If subjects could observe the full payoff structure of the game, then a pre-existing norm such as the Kalai–Smorodinsky, Nash, Egalitarian or Logit solutions could encourage subjects to play $i = 2, 3, 1, 2$ respectively. The outcome of the sessions suggests that subjects did not infer any correspondence between a pre-existing norm and the payoff structure of the game, as of the five sessions, three converged to convention E_3 (Nash), while one converged to E_2 (Kalai–Smorodinsky, Logit) and one to E_1 (Egalitarian).

¹⁰ In contrast, if there are only two strategies, then, for a given subject with a given best response, there is only one possible deviation, so it is impossible to effect any classification at the level of the individual deviation.

The observed short run convergence to some convention is predicted by the theoretical processes considered earlier in the paper. Note though that nothing can be inferred about long run norms from observing the conventions that were reached. For such an analysis, we would require observations over a longer timescale than is feasible within the constraints of the laboratory.¹¹ However, Theorem 1 can be used to comment on long run norms by extrapolating from the characteristics of observed non-best response behavior. Our observations suggest that non-best response play is intentional and payoff dependent. Therefore, the theory suggests that, of the four solutions considered in this paper, the Egalitarian solution is the most likely candidate for a long run social norm.

As noted earlier in this section, information about the payoffs of subjects in the other population could potentially facilitate the application of pre-existing norms that evolved outside of the laboratory. This seems to be what was observed in pilot sessions that were carried out prior to a decision on a final design. In these sessions, opponents' payoffs were observable. There were three of these sessions, two which used the same payoffs as are used here, and a third which used different payoffs.¹² In all of these sessions, there was convergence within a few periods to the convention corresponding to the Egalitarian solution. This provides some slight further support for egalitarianism, but too much should not be read into this, as each of these sessions can be understood as a single data point.

We note that implications of the experimental results of the paper apply specifically to the class of games, the experimental environment and context that we consider. Rates of best response behavior and the type of non-best response behavior observed may differ for other classes of games or other experimental environments. This is, in itself, an important research topic.

5. Conclusion

What has been presented here is a theory of the emergence of bargaining solutions as social norms that rests on behavior and not on the properties of the solutions themselves. Societal interactions take individual behavior as an input and give a social norm, or to put it another way, a social choice, as output. Thus, if we think of a society as similar to an organism with agency, we can regard the traditional, normative approaches to social choice as specifying behavioral rules for society itself, rather than for the individuals within society. As such, the results of the current paper and the rest of the Evolutionary Nash Program can be understood as a reconciliation of micro and macro behavioral theories: the presence of well developed norms in a group should help members of the group to face new problems, take decisions and adjust their behavior as if they were of one mind. Conversely, in recent years there has been considerable work on the effect of collective agency at a small group level on social norms (see Newton, 2018, Section 2, for a survey). The modeling of two way influence between norms and agency is an interesting avenue for future research.

Appendix A. Proofs

Denote by $P^\eta(x, y)$ the transition probability from state x to state y . Define the resistance of such a transition, $V(x, y)$, as

$$V(x, y) := \lim_{\eta \rightarrow 0} -\eta \ln P^\eta(x, y) \tag{A.1}$$

where V is defined over the set of all $x, y \in \Xi$ such that $P^\eta(x, y) > 0$ for some $\hat{\eta} > 0$ (see Beggs, 2005; Sandholm, 2010).

For **uniform deviations**, $V(x, y)$ equals the number of agents who switch to anything other than a best response.

$$V(x, y) = \sum_{\gamma \in \{\alpha, \beta\}} \sum_{l \notin \arg \max_{\tilde{l}} \pi_\gamma(\tilde{l}, x)} \max\{y_\gamma(l) - x_\gamma(l), 0\}. \tag{A.2}$$

For the **logit choice rule**, $V(x, y)$ equals the best response payoff minus the payoff from the chosen strategy, summed over all updating agents.

$$V(x, y) = \sum_{\gamma \in \{\alpha, \beta\}} \sum_{l \in \mathcal{I}} \max\{y_\gamma(l) - x_\gamma(l), 0\} \left(\max_{\tilde{l}} \pi_\gamma(\tilde{l}, x) - \pi_\gamma(l, x) \right). \tag{A.3}$$

¹¹ To see this via a rough calculation, note that when deviations are intentional, the easiest path from one convention to another will usually require 6 or more deviations in one of the populations. In a period in which all subjects in a given population have the opportunity to update their strategies, at deviation rates of 10% (the approximate frequency observed in this study), 6 or more deviations will occur with a probability of less than $\frac{1}{6000}$. Therefore to make a statement about long run norms using aggregate population data alone would require sessions to last for considerably more than 6000 periods.

¹² These pilot sessions also differed from the final design in that payoffs were presented to subjects as points on a bargaining frontier rather than in a game matrix. This was amended to remove any cues that might suggest bargaining to the subjects, in line with our interest in the emergence of bargaining solutions from coordination games, rather than from an explicit bargaining context. Instructions to the subjects in the pilot sessions are available from the authors on request.

Lemma 1. Each process, uniform or logit, unintentional or intentional, for given $\eta > 0$, has a unique stationary distribution, which we denote μ_η .

Proof. Note that for all $x \in \Xi$, $n \in \Delta_\alpha(x)$, so for all of our processes, from any $x \in \Xi$, we have that $P^\eta(x, y) > 0$ for some y such that $y_\alpha(n) = N$. From y , n is a best response for any β -agent, so $P^\eta(y, E_n) > 0$. Therefore, from any $x \in \Xi$, with positive probability E_n will be reached within two periods. As the state space is finite, standard results in Markov chain theory¹³ imply that for all $\eta > 0$, P^η has a unique recurrent class and μ_η exists and is unique. \square

In a similar way that $V(\cdot, \cdot)$ measures the rarity of single steps in the process, we will use a concept, overall cost, that measures the rarity of a transition between any two states over any number of periods. Let $\mathcal{P}(x, x')$ be the set of finite sequences of states $\{x^1, x^2, \dots, x^T\}$ such that $x^1 = x$, $x^T = x'$ and for some $\hat{\eta} > 0$, $P^{\hat{\eta}}(x^\tau, x^{\tau+1}) > 0$, $\tau = 1, \dots, T - 1$.

Definition 2. The overall cost of a transition between $x, x' \in \Xi$ is:

$$c(x, x') := \min_{\{x^1, \dots, x^T\} \in \mathcal{P}(x, x')} \sum_{\tau=1}^{T-1} V(x^\tau, x^{\tau+1}). \tag{A.4}$$

If there is no positive probability path between x and x' then let $c(x, x') = \infty$. We shall be interested in the cost of transitions between conventions. In the current setting, this quantity is always finite. Denote the overall cost functions for the uniform-unintentional, logit-unintentional, uniform-intentional and logit-intentional processes by c^U , c^L , c^{UI} , c^{LI} respectively.

Lemma 2. For $c \in \{c^U, c^L, c^{UI}, c^{LI}\}$, $i \in L$, let

$$F_i := \left\{ x \in \Xi : \text{For some } \gamma \in \{\alpha, \beta\}, \{i\} \neq \arg \max_{j \in \mathcal{I}} \pi_\gamma(j, x) \right\}.$$

Then, to calculate $\min_{x \in F_i} c(E_i, x)$ via the minimization in (A.4) it suffices to consider $\{x^1, \dots, x^T\} \in \mathcal{P}(E_i, x)$ such that, for $\tau < T$, x^τ and $x^{\tau+1}$ are identical except that for some $j \in \mathcal{I}$, $\gamma \in \{\alpha, \beta\}$, $x_\gamma^{\tau+1}(i) = x_\gamma^\tau(i) - 1$ and $x_\gamma^{\tau+1}(j) = x_\gamma^\tau(j) + 1$.

Proof. Let $\{x^1, \dots, x^T\}$, $x^1 = E_i$, $x^T \in F_i$, be such that

$$\min_{x \in F_i} c(E_i, x) = \sum_{\tau=1}^{T-1} V(x^\tau, x^{\tau+1}). \tag{A.5}$$

As $V(\cdot, \cdot) \geq 0$, we can, without loss of generality, assume that $x^t \notin F_i$ for $t < T$. For $t = 1 \dots, T - 1$, for all $\gamma \in \{\alpha, \beta\}$, define

$$\begin{aligned} y_\gamma^1 &= x_\gamma^1, \\ y_\gamma^{t+1}(j) &= y_\gamma^t(j) + \max\{x_\gamma^{t+1}(j) - x_\gamma^t(j), 0\} \quad \text{for } j \neq i, \\ y_\gamma^{t+1}(i) &= N - \sum_{j \neq i} y_\gamma^{t+1}(j). \end{aligned}$$

That is, $\{y^1, \dots, y^T\}$ differs from $\{x^1, \dots, x^T\}$ only in that all transitions to any $j \in \mathcal{I}$ at $t + 1$ are now by agents who played i at t .

Let t' be the smallest t such that $y^t \in F_i$. $t' \leq T$ as $y_\gamma^T(i) \leq x_\gamma^T(i)$, $y_\gamma^T(j) \geq x_\gamma^T(j)$ for $j \neq i$, $x^T \in F_i$ implies $y^T \in F_i$. By (A.2) or (A.3), $y^t \notin F_i$ implies $V(y^t, y^{t+1}) \leq V(x^t, x^{t+1})$. Therefore, if $t' < T$, then $c(y^1 = x^1, y^{t'}) \leq \sum_{t=1}^{t'-1} V(y^t, y^{t+1}) \leq \sum_{t=1}^{t'-1} V(x^t, x^{t+1}) < \sum_{t=1}^{T-1} V(x^t, x^{t+1})$, contradicting (A.5). So $t' = T$ and for all $t < T$, we have $y^t \notin F_i$ and

$$V(y^t, y^{t+1}) \leq V(x^t, x^{t+1}). \tag{A.6}$$

Now, if $\sum_\gamma y_\gamma^{t+1}(i) < \sum_\gamma y_\gamma^t(i) - 1$, then take some j, γ such that $y_\gamma^{t+1}(j) > y_\gamma^t(j)$ and define y^{t+} to be identical to y^t except that $y_\gamma^{t+}(i) = y_\gamma^t(i) - 1$ and $y_\gamma^{t+}(j) = y_\gamma^t(j) + 1$. Then, by (A.2) or (A.3) we have

$$V(y^t, y^{t+}) + V(y^{t+}, y^{t+1}) \leq V(y^t, y^{t+1}). \tag{A.7}$$

¹³ See, for example, "Probability" by Shiryaev (1995, p. 586, Theorem 4).

Now replace $\{y^1, \dots, y^t, y^{t+1}, \dots, y^T\}$ with $\{y^1, \dots, y^t, y^{t+}, y^{t+1}, \dots, y^T\}$ and iterate this procedure until we obtain $\{z^1, \dots, z^{T'}\}$ such that $z^1 = y^1$, $z^{T'} = y^T$, and either $z^{t+1} = z^t$ or $\sum_{\gamma} z_{\gamma}^{t+1}(i) = \sum_{\gamma} z_{\gamma}^t(i) - 1$ for $t = 1, \dots, T' - 1$. If $z^{t+1} = z^t$, then $V(z^t, z^{t+1}) = 0$, so we omit such transitions and renumber our sequence $\{z^1, \dots, z^{T'}\}$, which now satisfies the conditions in the statement of the lemma. Now,

$$\min_{x \in F_i} c(E_i, x) \underbrace{\leq}_{\text{by defn}} \sum_{\tau=1}^{\tilde{T}-1} V(z^{\tau}, z^{\tau+1}) \underbrace{\leq}_{\text{by iterating (A.7)}} \sum_{\tau=1}^{T-1} V(y^{\tau}, y^{\tau+1}) \underbrace{\leq}_{\text{by (A.6)}} \sum_{\tau=1}^{T-1} V(x^{\tau}, x^{\tau+1}) \underbrace{=}_{\text{by (A.5)}} \min_{x \in F_i} c(E_i, x),$$

which completes the proof. \square

Lemma 3. For $i \in L$,

$$c^U(E_i, E_j) = \min \left\{ \left\lceil N \frac{f(\delta i)}{f(\delta i) + \bar{s}_{\beta}} \right\rceil, \left\lceil N \frac{\delta i}{\delta i + \bar{s}_{\alpha}} \right\rceil \right\} \quad \text{for all } j \neq i, \tag{A.8}$$

$$c^L(E_i, E_j) \approx \min \left\{ \delta i \left\lceil N \frac{f(\delta i)}{f(\delta i) + \bar{s}_{\beta}} \right\rceil, f(\delta i) \left\lceil N \frac{\delta i}{\delta i + \bar{s}_{\alpha}} \right\rceil \right\} \quad \text{for all } j \neq i, \tag{A.9}$$

$$\min_{j \neq i} c^{U^I}(E_i, E_j) = \min \left\{ \left\lceil N \frac{f(\delta i)}{f(\delta(i+1)) + f(\delta i)} \right\rceil, \left\lceil N \frac{i}{2i-1} \right\rceil \right\}, \tag{A.10}$$

$$\min_{j \neq i} c^{L^I}(E_i, E_j) \approx \min \left\{ \delta i \left\lceil N \frac{f(\delta i)}{f(\delta(i+1)) + f(\delta i)} \right\rceil, f(\delta i) \left\lceil N \frac{i}{2i-1} \right\rceil \right\}, \tag{A.11}$$

where $a \approx b$ denotes $|a - b| \leq \max\{\bar{s}_{\alpha}, \bar{s}_{\beta}\}$.

Proof. Let ξ_i^{γ} be the lowest number of deviations by a γ -agent, $\gamma \in \{\alpha, \beta\}$, on any transition path from E_i , $i \in L$, to some E_j , $j \in \mathcal{I}$, $j \neq i$. At some point on any such path, some $j \neq i$ must become a best response. Therefore,

$$\xi_i^{\alpha} \max_{j \in C_i^{\alpha}} \pi_{\beta}(j, j) \geq (N - \xi_i^{\alpha}) \pi_{\beta}(i, i) \quad \text{and} \quad \xi_i^{\beta} \max_{j \in C_i^{\beta}} \pi_{\alpha}(j, j) \geq (N - \xi_i^{\beta}) \pi_{\alpha}(i, i),$$

where $C_i^{\gamma} = \mathcal{I}$ for unintentional processes and $C_i^{\gamma} = \Delta_{\gamma}(E_i) \setminus \{i\}$ for intentional processes. It follows that ξ_i^{α} is attained when α -agents make deviations and play $k \in \arg \max_{j \in C_i^{\alpha}} \pi_{\beta}(j, j)$, and ξ_i^{β} is attained when β -agents make deviations and play $k \in \arg \max_{j \in C_i^{\beta}} \pi_{\alpha}(j, j)$.

$$\xi_i^{\alpha} = \min_{j \in C_i^{\alpha}} \left\lceil N \frac{\pi_{\beta}(i, i)}{\pi_{\beta}(i, i) + \pi_{\beta}(j, j)} \right\rceil \quad \text{and} \quad \xi_i^{\beta} = \min_{j \in C_i^{\beta}} \left\lceil N \frac{\pi_{\alpha}(i, i)}{\pi_{\alpha}(i, i) + \pi_{\alpha}(j, j)} \right\rceil. \tag{A.12}$$

Now, $\pi_{\alpha}(i, i) = \delta i$ and $\pi_{\beta}(i, i) = f(\delta i)$. For unintentional processes

$$\max_{j \in C_i^{\alpha}} \pi_{\beta}(j, j) = \bar{s}_{\beta} \quad \text{and} \quad \max_{j \in C_i^{\beta}} \pi_{\alpha}(j, j) = \bar{s}_{\alpha}, \tag{A.13}$$

and for intentional processes

$$\max_{j \in C_i^{\alpha}} \pi_{\beta}(j, j) = f(\delta(i+1)) \quad \text{and} \quad \max_{j \in C_i^{\beta}} \pi_{\alpha}(j, j) = \delta(i-1). \tag{A.14}$$

For uniform deviations, each deviation adds 1 to the cost of the transition, therefore the least cost transition from E_i , $i \in L$, to some E_j , $j \in \mathcal{I}$, $j \neq i$, is one involving the fewest deviations. The cost of such a transition is then $\min\{\xi_i^{\alpha}, \xi_i^{\beta}\}$, which by (A.12) and (A.13), equals the RHS of (A.8) for unintentional deviations, and by (A.12) and (A.14), equals the RHS of (A.10) for intentional deviations.

For logit deviations, as each deviation is weighted by the expected payoff loss, the lowest cost from the transitions involving the fewest deviations is $\min\{\pi_{\alpha}(i, i) \xi_i^{\alpha}, \pi_{\beta}(i, i) \xi_i^{\beta}\}$. There may exist lower cost transitions, but as Lemma 2 tells us we can restrict attention to paths in which one agent switches at a time, we can invoke Theorem 1 of Hwang and Newton (2017) and, for intentional processes, Theorem 1 from Hwang and Newton (2014), to give

$$\min_{j \neq i} c(E_i, E_j) \geq \min\{\pi_{\alpha}(i, i)(\xi_i^{\alpha} - 1), \pi_{\beta}(i, i)(\xi_i^{\beta} - 1)\},$$

so we have the RHS of (A.9) and (A.11).

Finally, note that for unintentional deviations, by (A.13), lowest cost transitions out of E_i involve extreme deviations in which either α -agents switch to 0 until 0 is a best response for β -agents, or β -agents switch to n until n is a best response for α -agents. Consider α -agents making deviations until 0 is a best response for β -agents. It is then possible that all β -agents update their strategy to 0, to reach a state x such that $x_\beta(0) = N$. From such a state, any strategy is a best response for α -agents, so it is possible that they all choose some arbitrary strategy k , following which k becomes a best response for β -agents who can in turn switch to k so that E_k is reached. We see that any least cost exit path from E_i can reach E_k for arbitrary k without further mistakes. That is, $c(E_i, E_k) = \min_{j \neq i} c(E_i, E_j)$ for all $k \neq i$. \square

The next step is to characterize the stochastically stable states of our model for given δ and large population size.

Definition 3. The following expressions are the limits of $\frac{1}{N}$ multiplied by (A.8), (A.9), (A.10), (A.11) as $N \rightarrow \infty$, written as a function of $t = \delta i$.

$$\begin{aligned} \varphi_\delta^U(t) &:= \min \left\{ \frac{f(t)}{f(t) + \bar{s}_\beta}, \frac{t}{t + \bar{s}_\alpha} \right\}, \\ \varphi_\delta^L(t) &:= \min \left\{ t \frac{f(t)}{f(t) + \bar{s}_\beta}, f(t) \frac{t}{t + \bar{s}_\alpha} \right\}, \\ \varphi_\delta^{UI}(t) &:= \min \left\{ \frac{f(t)}{f(t + \delta) + f(t)}, \frac{t}{2t - \delta} \right\}, \\ \varphi_\delta^{LI}(t) &:= \min \left\{ t \frac{f(t)}{f(t + \delta) + f(t)}, f(t) \frac{t}{2t - \delta} \right\}. \end{aligned}$$

Lemma 4. For $c \in \{c^{UI}, c^{LI}\}$ and corresponding $\phi_\delta \in \{\phi_\delta^{UI}, \phi_\delta^{LI}\}$, let $i \in \arg \max_{j \in L} \varphi_\delta(j\delta)$. Then

$$\begin{aligned} \phi_\delta(j\delta) &= \lim_{N \rightarrow \infty} \frac{1}{N} c(E_j, E_{j+1}) \quad \text{for } j < i, \\ \phi_\delta(j\delta) &= \lim_{N \rightarrow \infty} \frac{1}{N} c(E_j, E_{j-1}) \quad \text{for } j > i. \end{aligned}$$

Proof. We can write $\phi_\delta(t) = \min\{a(t), b(t)\}$. As $a(t)$ is increasing in t , $b(t)$ is decreasing in t , $a(i\delta) \leq b(i\delta)$ implies

$$\begin{aligned} a(j\delta) &< b(j\delta) \quad \text{for } j < i, \\ a(j\delta) &> b(j\delta) \quad \text{for } j > i \quad [\text{otherwise } \phi_\delta(j\delta) > \phi_\delta(i\delta), \text{ contradicting } i \in \arg \max_{j \in L} \varphi_\delta(j\delta)]. \end{aligned}$$

If $a(i\delta) > b(i\delta)$, a similar argument implies the same conclusion. To conclude, note that by the proof of Lemma 3,

$$\begin{aligned} a(j\delta) < b(j\delta) &\implies \lim_{N \rightarrow \infty} \frac{1}{N} c(E_j, E_{j+1}) = \lim_{N \rightarrow \infty} \frac{1}{N} \min_{k \neq j} c(E_j, E_k) = \phi_\delta(j\delta), \\ a(j\delta) > b(j\delta) &\implies \lim_{N \rightarrow \infty} \frac{1}{N} c(E_j, E_{j-1}) = \lim_{N \rightarrow \infty} \frac{1}{N} \min_{k \neq j} c(E_j, E_k) = \phi_\delta(j\delta). \quad \square \end{aligned}$$

Definition 4. An i -graph is a directed graph on L such that every vertex except for i has exactly one exiting edge and the graph has no cycles. Let $\mathcal{G}(i)$ denote the set of i -graphs. For a graph g , let $(j \rightarrow k) \in g$ denote an edge from j to k in g . Define stochastic potential:

$$SP(i) := \min_{g \in \mathcal{G}(i)} \sum_{(j \rightarrow k) \in g} c(E_j, E_k). \tag{A.15}$$

We know from Freidlin and Wentzell (1984, chap.6), Young (1993a) that:

$$\mu(E_i) > 0 \iff i \in \arg \min_{j \in L} SP(j).$$

Lemma 5. For $c \in \{c^U, c^L, c^{UI}, c^{LI}\}$ and corresponding $\phi_\delta \in \{\phi_\delta^U, \phi_\delta^L, \phi_\delta^{UI}, \phi_\delta^{LI}\}$, there exists N_δ such that for all $N \geq N_\delta$, $\mu(E_i) > 0 \implies i \in \arg \max_{j \in L} \varphi_\delta(j\delta)$.

Proof. Let $l \notin \arg \max_{j \in L} \varphi_\delta(j\delta)$ and let $i \in \arg \max_{j \in L} \varphi_\delta(j\delta)$. Consider $g \in \mathcal{G}(i)$. For $c = c^U, c^L$, let $g = \{j \rightarrow i : j \in L, j \neq i\}$. For $c = c^{U1}, c^{L1}$, let $g = \{j \rightarrow j + 1 : j \in L, j < i\} \cup \{j \rightarrow j - 1 : j \in L, j > i\}$. Note that $(j \rightarrow k) \in g$ correspond to $c(E_j, E_k)$ that solve $\min_k c(E_j, E_k)$. For $c = c^{U1}, c^{L1}$, this follows from the proof of Lemma 3, $i \in \arg \max_{j \in L} \varphi_\delta(j\delta)$, and the fact that one term inside the minimization defining the corresponding $\varphi_\delta(t)$ is increasing t , the other term decreasing in t . Thus $(j \rightarrow k) \in g$ implies that $\lim_{N \rightarrow \infty} \frac{1}{N} c(E_j, E_k) = \varphi_\delta(j\delta)$. We have

$$\begin{aligned} \lim_{N \rightarrow \infty} \frac{1}{N} SP(l) &\underset{\text{by defn}}{\geq} \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{\substack{j \neq l \\ j \in L}} \min_{\substack{k \neq j \\ k \in L}} c(E_j, E_k) &\overset{\text{by Lemma 3 and } \frac{1}{N} \max\{\bar{s}_\alpha, \bar{s}_\beta\} \rightarrow 0 \text{ as } N \rightarrow \infty}{=}& \sum_{\substack{j \neq l \\ j \in L}} \varphi_\delta(j\delta) \\ &\underset{\text{by } \varphi_\delta(l\delta) < \varphi_\delta(i\delta)}{\geq} \sum_{\substack{j \neq l \\ j \in L}} \varphi_\delta(j\delta) &\overset{\text{by Lemmas 3,4}}{=} & \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{(j \rightarrow k) \in g} c(E_j, E_k) &\underset{\text{by defn of } SP(i)}{\geq} & \lim_{N \rightarrow \infty} \frac{1}{N} SP(i). \end{aligned}$$

By (A.15), this shows that for large enough N , if $l \in L$ does not maximize $\varphi_\delta(\cdot)$ then $\mu(E_l) = 0$. So $\mu(E_i) > 0$ must imply that $i \in \arg \max_{j \in L} \varphi_\delta(j\delta)$. □

This characterizes the stochastically stable states for large N . The principle theorem of the paper approximates these states for small δ , linking them to bargaining solutions. To prove the Theorem we use the following lemma.

Lemma 6. Suppose φ is a continuous function which admits a unique maximum. Suppose φ_δ such that φ_δ converges uniformly to φ as $\delta \rightarrow 0$. Let $t^* \in \arg \max \varphi(t)$ and $i^* \in \arg \max_i \varphi_\delta(i\delta)$. Then for all $\zeta > 0$, there exists $\bar{\delta} > 0$ such that for all $\delta < \bar{\delta}$, we have $|i^*\delta - t^*| < \zeta$.

Proof. By the definitions of $t^* \in \arg \max_t \varphi(t)$ and $i^* \in \arg \max_i \varphi_\delta(i\delta)$, we have $\varphi(t^*) \geq \varphi(i^*\delta)$ and $\varphi_\delta(i^*\delta) \geq \varphi_\delta(t)$. Let $\zeta > 0$. By uniform convergence we can choose $\delta < \bar{\delta}$, such that $|\varphi_\delta(t^*) - \varphi(t^*)| < \zeta$ and $|\varphi_\delta(i^*\delta) - \varphi(i^*\delta)| < \zeta$. For $\delta < \bar{\delta}$, we have $\varphi(i^*\delta) \leq \varphi(t^*) \leq \varphi_\delta(t^*) + \zeta \leq \varphi_\delta(i^*\delta) + \zeta < \varphi(i^*\delta) + 2\zeta$. Thus we have that

$$\text{For all } \tilde{\zeta} > 0, \text{ there exists } \bar{\delta} \text{ such that for all } \delta < \bar{\delta}, \text{ we have } |\varphi(t^*) - \varphi(i^*\delta)| < \tilde{\zeta}. \tag{A.16}$$

Without loss of generality we suppose that $i^*\delta < t^*$ and let $\zeta > 0$ be given. Then for $\zeta > 0$ we can choose $\bar{\rho}$ such that for all $\rho < \bar{\rho}$

$$\varphi(t^*) - \rho < y < \varphi(t^*) \text{ implies } |\varphi^{-1}(y) - t^*| < \zeta, \tag{A.17}$$

where φ^{-1} is the inverse function for φ defined in a neighborhood of t^* except t^* . Now let $\zeta > 0$. Choose $\bar{\rho}$ satisfying (A.17) first. Then for $\tilde{\zeta} = \rho < \bar{\rho}$, choose $\bar{\delta}$ satisfying (A.16). Then for ρ and for $\delta < \bar{\delta}$, we have $|\varphi(i^*\delta) - \varphi(t^*)| < \rho$. Also since $\rho < \bar{\rho}$, by (A.17) we have $|i^*\delta - t^*| < \zeta$. Thus we show that for all $\zeta > 0$, there exists $\bar{\delta} > 0$ such that for all $\delta < \bar{\delta}$, we have $|i^*\delta - t^*| < \zeta$. □

Proof of Theorem 1. Taking the limit of $\varphi_\delta^U(t)$ and $\varphi_\delta^L(t)$ as $\delta \rightarrow 0$ gives uniform convergence to

$$\varphi^U(t) := \min \left\{ \frac{f(t)}{f(t) + \bar{s}_\beta}, \frac{t}{t + \bar{s}_\alpha} \right\}, \tag{A.18}$$

$$\varphi^L(t) := \min \left\{ t \frac{f(t)}{f(t) + \bar{s}_\beta}, f(t) \frac{t}{t + \bar{s}_\alpha} \right\}, \tag{A.19}$$

respectively. These functions are maximized at t^{KS}, t^L , respectively. Lemmas 5 and 6 then complete the proof for the cases of uniform-unintentional and logit-unintentional deviations.

For the case of logit-intentional deviations, $\varphi_\delta^{LI}(t)$ takes the form $\min\{a(t), b(t)\}$, with $a(t) = t \frac{f(t)}{f(t+\delta) + f(t)}$ and $b(t) = f(t) \frac{t}{2t-\delta}$. Continuity of $f(\cdot)$ implies that there exist $\zeta > 0, \hat{\delta} > 0$ such that for all $\delta < \hat{\delta}, a(t) < b(t)$ for all $\delta \leq t < \zeta$, and $a(t) > b(t)$ for all $\bar{s}_\alpha - \delta \geq t > \bar{s}_\alpha - \zeta$. Therefore, the following function equals $\varphi_\delta^{LI}(t)$ at all $t = \delta i, i = 1, \dots, n - 1$.

$$\hat{\varphi}_\delta^{LI}(t) = \begin{cases} a(t) & \text{if } t < \zeta. \\ \min\{a(t), b(t)\} & \text{if } \zeta \leq t \leq \bar{s}_\alpha - \zeta. \\ b(t) & \text{if } t > \bar{s}_\alpha - \zeta. \end{cases}$$

As $\delta \rightarrow 0, \hat{\varphi}_\delta^{LI}$ converges uniformly to

$$\hat{\varphi}^{LI}(t) := \min \left\{ \frac{t}{2}, \frac{f(t)}{2} \right\}.$$

This expression is maximized at t^E . Lemmas 5 and 6 then complete the proof for logit-intentional deviations.

For the case of uniform-intentional deviations, we cannot apply Lemma 6, since the $\varphi_\delta^{UI}(t)$ does not converge to a function with a unique maximum as $\delta \rightarrow 0$. However, in the minimization that defines $\varphi_\delta^{UI}(t)$, one of the terms is increasing in t , the other is decreasing in t , and they intersect at a unique \tilde{t} . Therefore, for small δ , $i^* \in \arg \max_{i \in L} \varphi_\delta^{UI}(i\delta)$ is close to \tilde{t} , which is given by

$$\frac{\tilde{t}}{2\tilde{t} - \delta} = \frac{f(\tilde{t})}{f(\tilde{t} + \delta) + f(\tilde{t})} \Leftrightarrow \tilde{t} \frac{f(\tilde{t} + \delta) - f(\tilde{t})}{\delta} + f(\tilde{t}) = 0,$$

which approaches the first order condition for t^{NB} as $\delta \rightarrow 0$. Hence $\delta i^* \rightarrow t^{NB}$ (See detailed argument in Naidu et al., 2010). □

Appendix B. The Logit bargaining solution

The definition of t^L in Table 2 can be rewritten as

$$t^L = \arg \max_{0 \leq t \leq \bar{s}_\alpha} \min\{h_1(t), h_3(t)\}, \quad h_1(t) := \frac{tf(t)}{f(t) + \bar{s}_\beta}, \quad h_3(t) := \frac{tf(t)}{t + \bar{s}_\alpha}. \tag{B.1}$$

We denote the maximizers of $h_1(t), h_3(t)$ by t_1, t_3 respectively.

$$t_l := \arg \max_{0 \leq t \leq \bar{s}_\alpha} h_l(t), \quad l = 1, 3.$$

When $h_1(t)$ and $h_3(t)$ intersect for $0 \leq t \leq \bar{s}_\alpha$, that is for $\frac{1}{2} \leq \frac{\bar{s}_\alpha}{\bar{s}_\beta} \leq 2$, we let t_2 be the value of t for which this intersection occurs. That is, t_2 solves

$$t_2 + \bar{s}_\alpha = f(t_2) + \bar{s}_\beta. \tag{B.2}$$

Remark 1. The Logit bargaining solution solves

$$t^L := \begin{cases} t_1 & \text{if } h_1(t_1) < h_3(t_1), \quad (\text{Case 1}) \\ t_3 & \text{if } h_3(t_3) < h_1(t_3), \quad (\text{Case 3}) \\ t_2 & \text{otherwise.} \quad (\text{Case 2}) \end{cases}$$

The cases of the solution are numbered by the order in which they occur as the ratio $\frac{\bar{s}_\alpha}{\bar{s}_\beta}$ moves from low to high values. For low values of $\frac{\bar{s}_\alpha}{\bar{s}_\beta}$, the maximum of $h_1(\cdot)$ lies underneath the curve of $h_3(\cdot)$. This is when Case 1 holds. For high values of $\frac{\bar{s}_\alpha}{\bar{s}_\beta}$, the maximum of $h_3(\cdot)$ lies underneath the curve of $h_1(\cdot)$ and we are in Case 3. For intermediate values of $\frac{\bar{s}_\alpha}{\bar{s}_\beta}$, the maximizer of (B.1) is determined by the intersection of $h_1(\cdot)$ and $h_3(\cdot)$ and we are in Case 2.

In Case 1 and Case 3, the Logit solution is similar to the Nash solution, but adjusted to take into account the best possible outcome for one of the players. Comparing the first order condition for the Nash bargaining solution:

$$t^{NB} f'(t^{NB}) + f(t^{NB}) = 0$$

to the first order conditions for the Logit bargaining solution in Cases 1 and 3 respectively:

$$t^L f'(t^L) + \frac{\bar{s}_\beta + f(t^L)}{\bar{s}_\beta} f(t^L) = 0, \quad t^L f'(t^L) + \frac{\bar{s}_\alpha}{t^L + \bar{s}_\alpha} f(t^L) = 0.$$

We see that Player α obtains more in Case 1 and less in Case 3 than he does under the Nash solution. Moreover, in Case 3, an increase in \bar{s}_α results in Player α achieving a higher payoff. This increase of Player α 's payoff in his best possible payoff differs from the similar effect in the Kalai–Smorodinsky solution. The effect in the latter depends on the ratio of \bar{s}_α and \bar{s}_β , whereas in Case 3 of the Logit solution, changes in \bar{s}_β have no direct effect. Symmetrically, in Case 1 the solution depends on $f(\cdot)$ and \bar{s}_β , but not directly on \bar{s}_α .

When conditions for Case 2 are satisfied, we see from Equation (B.2) and the illustration in Fig. 1 that Player γ 's payoff decreases with \bar{s}_γ . In fact, the solution is an Equal-loss solution (Chun, 1988) with an ideal point of $(\bar{s}_\beta, \bar{s}_\alpha)$, effectively an Egalitarian solution with a notional disagreement point of $(\bar{s}_\beta, \bar{s}_\alpha)$. The players equalize their losses from this ideal point. This ideal point for a player is equal to the maximum attainable payoff of the other player (see Fig. 1).

Table B.6
Explicit expressions for the Logit bargaining solution when the frontier is linear.

Case	Condition	Solution
1	$\bar{s}_\alpha < \left(\frac{3\sqrt{2}}{2} - 1\right)^{-1} \bar{s}_\beta$	$t^L = (2 - \sqrt{2})\bar{s}_\alpha$
2	$\left(\frac{3\sqrt{2}}{2} - 1\right)^{-1} \bar{s}_\beta \leq \bar{s}_\alpha \leq \left(\frac{3\sqrt{2}}{2} - 1\right) \bar{s}_\beta$	$t^L = \frac{(2\bar{s}_\beta - \bar{s}_\alpha)\bar{s}_\alpha}{\bar{s}_\alpha + \bar{s}_\beta}$
3	$\bar{s}_\alpha > \left(\frac{3\sqrt{2}}{2} - 1\right) \bar{s}_\beta$	$t^L = (\sqrt{2} - 1)\bar{s}_\alpha$

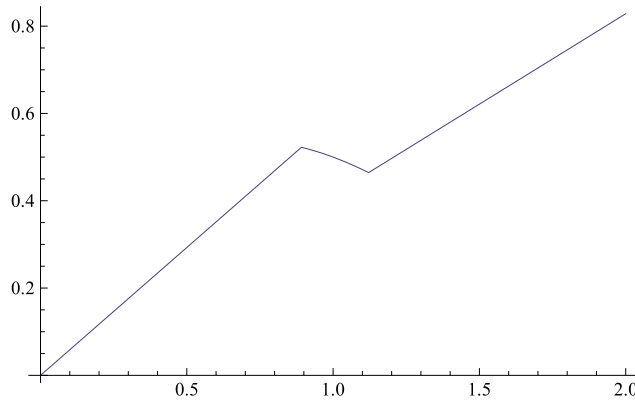


Fig. B.6. t^L by \bar{s}_α , keeping $\bar{s}_\beta = 1$.

Now consider a linear bargaining frontier given by the equation $f(t) = \bar{s}_\beta - t \frac{\bar{s}_\beta}{\bar{s}_\alpha}$. Conditions under which each case of the solution pertains and explicit solutions for each case are given in Table B.6. An increase in \bar{s}_α is equivalent to a stretch of the bargaining frontier parallel to the horizontal axis. It can be seen that when Case 2 pertains, an increase in \bar{s}_α results in a reduction in t^L . Fig. B.6 shows how, fixing \bar{s}_β , the payoff of Player α varies with \bar{s}_α .

Appendix C. Experiments – instructions and interface

C.1. Decision screen faced by subjects

Here we give the decision screen faced by subjects from the second round onwards. *Position 1* corresponds to the population of α -subjects and *Position 2* corresponds to the population of β -subjects. Subjects were informed of their own payoffs from successful coordination and the proportions of subjects in the other position who played each strategy in the preceding round. See Fig. C.7.

C.2. Instructions given to participants

INSTRUCTIONS

Welcome to the study. In the following two hours, you will participate in 200 rounds of strategic decision making. Please read these instructions carefully; the cash payment you will receive at the end of the study depends on how well you perform so it is important that you understand the instructions. If you have a question at any point, please raise your hand and wait for one of us to come over. We ask that you turn off your mobile phone and any other electronic devices. Communication of any kind with other participants is not allowed.

Your Cash Payment

For each participant, the experimenter *randomly* and *independently* selects 3 rounds to calculate the cash payment. (So it is in your best interest to take each round seriously.) Each round has an equal chance to be selected as a payment round for you. You will **not** be told which rounds are chosen to be the payment rounds for you until the end of the session. Your total cash payment at the end of the experiment will be the average earnings in the three selected rounds (translated into HKD as the exchange rate of 1 Token = 1 HKD) plus a 30 HKD show-up fee.

$$\text{Your total cash payment} = \text{HK\$ (The average of earnings in the 3 selected rounds)} + \text{HK\$ 30}$$

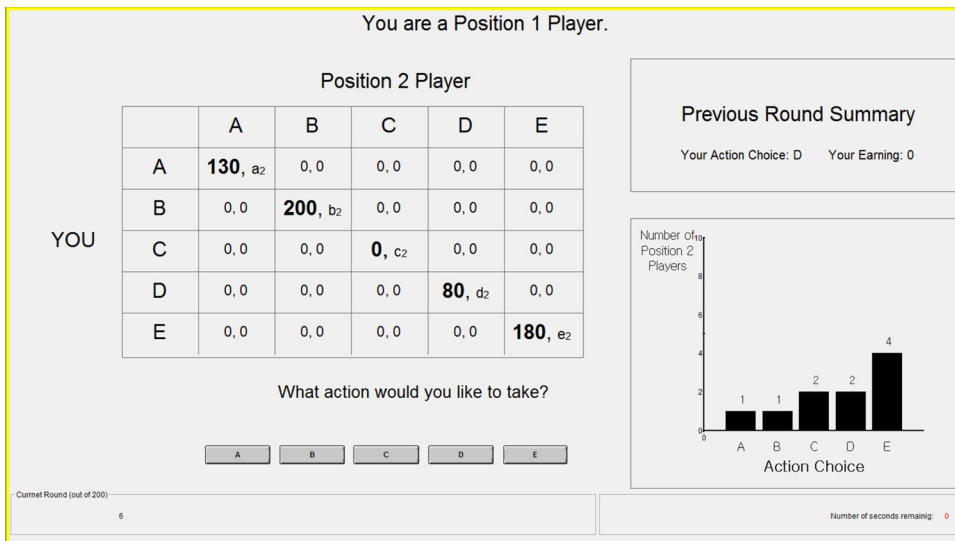


Fig. C.7. Screen faced by subjects in Position 1 (α -subjects).

Your Role and Decision Group

You are one of 20 participants in today’s session. At the beginning of the experiment, one half of the participants will be randomly assigned to be in Position 1 and the other half to be in Position 2. Your position will remain fixed throughout the experiment. In each round, all individuals are *randomly* paired so that each pair comprises one Position 1 player and one Position 2 player. Thus, in a round you will have an equal, 1 in 10 chance of being paired with any given participant in the other position. You will not be told the identity of the participant you are paired with in any round, nor will that participant be told your identity—even after the end of the experiment. Participants will be randomly re-paired after each round to form new pairs.

Your Decision in Each Round

In each round, you will play a 2-player game with the participant you are paired with. For each player, there are five available actions, labeled A, B, C, D, and E. You and the participant you are paired with simultaneously choose an action, and **only if the choices made by you and the other participant are the same, may you be able to get a positive earning** in the round.

Figure 1: Your Earnings

		Position 2 Player				
		A	B	C	D	E
Position 1 Player	A	a_1, a_2	0,0	0,0	0,0	0,0
	B	0,0	b_1, b_2	0,0	0,0	0,0
	C	0,0	0,0	c_1, c_2	0,0	0,0
	D	0,0	0,0	0,0	d_1, d_2	0,0
	E	0,0	0,0	0,0	0,0	e_1, e_2

In words this says,

- When you are a Position 1 player, if you and the other participant you are paired with have different actions, you each get 0. If you and the other player in your pair both choose
 - Action ‘A’, you get a_1 and the other gets a_2 ,
 - Action ‘B’, you get b_1 and the other gets b_2 ,
 - Action ‘C’, you get c_1 and the other gets c_2 ,
 - Action ‘D’, you get d_1 and the other gets d_2 , and
 - Action ‘E’, you get e_1 and the other gets e_2 .
- When you are a Position 2 player, if you and the other participant you are paired with have different actions, you each get 0. If you and the other player in your pair both choose

- (a) Action 'A', you get a_2 and the other gets a_1 ,
- (b) Action 'B', you get b_2 and the other gets b_1 ,
- (c) Action 'C', you get c_2 and the other gets c_1 ,
- (d) Action 'D', you get d_2 and the other gets d_1 , and
- (e) Action 'E', you get e_2 and the other gets e_1 .

You are prompted to choose an action by clicking one of the five buttons *A*, *B*, *C*, *D*, and *E* at the bottom of your screen. Your decision in the round is completed.

Do You Know Your Payoffs?

At the beginning of the first round, you will be assigned to a position. Then the payoff values, a_1 , b_1 , c_1 , d_1 , and e_1 , are revealed to the position 1 players and the payoff values, a_2 , b_2 , c_2 , d_2 , and e_2 , are revealed to the position 2 players. However, you will **not** be told the payoff values for the players in the other position, even after the end of the experiment.

An Opportunity to Change Your Action

In the first round, every participant is given the opportunity to choose an action out of the five available ones. In any round from the second round, the opportunity to change an action is given to each participant *independently* with **90%** chance only. That is, with 10% chance, any given participant is not allowed to change his/her action. In case that the opportunity to change your action is not given, you will see the following message in your decision screen:

“In this round, you are not given the opportunity to change your action choice.”

and will be assigned the same action as in the previous round. You will not be told whether the opportunity to change an action is given to the participant you are paired with, nor will the participant you are paired with be told whether such an opportunity is given to you.

Information Feedback

In each round, at the right-bottom corner of your screen, you will see the summary of the **previous round**. First, you will see **your action choice and your earning** from the previous round. Second, you will see the bar chart that reports how many people among the 10 participants in the other position chose each action in the previous round.

Rundown of the Study

1. At the beginning of the first round, you will be assigned to a position, and you will be shown the payoff values for yourself. In the main panel of your decision screen, you will be prompted to enter your choice of action. You must choose one of five actions *A*, *B*, *C*, *D*, and *E* within 30 seconds. If you do not choose an action, one action will be *randomly* assigned to you.
2. The first round is over after everybody has chosen an action. The screen will then show you a summary: (a) your choice of action in the first round, (b) your earning in the first round, and (c) how many players in the other position had each action in the first round.
3. You will be prompted to enter your choice of action for the second round, if you have the opportunity to change your action. The game does not change, so as before you must choose one of five actions. All future rounds are identical to the first round except for two **important** difference.
 - (a) The first difference concerns how much time you have to choose an action. In rounds 2–10, you have 20 seconds to make a decision. If you do not make a decision within the 20 second window, then you will be assigned whatever action you used in the previous round. For rounds 11–200, you have only 10 seconds in which to make a decision. Again, if you fail to choose an action in this timeframe, you will be assigned the same action as in the previous round.
 - (b) The second difference concerns whether the opportunity to change your action is given with 100% chance (round 1) or with 90% chance (rounds 2–200).

Administration

Your decisions as well as your cash payment will be kept completely confidential. Remember that you have to make your decisions entirely on your own; do not discuss your decisions with any other participants.

Upon completion of the study, you will receive your cash payment. You will be asked to sign your name to acknowledge your receipt of the payment. You are then free to leave.

If you have any questions, please raise your hand now. We will answer questions individually. If there are no questions, we will begin with the study.

Quiz

1. Suppose that you are a Position 1 player, and choose action A. It turns out that the participant you are paired with chooses action B. What is your earning? _____
2. Suppose that you are a Position 1 player, and choose action B. It turns out that the participant you are paired with chooses action B. What is your earning? _____

Appendix D. Experiments: additional material

D.1. Figures excluding early periods

Here we present versions of Fig. 5 that exclude some periods. Fig. D.8 excludes the first 50 periods. By excluding the first 50 periods during which convergence to a convention was observed in every session, much data for low best response payoffs is omitted. A similar, though less clear, pattern to Fig. 5 is observed. As well as the initial period of convergence, there were a considerable number of deviations observed in the final 20 periods of Session 3, auguring a possible switch in convention. Fig. D.9 gives Fig. 5 with both the first 50 and the last 20 periods omitted.

D.2. Regression

Letting i denote an individual subject, for periods $t = 2, \dots, 200$, let π_{it}^{BR} be the expected payoff to subject i from playing a best response to the strategy profile played in period $t - 1$. Similarly, let π_{it}^{NBR} be the highest expected payoff from any strategy that is not a best response. Using expression (1), it can be shown that, for small values of η , the probability of a deviation at period t by player i under the logit choice rule is approximately proportional to $\exp(-\eta^{-1}(\pi_{it}^{BR} - \pi_{it}^{NBR}))$.

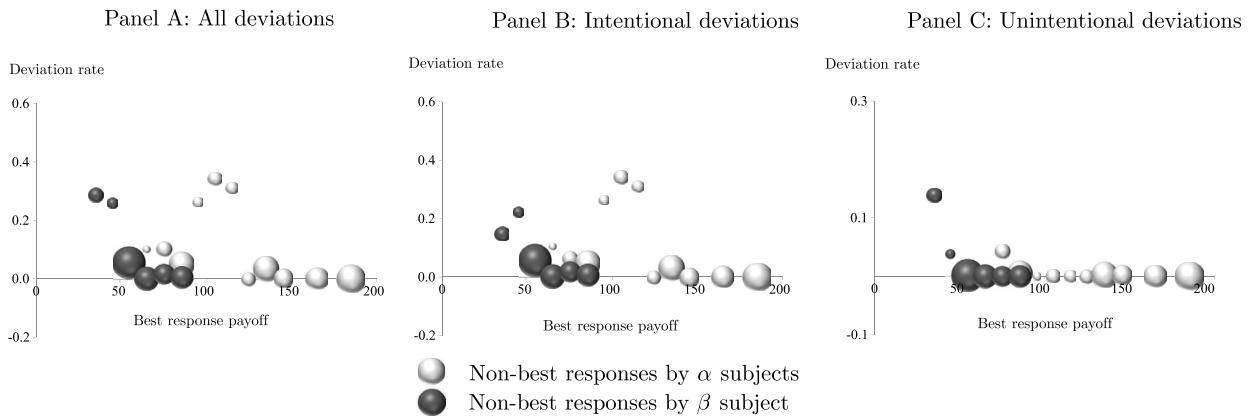


Fig. D.8. Rates of non-best response play grouped by expected payoff from the best response strategy for periods 51 to 200. Data are grouped into bins by best response payoff (0 to 10, 10 to 20, etc.). The area of each circle on the chart is proportional to the square root of the number of strategy updating opportunities it represents.

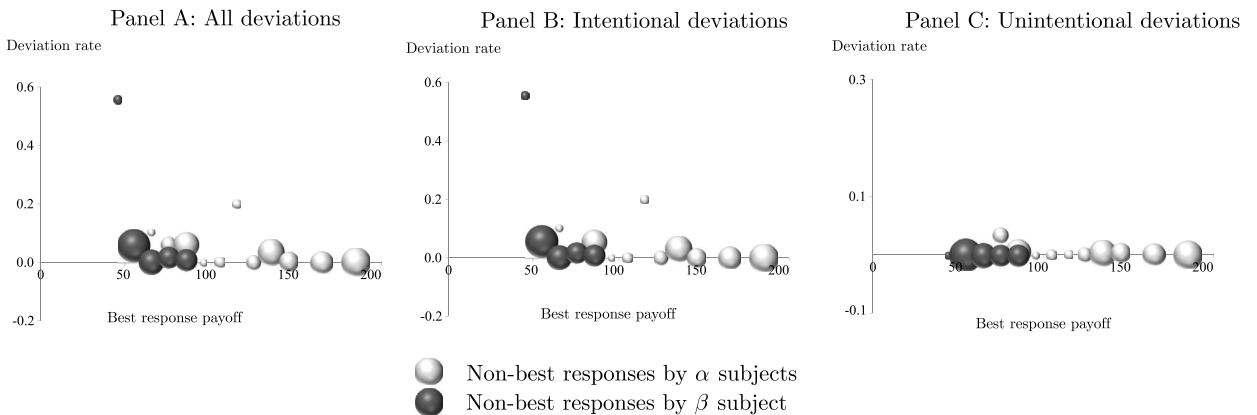


Fig. D.9. Rates of non-best response play grouped by expected payoff from the best response strategy for periods 51 to 180. Data are grouped into bins by best response payoff (0 to 10, 10 to 20, etc.). The area of each circle on the chart is proportional to the square root of the number of strategy updating opportunities it represents.

Table D.7
Logit regression with individual level fixed effect.

Revision opportunity	Session	# of Obs.	(1)		(2)	
			$U(\text{BR}) - U(\text{NBR})$ Coef.	p -V	Period Coef.	p -V
90%	1	3,591	-.0040936	.000	-.0135664	.000
	2	3,585	-.0038834	.000	-.0231305	.000
	3	3,381	-.0060386	.000	-.0050871	.000
50%	4	1,103	-.005016	.000	-.0185845	.000
	5	1,616	-.0031452	.000	-.0177666	.000

Note: Observations from 9 individuals (895 observations) and 4 individuals (402 observations) respectively in Session 4 and Session 5 are dropped because of no variation in their choices. Abbreviations: # of Obs. = Number of Observations; $U(\text{BR})$ = Payoff from the best response action; $U(\text{nBR})$ = Payoff from the second best action; Coef. = Coefficient; p -V = p -value.

Define the binary variable M_{it} as follows. $M_{it} = 1$ if subject i in period t does not play a best response to the strategies played in period $t - 1$, and $M_{it} = 0$ otherwise. A multinomial logit regression with individual fixed effects was carried out with dependent variable M_{it} and independent variables $\pi_{it}^{\text{BR}} - \pi_{it}^{\text{NBR}}$ and t , the latter to account for a possible change in deviation rates due to the passage of time. We label the coefficients of these variables β_1 , β_2 respectively.

$$Pr[M_{it} = 1] = \exp(\beta_1 (\pi^{\text{BR}} - \pi^{\text{NBR}}) + \beta_2 t) \quad (\text{D.1})$$

The sign of each coefficient is easy to interpret. If deviations are less frequent when more payoff is lost as a consequence, then β_1 should be negative. If subjects deviate less as a session progresses, then β_2 should be negative. Table D.7 presents the results. From Column (1), we see that β_1 is negative and significant at the 1% level across all sessions. Column (2) shows that across all sessions, subjects deviate less often as time proceeds. Estimated marginal effects can be determined as follows. The coefficient -0.0040936 in Session 1 indicates that, holding t fixed, a one unit decrease in $\pi^{\text{BR}} - \pi^{\text{NBR}}$ leads to an increase in the probability of a deviation by a factor of $\exp(0.0040936) = 1.004102$.

References

- Agastya, M., 1999. Perturbed adaptive dynamics in coalition form games. *J. Econ. Theory* 89, 207–233.
- Alexander, J., Skyrms, B., 1999. Bargaining with neighbors: is justice contagious? *J. Philos.* 96, 588–598.
- Alós-Ferrer, C., Netzer, N., 2010. The logit-response dynamics. *Games Econ. Behav.* 68, 413–427.
- Beggs, A., 2005. Waiting times and equilibrium selection. *Econ. Theory* 25, 599–628. <https://doi.org/10.1007/s00199-003-0444-6>.
- Binmore, K.G., 1998. *Game Theory and the Social Contract: Just Playing*, vol. 2. MIT Press.
- Binmore, K., 2005. *Natural Justice*. Oxford University Press.
- Binmore, K., Samuelson, L., Young, P., 2003. Equilibrium selection in bargaining models. *Games Econ. Behav.* 45, 296–328. [https://doi.org/10.1016/S0899-8256\(03\)00146-5](https://doi.org/10.1016/S0899-8256(03)00146-5).
- Block, H., Marschak, J., 1959. *Random Orderings and Stochastic Theories of Response*. Technical Report. Cowles Foundation for Research in Economics, Yale University.
- Blume, L.E., 1993. The statistical mechanics of strategic interaction. *Games Econ. Behav.* 5, 387–424. <https://doi.org/10.1006/game.1993.1023>.
- Blume, L.E., 1996. *Population Games*. Working Papers 96-04-022. Santa Fe Institute.
- Bowles, S., 2005. Is inequality a human universal? In: Barrett, C.B. (Ed.), *The Social Economics of Poverty*. Routledge, pp. 125–145.
- Bowles, S., 2006. Institutional poverty traps. In: Bowles, S., Durlauf, S.N., Hoff, K. (Eds.), *Poverty Traps*. Princeton University Press, pp. 116–138.
- Cao, X., 1982. Preference functions and bargaining solutions. In: 1982 21st IEEE Conference on Decision and Control. IEEE, pp. 164–171.
- Chun, Y., 1988. The equal-loss principle for bargaining problems. *Econ. Lett.* 26, 103–106.
- Cox, D.R., 1958. The regression analysis of binary sequences. *J. R. Stat. Soc. B Methodol.* 20, 215–242.
- Fischbacher, U., 2007. z-tree: Zurich toolbox for ready-made economic experiments. *Exp. Econ.* 10, 171–178.
- Freidlin, M.I., Wentzell, A.D., 1984. *Random Perturbations of Dynamical Systems*. Springer. ISBN 9780387983622. 430 pp. 2nd ed., 1998.
- Hwang, S.H., Newton, J., 2014. *A Classification of Bargaining Solutions by Evolutionary Origin*. Working Papers 2014-02. University of Sydney, School of Economics.
- Hwang, S.H., Newton, J., 2017. Payoff-dependent dynamics and coordination games. *Econ. Theory* 64, 589–604. <https://doi.org/10.1007/s00199-016-0988-x>.
- Hwang, S.H., Rey-Bellet, L., 2017. Positive feedback in coordination games: stochastic evolutionary dynamics and the logit choice rule. arXiv:1701.04870.
- Kalai, E., 1977. Proportional solutions to bargaining situations: interpersonal utility comparisons. *Econometrica* 45, 1623–1630.
- Kalai, E., Smorodinsky, M., 1975. Other solutions to Nash's bargaining problem. *Econometrica* 43, 513–518.
- Kandori, M., Mailath, G.J., Rob, R., 1993. Learning, mutation, and long run equilibria in games. *Econometrica* 61, 29–56.
- Klaus, B., Newton, J., 2016. Stochastic stability in assignment problems. *J. Math. Econ.* 62, 62–74. <https://doi.org/10.1016/j.jmateco.2015.11.002>.
- Lim, W., Neary, P.R., 2016. An experimental investigation of stochastic adjustment dynamics. *Games Econ. Behav.* 100, 208–219.
- Marden, J.R., Young, H.P., Pao, L.Y., 2014. Achieving Pareto optimality through distributed learning. *SIAM J. Control Optim.* 52, 2753–2770.
- Mäs, M., Nax, H.H., 2016. A behavioral study of “noise” in coordination games. *J. Econ. Theory* 162, 195–208. <https://doi.org/10.1016/j.jet.2015.12.010>.
- Naidu, S., Hwang, S.H., Bowles, S., 2010. Evolutionary bargaining with intentional idiosyncratic play. *Econ. Lett.* 109, 31–33. <https://doi.org/10.1016/j.econlet.2010.07.005>.
- Nash, J., 1953. Two-person cooperative games. *Econometrica* 21, 128–140.
- Nash, John F. J., 1950. The bargaining problem. *Econometrica* 18, 155–162.
- Nax, H.H., Burton-Chellew, M.N., West, S.A., Young, H.P., 2016. Learning in a black box. *J. Econ. Behav. Organ.* 127, 1–15.
- Nax, H.H., Pradelski, B.S.R., 2015. Evolutionary dynamics and equitable core selection in assignment games. *Int. J. Game Theory* 44, 903–932. <https://doi.org/10.1007/s00182-014-0459-1>.

- Neary, P.R., 2012. Competing conventions. *Games Econ. Behav.* 76, 301–328.
- Newton, J., 2012a. Coalitional stochastic stability. *Games Econ. Behav.* 75, 842–854. <https://doi.org/10.1016/j.geb.2012.02.014>.
- Newton, J., 2012b. Recontracting and stochastic stability in cooperative games. *J. Econ. Theory* 147, 364–381. <https://doi.org/10.1016/j.jet.2011.11.007>.
- Newton, J., 2018. Evolutionary game theory: a renaissance. *Games* 9, 31. <https://doi.org/10.3390/g9020031>.
- Newton, J., Sawa, R., 2015. A one-shot deviation principle for stability in matching problems. *J. Econ. Theory* 157, 1–27. <https://doi.org/10.1016/j.jet.2014.11.015>.
- Pradelski, B.S., Young, H.P., 2012. Learning efficient Nash equilibria in distributed systems. *Games Econ. Behav.* 75, 882–897.
- Rachmilevitch, S., 2015. The Nash solution is more utilitarian than egalitarian. *Theory Decis.* 79, 463–478.
- Rachmilevitch, S., 2016. Egalitarian–utilitarian bounds in Nash's bargaining problem. *Theory Decis.* 80, 427–442.
- Sandholm, W.H., 2010. *Population Games and Evolutionary Dynamics*. Economic Learning and Social Evolution. MIT Press, Cambridge, Mass.
- Shiryayev, A.N., 1995. *Probability*, 2nd ed.. Springer-Verlag, New York, Inc., Secaucus, NJ, USA.
- Young, H.P., 1993a. The evolution of conventions. *Econometrica* 61, 57–84.
- Young, H.P., 1993b. An evolutionary model of bargaining. *J. Econ. Theory* 59, 145–168. <https://doi.org/10.1006/jeth.1993.1009>.
- Young, H.P., 1998a. Conventional contracts. *Rev. Econ. Stud.* 65, 773–792.
- Young, H.P., 1998b. *Individual Strategy and Social Structure*. Princeton University Press.