# Supplementary Appendices for
# "Conventions under heterogeneous behavioral rules"

Jonathan Newton*

August 24, 2020

## D. Proofs for imitative choice

This section formalizes the arguments of Section 5 of the main manuscript. The proofs of Propositions 2 and 3 proceed by considering weak asymmetry and supermodularity as defined in Appendix B. First, we define a large class of imitative rules and show conditions under which weak asymmetry holds for the entire class. Second, we show supermodularity of condition dependent and imitate-the-best rules under these conditions. Weak asymmetry together with supermodularity then implies asymmetry by Lemma 4.

### D.1 Formal definition of imitative rules

Let $C \subseteq V$ be player $i$'s *comparison set*. When player $i$ considers changing his strategy, his switching probability will depend on the current payoffs of the players in his comparison set. Define a function $h^C : \{S : S \subseteq C\} \times \mathbb{R}^V \to \mathbb{R}$ such that, for given $S \subseteq C$,

$$h^C(S,x) \text{ is } \begin{cases} \text{non-decreasing in } x_j & \text{if } j \in S, \\ \text{non-increasing in } x_j & \text{if } j \in C \setminus S, \\ \text{constant in } x_j & \text{if } j \notin C. \end{cases}$$

Using this function, we define a statistic $\Delta_i^\sigma$ that measures, at strategy profile $\sigma$, how well players in $C$ who play the same strategy as player $i$ perform relative to players who play the alternative strategy.

$$\Delta_i^\sigma := h^C\left(V_{\sigma(i)}(\sigma) \cap C\,,\, (U_j(\sigma))_{j\in V}\right).$$

$\Delta_i^\sigma$ is non-decreasing in the payoffs of players in $C$ who play the same strategy as player $i$, non-increasing in the payoffs of players in $C$ who play a different strategy to player $i$, and constant in the payoff of players outside of $C$.

*Institute of Economic Research, University of Kyoto, Yoshida-honmachi, Sakyo-ku, Kyoto 606-8501. e-mail: newton@kier.kyoto-u.ac.jp; website: jonathannewton.net.

Let an *imitative* rule for player $i$ be defined as follows. For non-decreasing $\Upsilon_i^{Im}:\mathbb{R}\to\mathbb{R}_+$ and constant $d_i^\sigma\in(0,1)$, $\sigma\in\Sigma$, let

$$\text{(D.1)}\qquad P^{\varepsilon}_{\{i\}}(\sigma,\sigma)=1-d_i^\sigma\varepsilon^{\Upsilon_i^{Im}(\Delta_i^\sigma)}\quad\text{and}\quad P^{\varepsilon}_{\{i\}}(\sigma,\sigma^{(i)})=d_i^\sigma\varepsilon^{\Upsilon_i^{Im}(\Delta_i^\sigma)},$$

with the convention that $0^0=1$ so that $P^{\varepsilon}_{\{i\}}$ is continuous in $\varepsilon$ at $\varepsilon=0$. Such rules satisfy the restriction on behavior that the probability that a strategy is chosen is non-decreasing in the payoffs of those who currently play that strategy.

## D.2 Weak asymmetry of imitative rules

For readability, in this section and the remainder of this appendix we write $U(\sigma):=(U_j(\sigma))_{j\in V}$.

**Lemma 7.** *If player i follows an imitative rule, A is $PD_{jk}$ and $MM_{jk}$ for all $j\in C$, k, then $c_{\{i\}}(\cdot,\cdot)$ is weakly asymmetric towards A.*

*Proof of Lemma 7.* Let $\sigma$, $\hat\sigma$ be such that $V_A(\sigma)=V_B(\hat\sigma)$, $\sigma(i)=A$. Note that

$$\text{(D.2)}\qquad V_{\sigma(i)}(\sigma)=V_A(\sigma)=V_B(\hat\sigma)=V_{\hat\sigma(i)}(\hat\sigma).$$

Consider the elements of $U(\sigma)$,

$$\text{(D.3)}\qquad U_j(\sigma)=\begin{cases}\sum_{k\in V_A(\sigma)\setminus\{j\}}u_{jk}(A,A)+\sum_{k\in V_B(\sigma)\setminus\{j\}}u_{jk}(A,B) & \text{if }\sigma(j)=A,\\ \sum_{k\in V_A(\sigma)\setminus\{j\}}u_{jk}(B,A)+\sum_{k\in V_B(\sigma)\setminus\{j\}}u_{jk}(B,B) & \text{if }\sigma(j)=B,\end{cases}$$

and the elements of $U(\hat\sigma)$,

$$\text{(D.4)}\qquad U_j(\hat\sigma)=\begin{cases}\sum_{k\in V_A(\hat\sigma)\setminus\{j\}}u_{jk}(A,A)+\sum_{k\in V_B(\hat\sigma)\setminus\{j\}}u_{jk}(A,B) & \\ =\sum_{k\in V_B(\sigma)\setminus\{j\}}u_{jk}(A,A)+\sum_{k\in V_A(\sigma)\setminus\{j\}}u_{jk}(A,B) & \text{if }\hat\sigma(j)=A,\\ \sum_{k\in V_A(\hat\sigma)\setminus\{j\}}u_{jk}(B,A)+\sum_{k\in V_B(\hat\sigma)\setminus\{j\}}u_{jk}(B,B) & \\ =\sum_{k\in V_B(\sigma)\setminus\{j\}}u_{jk}(B,A)+\sum_{k\in V_A(\sigma)\setminus\{j\}}u_{jk}(B,B) & \text{if }\hat\sigma(j)=B.\end{cases}$$

By (D.2), if $\sigma(j)=A$, then $\hat\sigma(j)=B$, and if $\sigma(j)=B$, then $\hat\sigma(j)=A$. Consequently, (D.3) and (D.4), together with $\text{PD}_{jk}$ ($u_{jk}(A,A)\geq u_{jk}(B,B)$) and $\text{MM}_{jk}$ ($u_{jk}(A,B)\geq u_{jk}(B,A)$) for all $j\in C$, $k$ imply

$$\text{(D.5)}\qquad \begin{aligned}&\text{For all } j\in V_A(\sigma)\cap C,\quad U_j(\sigma)\geq U_j(\hat\sigma),\\ &\text{For all } j\in V_B(\sigma)\cap C,\quad U_j(\sigma)\leq U_j(\hat\sigma).\end{aligned}$$

Then

$$\text{(D.6)}\qquad \begin{aligned}\Delta_i^\sigma&=h^C(V_{\sigma(i)}(\sigma)\cap C,U(\sigma)) && [\text{by defn of }\Delta_i^\sigma]\\ &=h^C(V_A(\sigma)\cap C,U(\sigma)) && [\text{by (D.2)}]\end{aligned}$$

$$
\begin{aligned}
&\geq h^C(V_A(\sigma)\cap C, U(\hat{\sigma})) && \text{[by (D.5) and defn of } h^C\text{]}\\
&= h^C(V_B(\hat{\sigma})\cap C, U(\hat{\sigma})) && \text{[by (D.2)]}\\
&= h^C(V_{\hat{\sigma}(i)}(\hat{\sigma})\cap C, U(\hat{\sigma})) && \text{[by (D.2)]}\\
&= \Delta_i^{\hat{\sigma}}. && \text{[by defn of } \Delta_i^{\hat{\sigma}}\text{]}
\end{aligned}
$$

As $\Upsilon_i^{Im}$ is non-decreasing, (D.6) implies that $\Upsilon_i^{Im}(\Delta_i^{\sigma}) \geq \Upsilon_i^{Im}(\Delta_i^{\hat{\sigma}})$ and therefore, by (4.1), $c_{\{i\}}(\sigma, \sigma^{(i)}) \geq c_{\{i\}}(\hat{\sigma}, \hat{\sigma}^{(i)})$. That is, $c_{\{i\}}(\cdot,\cdot)$ is weakly asymmetric, proving Lemma 7.

□

### D.3 Condition dependence

If $C = \{i\}$, then the switching probability for a player $i$ decreases in his current payoff $U_i(\sigma)$ and is independent of the payoffs of the other players. This is known as *condition dependence*.

**Lemma 8.** *If player i follows a condition dependent rule and $u_{ij}(B,B) \geq u_{ij}(B,A)$ for all j, then $c_{\{i\}}(\cdot,\cdot)$ is supermodular towards A.*

*Proof of Lemma 8.* Let $\hat{\sigma}$, $\tilde{\sigma}$ be such that $\hat{\sigma}(i) = \tilde{\sigma}(i) = B$, $V_A(\hat{\sigma}) \subseteq V_A(\tilde{\sigma})$. From (D.3), if $u_{ik}(B,B) \geq u_{ik}(B,A)$ for all $k \neq i$, then $U_i(\hat{\sigma}) \geq U_i(\tilde{\sigma})$. Then,

$$
\text{(D.7)}\qquad
\begin{aligned}
\Delta_i^{\hat{\sigma}} &= h^C(V_{\hat{\sigma}(i)}(\hat{\sigma})\cap C, U(\hat{\sigma})) && \text{[by defn of } \Delta_i^{\hat{\sigma}}\text{]}\\
&= h^C(\{i\}, U(\hat{\sigma})) && \text{[by condition dependence, } C=\{i\}\text{]}\\
&\geq h^C(\{i\}, U(\tilde{\sigma})) && \text{[by } U_i(\hat{\sigma}) \geq U_i(\tilde{\sigma}) \text{ and defn of } h^C\text{]}\\
&= h^C(V_{\tilde{\sigma}(i)}(\tilde{\sigma})\cap C, U(\tilde{\sigma})) && \text{[by condition dependence, } C=\{i\}\text{]}\\
&= \Delta_i^{\tilde{\sigma}}. && \text{[by defn of } \Delta_i^{\tilde{\sigma}}\text{]}
\end{aligned}
$$

As $\Upsilon_i^{Im}$ is non-decreasing, (D.7) implies that $\Upsilon_i^{Im}(\Delta_i^{\hat{\sigma}}) \geq \Upsilon_i^{Im}(\Delta_i^{\tilde{\sigma}})$ and therefore, by (4.1), $c_{\{i\}}(\hat{\sigma}, \hat{\sigma}^{(i)}) \geq c_{\{i\}}(\tilde{\sigma}, \tilde{\sigma}^{(i)})$. That is, $c_{\{i\}}(\cdot,\cdot)$ is supermodular, proving Lemma 8. □

Note that if $A$ is $\text{PD}_{ij}$ and $\text{MM}_{ij}$, then $u_{ij}(B,B) \geq u_{ij}(B,A)$, otherwise (2.2) would be violated. Therefore, the conditions of Lemma 7 imply the condition of Lemma 8. By Lemma 4, weak asymmetry and supermodularity suffice for asymmetry.

*Proof of Proposition 2.* By definition of condition dependence, the process is independent of the payoffs of players other than $i$, therefore $\text{PD}_{ik}$ and $\text{MM}_{ik}$ for all $k \neq i$ suffices for Lemma 7 to imply that $c_{\{i\}}(\cdot,\cdot)$ is weakly asymmetric. Furthermore, $\text{PD}_{ik}$ and $\text{MM}_{ik}$ for all $k \neq i$, together with (2.2) implies the payoff ordering $u_{ik}(A,A) \geq u_{ik}(B,B) \geq u_{ik}(A,B) \geq u_{ik}(B,A)$ for all $k \neq i$. In particular, $u_{ik}(B,B) \geq u_{ik}(B,A)$. Therefore, by Lemma 8, $c_{\{i\}}(\cdot,\cdot)$ is supermodular. Consequently, by Lemma 4, $c_{\{i\}}(\cdot,\cdot)$ is asymmetric. □

### D.4 Imitate the best

Consider a player $i$ whose choice probabilities are a function of the highest payoff obtained amongst all of the players who play $A$ and the highest payoff obtained amongst all of the players who play $B$. To pick out the highest payoff obtained by some player in a set of players $S$, define, for $S \subseteq C$, functions $M^S : \mathbb{R}^V \to \mathbb{R}$,

$$M^S(x) = \max\, \{\underline{h}\} \cup \{x_j : j \in S\},$$

where $\underline{h} \in \mathbb{R}$ is a constant that is independent of $S$. That is, $M^S(x)$ equals the maximum value of $x_j$ for $j \in S$, except in the cases when this maximum is less than $\underline{h}$, or $S$ is empty, in which case $M^S(x)$ equals $\underline{h}$.

If $h^C$ is such that $h^C(S,x)$ can be written as

$$h^C(S,x) = f(M^S(x), M^{C \setminus S}(x)),$$

for a function $f$ that is non-decreasing in its first argument and non-increasing in its second argument, we say the rule is an *imitate-the-best* rule.

Let $\hat{\sigma}$, $\tilde{\sigma}$ be such that $V_A(\hat{\sigma}) \subseteq V_A(\tilde{\sigma})$ and $\tilde{\sigma}(i) = B$ as in the definition of supermodularity. By similar arguments to the case of condition dependence, to ensure that the maximum payoff amongst players who play $B$ is at least as high at $\hat{\sigma}$ as at $\tilde{\sigma}$, we require that $u_{jk}(B,B) \geq u_{jk}(B,A)$ for all $j,k$. Similarly, to ensure that the maximum payoff amongst players who play $A$ is no higher at $\hat{\sigma}$ than at $\tilde{\sigma}$, we require that $u_{jk}(A,A) \geq u_{jk}(A,B)$ for all $j,k$.

**Lemma 9.** *If player $i$ follows an imitate-the-best rule, $u_{jk}(B,B) \geq u_{jk}(B,A)$ and $u_{jk}(A,A) \geq u_{jk}(A,B)$ for all $j \in C$, $k$, then $c_{\{i\}}(\cdot,\cdot)$ is supermodular towards $A$.*

*Proof of Lemma 9.* Let $\hat{\sigma}$, $\tilde{\sigma}$ be such that $\hat{\sigma}(i) = \tilde{\sigma}(i) = B$, $V_A(\hat{\sigma}) \subseteq V_A(\tilde{\sigma})$. Together with (D.3), $u_{jk}(A,A) \geq u_{jk}(A,B)$, $u_{jk}(B,B) \geq u_{jk}(B,A)$ for all $j \in C$, $k$, this implies the following inequalities.

(D.8)
$$\begin{aligned} &\text{For all } j \in V_A(\hat{\sigma}) \cap C, \quad U_j(\hat{\sigma}) \leq U_j(\tilde{\sigma}), \\ &\text{For all } j \in V_B(\tilde{\sigma}) \cap C, \quad U_j(\hat{\sigma}) \geq U_j(\tilde{\sigma}). \end{aligned}$$

Note that, as $V_A(\hat{\sigma}) \subseteq V_A(\tilde{\sigma})$ and $V_B(\tilde{\sigma}) \subseteq V_B(\hat{\sigma})$, (D.8) only relates to $j$ for whom $\hat{\sigma}(j) = \tilde{\sigma}(j)$. Then,

(D.9)
$$\begin{aligned} \Delta_i^{\hat{\sigma}} &= h^C(V_{\hat{\sigma}(i)}(\hat{\sigma}) \cap C, U(\hat{\sigma})) && [\text{by defn of } \Delta_i^{\hat{\sigma}}] \\ &= h^C(V_B(\hat{\sigma}) \cap C, U(\hat{\sigma})) && [\text{as } \hat{\sigma}(i) = B] \\ &= f(M^{V_B(\hat{\sigma}) \cap C}(\hat{\sigma}), M^{V_A(\hat{\sigma}) \cap C}(\hat{\sigma})) && [\text{by defn of } h^C \text{ under imitate-the-best}] \\ &\geq f(M^{V_B(\tilde{\sigma}) \cap C}(\hat{\sigma}), M^{V_A(\hat{\sigma}) \cap C}(\hat{\sigma})) && [\text{as } V_B(\tilde{\sigma}) \subseteq V_B(\hat{\sigma}) \text{ and } f \text{ non-decreasing in first argument}] \\ &\geq f(M^{V_B(\tilde{\sigma}) \cap C}(\hat{\sigma}), M^{V_A(\hat{\sigma}) \cap C}(\tilde{\sigma})) && [\text{by (D.8) and } f \text{ non-increasing in second argument}] \\ &\geq f(M^{V_B(\tilde{\sigma}) \cap C}(\hat{\sigma}), M^{V_A(\tilde{\sigma}) \cap C}(\tilde{\sigma})) && [\text{as } V_A(\hat{\sigma}) \subseteq V_A(\tilde{\sigma}) \text{ and } f \text{ non-increasing in second argument}] \\ &\geq f(M^{V_B(\tilde{\sigma}) \cap C}(\tilde{\sigma}), M^{V_A(\tilde{\sigma}) \cap C}(\tilde{\sigma})) && [\text{by (D.8) and } f \text{ non-decreasing in first argument}] \end{aligned}$$

$$
\begin{aligned}
&= h^C(V_B(\tilde{\sigma}) \cap C, U(\tilde{\sigma})) && \text{[by defn of } h^C \text{ under imitate-the-best]} \\
&= h^C(V_{\tilde{\sigma}(i)}(\tilde{\sigma}) \cap C, U(\tilde{\sigma})) && \text{[as } \tilde{\sigma}(i) = B\text{]} \\
&= \Delta_i^{\tilde{\sigma}}. && \text{[by defn of } \Delta_i^{\tilde{\sigma}}\text{]}
\end{aligned}
$$

As $\Upsilon_i^{Im}$ is non-decreasing, (D.9) implies that $\Upsilon_i^{Im}(\Delta_i^{\hat{\sigma}}) \geq \Upsilon_i^{Im}(\Delta_i^{\tilde{\sigma}})$ and therefore, by (4.1), $c_{\{i\}}(\hat{\sigma}, \hat{\sigma}^{(i)}) \geq c_{\{i\}}(\tilde{\sigma}, \tilde{\sigma}^{(i)})$. That is, $c_{\{i\}}(\cdot,\cdot)$ is supermodular, proving Lemma 9. □

Note that if $A$ is $\text{PD}_{jk}$ and $\text{MM}_{jk}$, then $u_{jk}(B,B) \geq u_{jk}(B,A)$ and $u_{jk}(A,A) \geq u_{jk}(A,B)$, otherwise (2.2) would be violated. So, under the conditions of Lemma 7, Lemma 9 also applies. By Lemma 4, weak asymmetry and supermodularity suffice for asymmetry.

*Proof of Proposition 3.* $\text{PD}_{jk}$ and $\text{MM}_{jk}$ for all $j \in C$, $k$, together with (2.2) implies the payoff ordering $u_{jk}(A,A) \geq u_{jk}(B,B) \geq u_{jk}(A,B) \geq u_{jk}(B,A)$ for all $j \in C$, $k$. In particular, $u_{jk}(A,A) \geq u_{jk}(A,B)$ and $u_{jk}(B,B) \geq u_{jk}(B,A)$. Therefore, by Lemmas 7 and 9, $c_{\{i\}}(\cdot,\cdot)$ is weakly asymmetric and supermodular, so by Lemma 4, $c_{\{i\}}(\cdot,\cdot)$ is asymmetric. □

## E. Further examples of payoff-difference based rules

Here give some further examples of behavioral rules that satisfy the definition of payoff-difference based rules given in Section 4.

### E.0.1 Own-payoff based rules

A player $i$ follows an *own-payoff based* best response rule (Peski, 2010) if, for some strictly increasing function $f : \mathbb{R}_+ \to \mathbb{R}_+$ such that $f(0) = 0$, we have that

(E.1) $$\Upsilon_i(x) = f([x_i]_+).$$

A special case is $f(z) = z$, which again gives best response with log-linear deviations. Another special case is $f(z) = z^2$, in which case we have *best response with log-quadratic deviations*, which for small $\varepsilon$ approximates the probit choice rule in two strategy environments such as the one in the current paper (Dokumaci and Sandholm, 2011).

### E.0.2 Hippocratic rules

A player $i$ follows a *Hippocratic* rule if, for some nonnegative vector $\lambda \in \mathbb{R}_+^V$,

(E.2) $$\Upsilon_i(x) = \lambda \cdot [x]_+,$$

so that the probability of player $i$ changing his strategy is decreasing in a weighted sum of payoff losses when he does so. Unlike the utilitarian rule, any gains in payoff are disregarded. If $\lambda_i = 1$ and $\lambda_j = 0$ for $j \neq i$, this is once again best response with log-linear deviations.

E.0.3 Best response with switching costs

A player $i$ follows best response with *switching costs* and uniform deviations (Norman, 2009) if, for some strictly positive $\delta > 0$,

$$\text{(E.3)} \qquad \Upsilon_i(x) = [sgn(x_i + \delta)]_+,$$

so that, for small $\varepsilon$, player $i$ will rarely change his strategy unless his payoff increases by at least $\delta$ as a consequence.

E.0.4 Disjunction and conjunction

Consider rules similar to (E.3) in that $\Upsilon_i$ takes values on $\{0, 1\}$. These $\Upsilon_i$ are *truth functions* that output a value of 1 if a condition is satisfied and output 0 if it is not satisfied. Another example is

$$\text{(E.4)} \qquad \Upsilon'_i(x) = \begin{cases} 1 & \text{if } \sum_{k \in V}[\text{sgn}(x_k)]_+ > 3, \\ 0 & \text{otherwise.} \end{cases}$$

which corresponds to a process in which, for small $\varepsilon$, player $i$ will rarely change his strategy unless by doing so he harms no more than three players.

Any two truth functions can be combined through logical conjunction, which corresponds to taking the minimum of the functions, or logical disjunction, which corresponds to taking the maximum of the functions. For example, in the case of $\Upsilon_i$ given by (E.3) and $\Upsilon'_i$ given by (E.4), the truth function

$$\text{(E.5)} \qquad \Upsilon^*_i = \max\{\Upsilon_i, \Upsilon'_i\},$$

corresponds to a process in which, for small $\varepsilon$, player $i$ rarely changes his strategy unless, as a consequence, his payoff increases by at least $\delta$ and the payoff of no more than three players decreases. Note that $\Upsilon^*_i$ inherits the non-decreasing property from $\Upsilon_i$, $\Upsilon'_i$. Furthermore, given any set of primitive truth functions, the set of truth functions that can be constructed in this way has a lattice structure with a maximal and minimal element.

## F. Coalitional choice

This section formalizes the arguments of Section 6.1 of the main manuscript. From Theorem 3, we know that asymmetric $c_S$ can arise from independent, simultaneous choice by $i \in S$ who follow rules with asymmetric $c_{\{i\}}$. In this section, we consider choice by $S$ as a coalition and study a coalitional variant of payoff-difference based rules. Let

$$E^\sigma_S = \left(U_j(\sigma^A_S, \sigma_{V \setminus S}) - U_j(\sigma^B_S, \sigma_{V \setminus S})\right)_{j \in V} \in \mathbb{R}^V.$$

That is, starting from profile $\sigma$ and keeping the strategies of players in $V \setminus S$ constant, $(E_S^\sigma)_j$ equals the difference between the payoff of player $j$ when $S$ plays $\sigma_S^A$ and the payoff of player $j$ when $S$ plays $\sigma_S^B$.

Let a *coalitional payoff-difference based* rule for $S$ be a rule that gives the following cost function. For non-decreasing $\Upsilon_S^C(\cdot) : \mathbb{R}^V \to \mathbb{R}_+$,

$$c_S(\sigma, \sigma') = \begin{cases} 0, & \text{if } \sigma' = \sigma, \\ \Upsilon_S^C(-E_S^\sigma), & \text{if } \sigma' = (\sigma_S^A, \sigma_{V \setminus S}) \neq \sigma, \\ \Upsilon_S^C(E_S^\sigma), & \text{if } \sigma' = (\sigma_S^B, \sigma_{V \setminus S}) \neq \sigma, \\ \infty, & \text{otherwise.} \end{cases} \tag{F.1}$$

That is, greater values of $E_S^\sigma$ make it more likely that $S$ will choose $\sigma_S^A$ and less likely that $S$ will choose $\sigma_S^B$. Note that if $S = \{i\}$, then the cost function (F.1) reduces to the cost function (4.1).[1] That is, the individualistic payoff-difference based models of Section 4 are a special case of the models of this section.

## F.1 Examples of coalitional rules

Coalitional versions of the rules in Section 4.2 and Appendix E can be considered. For example, $S$ follows a Hippocratic rule if, for some nonnegative $\lambda \in \mathbb{R}_+^V$,

$$\Upsilon_S(x) = \lambda \cdot [x]_+. \tag{F.2}$$

Under this rule, the probability of $S$ switching to $\sigma_S^A$ depends on a weighted sum of payoff losses relative to when $S$ switches to strategy $\sigma_S^B$. If $\lambda_i = 1$ for all $i \in S$ and $\lambda_j = 0$ for $j \notin S$, then we have a *coalitional logit* rule (Sawa, 2014), which can be understood as the rule that arises when each member of $S$ votes for $S$ to switch to $\sigma_S^A$ or $\sigma_S^B$ according to the (individualistic) logit choice rule based on payoffs at $(\sigma_S^A, \sigma_{V \setminus S})$ and $(\sigma_S^B, \sigma_{V \setminus S})$, with a switch being implemented only if the vote is unanimously in favor. This rule is self-regarding in the following sense.

*Definition 10.* A rule $\Upsilon_S$ is *self-regarding* if $\Upsilon_S(x) = f(x_S)$ for some non-decreasing function $f : \mathbb{R}^S \to \mathbb{R}_+$.

A class of rules that only makes sense in a non-individualistic setup is the class of *coalitional stochastic stability* rules (Newton, 2012), where the likelihood of strategic change by coalition $S$ depends on the size of $S$. For example, if, for some constant $\kappa \in \mathbb{R}_{++}$, nonnegative $\lambda \in \mathbb{R}_+^V$,

$$\Upsilon_S(x) = \kappa |S| + [\lambda \cdot x]_+, \tag{F.3}$$

then we have an augmented utilitarian rule in which the larger a coalition is, the less likely it is to change its strategies.

[1] To see this, observe that if $S = \{i\}$, then when $\sigma(i) = A$, $D_i^\sigma = E_S^\sigma$ and when $\sigma(i) = B$, $D_i^\sigma = -E_S^\sigma$.

## F.2 Asymmetry of coalitional rules

When it comes to conditions for asymmetry, the differences between coalitional and individualistic payoff-difference based rules can be concisely explained. First, consider $i \in S$, $j \notin S$. Note that the strategy of player $i$ affects the payoff of players $i$ and $j$ in exactly the same way as it would if player $i$ were updating his strategy as an individual. This creates the need for risk dominance and altruistic risk dominance conditions similar to those of Proposition 1.

*Definition 11.* Strategy $A$ is $\mathrm{RD}_{iT}$ (risk dominant for $i$ against $T$) if

$$\sum_{j \in T \setminus \{i\}} u_{ij}(A,A) + u_{ij}(A,B) \geq \sum_{j \in T \setminus \{i\}} u_{ij}(B,A) + u_{ij}(B,B);$$

$\mathrm{ARD}_{Sj}$ (altruistically risk dominant for $S$ against $j$) if

$$\sum_{i \in S \setminus \{j\}} u_{ji}(A,A) + u_{ji}(B,A) \geq \sum_{i \in S \setminus \{j\}} u_{ji}(A,B) + u_{ji}(B,B).$$

Our previous risk dominance condition summed over all $j \neq i$. Now, the relevant summation is over players outside of $S$, that is $T = V \setminus S$ in the definition above. $\mathrm{ARD}_{Sj}$ simply aggregates $\mathrm{ARD}_{ij}$ over all $i \in S$.

Second, note that there is an additional consideration present in the coalitional case, which is the payoff that players in $S$ obtain from interacting with one another. When players in $S$ all play $A$, interaction between $i, j \in S$ will generate payoff of $u_{ij}(A,A)$ for player $i$ and $u_{ji}(A,A)$ for player $j$. When players in $S$ all play $B$, these payoffs will be $u_{ij}(B,B)$ and $u_{ji}(B,B)$ respectively. Consequently, to ensure that within-coalition incentives to play $A$ outweigh within-coalition incentives to play $B$, we require a payoff dominance condition.

*Definition 12.* Strategy $A$ is $\mathrm{PD}_{iS}$ (payoff dominant for $i$ against $S$) if

$$\sum_{j \in S \setminus \{i\}} u_{ij}(A,A) \geq \sum_{j \in S \setminus \{i\}} u_{ij}(B,B).$$

Combining the above arguments, we obtain the following proposition.

**Proposition 4.** *If $S$ follows a coalitional payoff-difference based rule, $A$ is $RD_{i(V \setminus S)}$ and $PD_{iS}$ for all $i \in S$, and*

*(i) $\Upsilon_S$ is self-regarding, or*

*(ii) $A$ is $ARD_{Sj}$ for all $j \notin S$,*

*then $c_S(\cdot,\cdot)$ is asymmetric towards $A$.*

Finally, note that the coalitional rules we have considered in this section involve coalition $S$ comparing $(\sigma_S^A, \sigma_{V \setminus S})$ to $(\sigma_S^B, \sigma_{V \setminus S})$. Another possibility is that a coalition would compare an alternative profile to the status quo $\sigma$. This leads to difficulties similar to violations of supermodularity discussed in Section 5. This is not pursued further here, although a detailed study of the intricacies of such rules would certainly be an interesting topic for further study.

### F.3 Proofs for coalitional choice

**Lemma 10.** *Let $\sigma, \hat{\sigma}$ be such that $V_B(\sigma) = V_A(\hat{\sigma})$. If $S$ follows a coalitional payoff-difference based rule, $A$ is $RD_{i(V\setminus S)}$ and $PD_{iS}$ for all $i \in S$, and*

*(i) $\Upsilon_S$ is self-regarding, or*

*(ii) $A$ is $ARD_{Sj}$ for all $j \notin S$,*

*then $c_S(\sigma, (\sigma_S^B, \sigma_{V\setminus S})) \geq c_S(\hat{\sigma}, (\sigma_S^A, \hat{\sigma}_{V\setminus S}))$.*

*Proof of Lemma 10.* If $\sigma_S = \sigma_S^B$, then $\hat{\sigma}_S = \hat{\sigma}_S^A$, so $c_S(\sigma, (\sigma_S^B, \sigma_{V\setminus S})) = c_S(\hat{\sigma}, (\sigma_S^A, \hat{\sigma}_{V\setminus S})) = 0$.
If $\sigma_S \neq \sigma_S^B$, then $\hat{\sigma}_S \neq \hat{\sigma}_S^A$. Consider the elements of $E_S^{\sigma}$,

$$(E_S^{\sigma})_j = U_j(\sigma_S^A, \sigma_{V\setminus S}) - U_j(\sigma_S^B, \sigma_{V\setminus S}) \tag{F.4}$$
$$= \begin{cases} \sum_{i\in S} \left(u_{ji}(A,A) - u_{ji}(A,B)\right) & \text{if } j \notin S,\ \sigma(j) = A, \\ -\sum_{i\in S} \left(u_{ji}(B,B) - u_{ji}(B,A)\right) & \text{if } j \notin S,\ \sigma(j) = B, \\ \sum_{\substack{k\notin S \\ \sigma(k)=A}} \big(u_{jk}(A,A) - u_{jk}(B,A)\big) & \\ \quad -\sum_{\substack{k\notin S \\ \sigma(k)=B}} \big(u_{jk}(B,B) - u_{jk}(A,B)\big) & \\ \quad +\sum_{\substack{i\in S \\ i\neq j}} \left(u_{ji}(A,A) - u_{ji}(B,B)\right) & \text{if } j \in S, \end{cases}$$

and the elements of $-E_S^{\hat{\sigma}}$,

$$(-E_S^{\hat{\sigma}})_j = -U_j(\sigma_S^A, \hat{\sigma}_{V\setminus S}) + U_j(\sigma_S^B, \hat{\sigma}_{V\setminus S}) \tag{F.5}$$
$$= \begin{cases} -\sum_{i\in S} \left(u_{ji}(A,A) - u_{ji}(A,B)\right) & \text{if } j \notin S,\ \hat{\sigma}(j) = A, \\ \sum_{i\in S} \left(u_{ji}(B,B) - u_{ji}(B,A)\right) & \text{if } j \notin S,\ \hat{\sigma}(j) = B, \\ -\sum_{\substack{k\notin S \\ \hat{\sigma}(k)=A}} \big(u_{jk}(A,A) - u_{jk}(B,A)\big) & \\ \quad +\sum_{\substack{k\notin S \\ \hat{\sigma}(k)=B}} \big(u_{jk}(B,B) - u_{jk}(A,B)\big) & \\ \quad -\sum_{\substack{i\in S \\ i\neq j}} \left(u_{ji}(A,A) - u_{ji}(B,B)\right) & \text{if } j \in S. \end{cases}$$

Noting that $\sigma(j) = A$ if and only if $\hat{\sigma}(j) = B$, $\sigma(j) = B$ if and only if $\hat{\sigma}(j) = A$, we can subtract (F.5) from (F.4) to get

$$(E_S^{\sigma} - (-E_S^{\hat{\sigma}}))_j = (E_S^{\sigma} + E_S^{\hat{\sigma}})_j = \tag{F.6}$$

$$= \begin{cases} \sum_{i\in S}\left((u_{ji}(A,A)-u_{ji}(A,B))-(u_{ji}(B,B)-u_{ji}(B,A))\right) & \text{if } j\notin S,\ \sigma(j)=A,\\ \sum_{i\in S}\left((u_{ji}(A,A)-u_{ji}(A,B))-(u_{ji}(B,B)-u_{ji}(B,A))\right) & \text{if } j\notin S,\ \sigma(j)=B,\\ \sum_{\substack{k\notin S\\ \sigma(k)=A}}\Big(\big(u_{jk}(A,A)-u_{jk}(B,A)\big)-\big(u_{jk}(B,B)-u_{jk}(A,B)\big)\Big) & \\ \quad+\sum_{\substack{k\notin S\\ \sigma(k)=B}}\big(u_{jk}(A,A)-u_{jk}(B,A)\big)-\big(u_{jk}(B,B)-u_{jk}(A,B)\big) & \\ \quad+\sum_{\substack{i\in S\\ i\neq j}}\left((u_{ji}(A,A)-u_{ji}(B,B))+(u_{ji}(A,A)-u_{ji}(B,B))\right) & \text{if } j\in S, \end{cases}$$

If $A$ is $\mathrm{RD}_{j(V\smallsetminus S)}$ for $j\in S$, then the sum of the first and second lines of the third case of (F.6) is nonnegative. If $A$ is $\mathrm{PD}_{jS}$ for $j\in S$, then the third line of the third case of (F.6) is nonnegative. Therefore, if $A$ is $\mathrm{RD}_{j(V\smallsetminus S)}$ and $\mathrm{PD}_{jS}$, the third case of (F.6) is nonnegative. That is, if $j\in S$, then $(E_S^{\sigma}+E_S^{\hat\sigma})_j\geq 0$, so $(E_S^{\sigma})_j\geq(-E_S^{\hat\sigma})_j$.

If $\Upsilon_S^C$ is self-regarding, then $(E_S^{\sigma})_j\geq(-E_S^{\hat\sigma})_j$ for all $j\in S$ implies that $\Upsilon_S^C(E_S^{\sigma})\geq\Upsilon_S^C(-E_S^{\hat\sigma})$ and therefore, by (F.1), $c_S(\sigma,(\sigma_S^B,\sigma_{V\smallsetminus S}))\geq c_S(\hat\sigma,(\sigma_S^A,\hat\sigma_{V\smallsetminus S}))$.

If A is $\mathrm{ARD}_{Sj}$ for all $j\notin S$, then the first and second cases of (F.6) are nonnegative. That is, if $j\notin S$, $(E_S^{\sigma}+E_S^{\hat\sigma})_j\geq 0$, so $(E_S^{\sigma})_j\geq(-E_S^{\hat\sigma})_j$. Therefore $E_S^{\sigma}\geq E_S^{\hat\sigma}$, and as $\Upsilon_S^C$ is non-decreasing, $\Upsilon_S^C(E_S^{\sigma})\geq\Upsilon_S^C(-E_S^{\hat\sigma})$ and therefore, by (F.1), $c_S(\sigma,(\sigma_S^B,\sigma_{V\smallsetminus S}))\geq c_S(\hat\sigma,(\sigma_S^A,\hat\sigma_{V\smallsetminus S}))$. ☐

**Lemma 11.** *Let $\hat\sigma,\tilde\sigma$ be such that $V_A(\hat\sigma)\subseteq V_A(\tilde\sigma)$. If $S$ follows a coalitional payoff-difference based rule, then $c_S(\hat\sigma,(\sigma_S^A,\hat\sigma_{V\smallsetminus S}))\geq c_S(\tilde\sigma,(\sigma_S^A,\tilde\sigma_{V\smallsetminus S}))$.*

*Proof of Lemma 11.* If $\tilde\sigma_S=\sigma_S^A$, then $c_S(\hat\sigma,(\sigma_S^A,\hat\sigma_{V\smallsetminus S}))\geq c_S(\tilde\sigma,(\sigma_S^A,\tilde\sigma_{V\smallsetminus S}))=0$.
If $\tilde\sigma_S\neq\sigma_S^A$, then $\hat\sigma_S\neq\sigma_S^A$. Using (F.5) for both $E_S^{\hat\sigma}$ and $E_S^{\tilde\sigma}$ gives

$$\text{(F.7)}\qquad ((-E_S^{\hat\sigma})-(-E_S^{\tilde\sigma}))_j$$

$$= \begin{cases} 0 & \text{if } j\notin S,\ \hat\sigma(j)=A,\\ 0 & \text{if } j\notin S,\ \tilde\sigma(j)=B,\\ \sum_{i\in S}\left((u_{ji}(A,A)-u_{ji}(A,B))+(u_{ji}(B,B)-u_{ji}(B,A))\right) & \text{if } j\notin S,\ \tilde\sigma(j)=A,\ \hat\sigma(j)=B,\\ \sum_{\substack{k\notin S\\ \tilde\sigma(k)=A}}\Big(u_{ik}(A,A)-u_{ik}(B,A)\Big) & \\ \quad-\sum_{\substack{k\notin S\\ \hat\sigma(k)=A}}\Big(u_{ik}(A,A)-u_{ik}(B,A)\Big) & \\ \quad+\sum_{\substack{k\notin S\\ \hat\sigma(k)=B}}\Big(u_{ik}(B,B)-u_{ik}(A,B)\Big) & \\ \quad-\sum_{\substack{k\notin S\\ \tilde\sigma(k)=B}}\Big(u_{ik}(B,B)-u_{ik}(A,B)\Big) & \text{if } j\in S. \end{cases}$$

The third case of of (F.7) is nonnegative by (2.2). The first two lines of the fourth case, taken together, are nonnegative as $V_A(\hat\sigma)\subseteq V_A(\tilde\sigma)$. The final two lines of the fourth case, taken together, are nonnegative as $V_B(\tilde\sigma)\subseteq V_B(\hat\sigma)$. So every element of $((-E_S^{\hat\sigma})-(-E_S^{\tilde\sigma}))_j$ is nonnegative and $-E_S^{\hat\sigma}\geq -E_S^{\tilde\sigma}$. As $\Upsilon_S^C$ is non-decreasing, $\Upsilon_S^C(-E_S^{\hat\sigma})\geq\Upsilon_S^C(-E_S^{\tilde\sigma})$ and therefore, by (F.1), $c_S(\hat\sigma,(\sigma_S^A,\hat\sigma_{V\smallsetminus S}))\geq c_S(\tilde\sigma,(\sigma_S^A,\tilde\sigma_{V\smallsetminus S}))$. ☐

**Lemma 12.** *If $S$ follows a coalitional payoff-difference based rule, then $c_S$ is asymmetric towards $A$ if and only if, for all $\sigma$, $\tilde{\sigma}$ such that $V_B(\sigma) \subseteq V_A(\tilde{\sigma})$, we have that $c_S(\sigma, (\sigma_S^B, \sigma_{V\setminus S})) \geq c_S(\tilde{\sigma}, (\tilde{\sigma}_S^A, \tilde{\sigma}_{V\setminus S}))$.*

*Proof of Lemma 12.* Consider $\sigma$, $\sigma'$, $\tilde{\sigma}$ such that $V_B(\sigma) \subseteq V_A(\tilde{\sigma})$.

If $\sigma'$ is not equal to $\sigma$, $(\sigma_S^B, \sigma_{V\setminus S})$ or $(\sigma_S^A, \sigma_{V\setminus S})$, then by (F.1), $c_S(\sigma, \sigma') = \infty$, so setting $\bar{\sigma} = \sigma^A$, we have that $V_B(\sigma') \subseteq V_A(\bar{\sigma})$, $V_A(\tilde{\sigma}) \subseteq V_A(\bar{\sigma})$, and $c_S(\sigma, \sigma') \geq c_S(\tilde{\sigma}, \bar{\sigma})$, satisfying the condition in Definition 3.

If $\sigma = \sigma'$, then, by (F.1), $c_S(\sigma, \sigma') = 0$. Letting $\bar{\sigma} = \tilde{\sigma}$, we have $V_B(\sigma') = V_B(\sigma) \subseteq V_A(\tilde{\sigma}) = V_A(\bar{\sigma})$ and, by (F.1), $c_S(\tilde{\sigma}, \bar{\sigma}) = 0$, so $c_S(\sigma, \sigma') \geq c_S(\tilde{\sigma}, \bar{\sigma}) = 0$, satisfying the condition in Definition 3.

If $\sigma \neq \sigma' = (\sigma_S^A, \sigma_{V\setminus S})$, let $\bar{\sigma} = \tilde{\sigma}$. Then we have $V_B(\sigma') \subset V_B(\sigma) \subseteq V_A(\tilde{\sigma}) = V_A(\bar{\sigma})$ and, by (F.1), $c_S(\tilde{\sigma}, \bar{\sigma}) = 0$, so $c_S(\sigma, \sigma') \geq c_S(\tilde{\sigma}, \bar{\sigma}) = 0$, satisfying the condition in Definition 3.

The only remaining case is $\sigma \neq \sigma' = (\sigma_S^B, \sigma_{V\setminus S})$. For $\bar{\sigma}$ to satisfy $V_B(\sigma') \subseteq V_A(\bar{\sigma})$, it must be that $\bar{\sigma}_S = \sigma_S^A$, and for $c_S(\tilde{\sigma}, \bar{\sigma}) < \infty$, it must be that $\bar{\sigma}_{V\setminus S} = \tilde{\sigma}_{V\setminus S}$. Hence, the condition in Definition 3 will be satisfied if and only if $c_S(\sigma, \sigma') \geq c_S(\tilde{\sigma}, \bar{\sigma}) = c_S(\tilde{\sigma}, (\sigma_S^A, \tilde{\sigma}_{V\setminus S}))$, the condition in the statement of the lemma.

□

*Proof of Proposition 4.* Let $\sigma$, $\tilde{\sigma}$ be such that $V_B(\sigma) \subseteq V_A(\tilde{\sigma})$. Define $\hat{\sigma}$ so that $V_B(\sigma) = V_A(\hat{\sigma})$. Note that $V_A(\hat{\sigma}) \subseteq V_A(\tilde{\sigma})$. Then,

$$c_S(\sigma, (\sigma_S^B, \sigma_{V\setminus S})) \underbrace{\geq}_{\text{by Lemma 10}} c_S(\hat{\sigma}, (\sigma_S^A, \hat{\sigma}_{V\setminus S})) \underbrace{\geq}_{\text{by Lemma 11}} c_S(\tilde{\sigma}, (\sigma_S^A, \tilde{\sigma}_{V\setminus S})),$$

satisfying the condition for asymmetry given in Lemma 12. □

## G. Payoff transformations

This section formalizes the arguments of Sections 6.2 and 6.3 of the main manuscript. Sometimes a transformation of payoffs can carry conceptual weight. In such cases, it can be instructive to consider the implications of the transformation with respect to conditions on the underlying payoffs. For example, we can subject the payoffs of player $i$ to a *Homo Moralis* transformation (Bergstrom, 1995; Alger and Weibull, 2013, 2016),

$$\text{(G.1)} \qquad u_{ij}^{HM}\Big(\sigma(i), \sigma(j)\Big) = (1-\sigma)\, u_{ij}\Big(\sigma(i), \sigma(j)\Big) + \sigma\, u_{ij}\Big(\sigma(i), \sigma(i)\Big),$$

where $\sigma \in [0,1]$ is a parameter that weighs the payoff maximizing first term against the *Kantian* second term.

Consider a player $i$ who follows a self-regarding payoff-difference based rule according to the transformed payoffs. For this rule to be asymmetric, we require risk dominance of $A$ under the transformed payoffs. Using payoffs $u_{ij}^{HM}$ in the definition of $\text{RD}_i$ and substituting, we obtain

$$(G.2)\qquad (1-\sigma)\underbrace{\sum_{j\in V\setminus\{i\}} u_{ij}(A,A)+u_{ij}(A,B)-u_{ij}(B,A)-u_{ij}(B,B)}_{\geq 0 \text{ if and only if } A \text{ is } \mathrm{RD}_i}$$
$$+2\sigma\underbrace{\sum_{j\in V\setminus\{i\}} u_{ij}(A,A)-u_{ij}(B,B)}_{\geq 0 \text{ if and only if } A \text{ is } \mathrm{PD}_{iV}}\geq 0.$$

If $\sigma = 0$, then (G.2) is the risk dominance condition of Proposition 1[i]. If $\sigma = 1$, then (G.2) is the component of the payoff dominance condition of Proposition 4[i] that relates to the incentives of player $i$ under coalitional choice by the entire player set $V$. If both terms under the summations are greater than zero, then the condition holds regardless of the value of $\sigma$ and so asymmetry will continue to hold even when $\sigma$ changes (see Newton, 2017; Nax and Rigos, 2016; Wu, 2017).

It is similarly possible to subject the payoffs of player $i$ to an altruistic transformation,

$$u^A_{ij}(\sigma(i),\sigma(j)) = (1-\alpha)\,u_{ij}(\sigma(i),\sigma(j)) + \alpha\, u_{ji}(\sigma(j),\sigma(i)),$$

where $\alpha \in [0,1]$ is a parameter that weights the payoff maximizing first term against the altruistic second term. This approach to altruistic behavior is less flexible than the approach taken in Section 4. However, it is common, so it is worth noting that it can easily fit into our framework.

Again consider a player $i$ who follows a self-regarding payoff-difference based rule according to the transformed payoffs. For this rule to be asymmetric, we require risk dominance of $A$ under the transformed payoffs. Using payoffs $u^A_{ij}$ in the definition of $\mathrm{RD}_i$ and substituting, we obtain a convex combination of risk dominance and altruistic risk dominance,

$$(G.3)\qquad (1-\alpha)\underbrace{\sum_{j\in V\setminus\{i\}} u_{ij}(A,A)+u_{ij}(A,B)-u_{ij}(B,A)-u_{ij}(B,B)}_{\geq 0 \text{ if and only if } A \text{ is } \mathrm{RD}_i}$$
$$+\alpha\sum_{j\in V\setminus\{i\}}\underbrace{u_{ji}(A,A)+u_{ji}(B,A)-u_{ji}(A,B)-u_{ji}(B,B)}_{\geq 0 \text{ if and only if } A \text{ is } \mathrm{ARD}_{ij}}\geq 0.$$

# H. Adding perturbations to an unperturbed process

This section relates to Footnote 10 of the main manuscript, which notes that models in the literature that additively combine an unperturbed process with perturbations can be analyzed using a variant of the convex combinations considered in Theorem 1. Consider a family of behavioral rules $P_S$ defined by a convex combination of $\bar{P}_S$ and $\bar{\bar{P}}_S$,

$$(H.1)\qquad P^{\varepsilon}_S = (1-\varepsilon)\bar{P}^{\varepsilon}_S + \varepsilon\bar{\bar{P}}^{\varepsilon}_S.$$

That is, the share of $\bar{\bar{P}}_S$ in $P_S$ vanishes as $\varepsilon \to 0$. Note that for $\varepsilon = 0$, we have $P^0_S = \bar{P}^0_S$. Following the same steps as the Proof of Theorem 1, we obtain

(H.2) $$c_S = \min\{\bar{c}_S, \bar{\bar{c}}_S + 1\},$$

where the additional 1 arises from the multiplicative $\varepsilon$ in the final term of (H.1). Note that even though $\bar{\bar{c}}_S + 1$ is not a cost function of a process, the definition of asymmetry can still be applied and the proof of Lemma 2 continues to hold. It follows from Definition 3 that if $\bar{\bar{c}}_S$ is asymmetric, then $\bar{\bar{c}}_S + 1$ is asymmetric. Therefore, if $\bar{c}_S$ and $\bar{\bar{c}}_S$ are asymmetric, then Lemma 2 implies that $c_S$ is asymmetric. That is, we have shown a result similar to Theorem 1.

**Proposition 5.** *If $\bar{P}_S$ and $\bar{\bar{P}}_S$ are asymmetric towards A, then $P_S$ defined by $P_S^\varepsilon = (1-\varepsilon)\bar{P}_S^\varepsilon + \varepsilon\bar{\bar{P}}_S^\varepsilon$, is asymmetric towards A.*

Now, consider the special case $\bar{P}_S^\varepsilon \equiv \bar{P}_S^0$ for all $\varepsilon \in [0,1)$. That is, $\bar{P}_S$ does not depend on $\varepsilon$. When this is the case, we can consider $P_S$ as being formed by taking $\bar{P}_S$ as a starting point and adding a small probability $\varepsilon$ of perturbations occuring according to $\bar{\bar{P}}_S$. In such cases, for $P_S$ to be asymmetric, it must be that $\bar{P}_S$ is also asymmetric.

**Proposition 6.** *Let $P_S$ be defined by $P_S^\varepsilon = (1-\varepsilon)\bar{P}_S^\varepsilon + \varepsilon\bar{\bar{P}}_S^\varepsilon$, and let $\bar{P}_S^\varepsilon \equiv \bar{P}_S^0$ for all $\varepsilon \in [0,1)$. If $P_S$ is asymmetric towards A, then $\bar{P}_S$ is asymmetric towards A.*

As an example, consider $S = \{i\}$. Let $\bar{P}_{\{i\}}$ be a rule according to which $i$ sometimes remains playing his current strategy and sometimes plays a best response, randomizing if there are multiple best responses. Let $\bar{\bar{P}}_{\{i\}}$ be a rule that, independently of $\varepsilon$, selects each strategy with strictly positive probability. Combining these according to (H.1), $P_{\{i\}}$ is then a best response with uniform deviations rule as defined in Section 4.2. Proposition 6 tells us that for this rule to be asymmetric, it must be the case that the unperturbed best response rule is also asymmetric.

*Proof of Proposition 6.* If $\bar{P}_S^\varepsilon \equiv \bar{P}_S^0$ for all $\varepsilon \in [0,1)$, then transition probabilities under $\bar{P}_S$ are either zero for all $\varepsilon \in [0,1)$ or positive for all $\varepsilon \in [0,1)$. Considering (2.5), this implies that

(H.3) The range of $\bar{c}_S$ is $\{0,\infty\}$.

We prove the proposition by contrapositive. Assume that $\bar{c}_S$ is not asymmetric. There must exist $\sigma$, $\sigma'$, $\tilde{\sigma}$ as per Definition 3, such that for any $\tilde{\sigma}'$ that satisfies the inclusion relations described in Definition 3, we have $\bar{c}_S(\sigma,\sigma') \not\geq \bar{c}_S(\tilde{\sigma},\tilde{\sigma}')$. Together with (H.3), this implies

(H.4) $$\bar{c}_S(\sigma,\sigma') = 0 \quad \text{and} \quad \bar{c}_S(\tilde{\sigma},\tilde{\sigma}') = \infty.$$

Combining (H.2) and (H.4), we obtain

(H.5) $$c_S(\sigma,\sigma') = \min\{0, \bar{\bar{c}}_S(\sigma,\sigma') + 1\} = 0 < 1 \leq \min\{\infty, \bar{\bar{c}}_S(\tilde{\sigma},\tilde{\sigma}') + 1\} = c_S(\tilde{\sigma},\tilde{\sigma}').$$

Therefore, $c_S$ is not asymmetric. This concludes the proof. □

As a final observation that may be useful to practitioners, we note that a rule that, independently of $\varepsilon$, selects each strategy with strictly positive probability is trivially asymmetric towards any given strategy. Therefore, if $\bar{P}_{\{i\}}$ satisfies the conditions of Proposition 6 and deviations are uniform in the sense that the process follows (H.1) and $\bar{\bar{P}}_{\{i\}}$ is independent of $\varepsilon$ and selects each strategy with strictly positive probability, then asymmetry of $P_{\{i\}}$ depends only on $\bar{P}_{\{i\}}$.

**Proposition 7.** *Let $P_{\{i\}}$ be defined by $P^{\varepsilon}_{\{i\}} = (1-\varepsilon)\bar{P}^{\varepsilon}_{\{i\}} + \varepsilon\bar{\bar{P}}^{\varepsilon}_{\{i\}}$. Let $\bar{P}^{\varepsilon}_{\{i\}} \equiv \bar{P}^{0}_{\{i\}}$ and $\bar{\bar{P}}^{\varepsilon}_{\{i\}} \equiv \bar{\bar{P}}^{0}_{\{i\}}$ for all $\varepsilon \in [0,1)$. Let $\bar{\bar{P}}_{\{i\}}(\sigma,\sigma) > 0$ and $\bar{\bar{P}}_{\{i\}}(\sigma,\sigma^{(i)}) > 0$ for all $\sigma$. Then $P_{\{i\}}$ is asymmetric towards A if and only if $\bar{P}_{\{i\}}$ is asymmetric towards A.*

*Proof.* $\bar{\bar{P}}_{\{i\}}$ is trivially asymmetric, so if $\bar{P}_{\{i\}}$ is asymmetric, by Proposition 5 we have that $P_{\{i\}}$ is asymmetric. In the other direction, if $P_{\{i\}}$ is asymmetric, then $\bar{P}_{\{i\}}$ is asymmetric by Proposition 6 □

# I. Asymmetry with more than two strategies

This section shows how the results of Section 3 extend to environments with more than two strategies. The case of more than two strategies is very similar to the two strategy case. The fundamental difference is that instead of thinking in terms of strategy $A$ versus strategy $B$, we think in terms of strategy $A$ versus all strategies other than $A$.

## I.1 Amended model and definition

Amending the model of Section 2, let $\sigma \in \Sigma := \{A, B_1, B_2, \ldots\}^V$, where $\{A, B_1, B_2, \ldots\}$ is a finite set of strategies. Denote by $V_{\neg A}(\sigma) \subseteq V$ the set of players who play any strategy other than $A$ at profile $\sigma$.

Asymmetry towards $A$ is then defined by replacing $V_B(\cdot)$ in Definition 3 by $V_{\neg A}(\cdot)$. That is, asymmetry towards $A$ is defined with respect to all other strategies rather than just $B$.

*Definition 13.* $c_S(\cdot,\cdot)$ is *asymmetric* towards $A$ if, for any $\sigma, \sigma', \tilde{\sigma} \in \Sigma$, such that $V_{\neg A}(\sigma) \subseteq V_A(\tilde{\sigma})$, there exists $\tilde{\sigma}' \in \Sigma$ such that $V_A(\tilde{\sigma}) \subseteq V_A(\tilde{\sigma}')$, $V_{\neg A}(\sigma') \subseteq V_A(\tilde{\sigma}')$ and $c_S(\sigma,\sigma') \geq c_S(\tilde{\sigma},\tilde{\sigma}')$.

Under this amended model and definition of asymmetry, Lemmas 1, 2 and Theorems 1, 2, 3 still hold. Substituting $V_{\neg A}(\cdot)$ for $V_B(\cdot)$, the proofs of Section A continue to apply.

## I.2 Individual behavioral rules

Consider individual behavioral rules, that is $S = \{i\}$. The condition for asymmetry of $c_{\{i\}}$ in Definition 13 requires three cases to be checked, the first two of which are trivial.

**Case 1.** $\sigma_j \neq \sigma'_j$ for some $j \neq i$.

Let $\tilde{\sigma}' = \sigma^A$, so that $V_A(\tilde{\sigma}) \subseteq V_A(\tilde{\sigma}')$ and $V_{\neg A}(\sigma') \subseteq V_A(\tilde{\sigma}')$ are satisfied. Further, note that $c_i(\sigma,\sigma') = \infty$, so $c_{\{i\}}(\sigma,\sigma') \geq c_{\{i\}}(\tilde{\sigma},\tilde{\sigma}')$.

| | | Player $j$ | | | |
|---|---|---|---|---|---|
| | | $A$ | $B_1$ | $B_2$ | $\cdots$ |
| | $A$ | $a$ | 0 | 0 | $\cdots$ |
| Player $i$ | $B_1$ | 0 | $b_1$ | 0 | $\cdots$ |
| | $B_2$ | 0 | 0 | $b_2$ | $\cdots$ |
| | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\ddots$ |

Figure 1: **Payoffs.** Assume $a \geq b_1 \geq b_2 \geq \dots$ without loss of generality. For each combination of strategies, entries give payoffs for player $i$.

**Case 2.** $\sigma_j = \sigma'_j$ for all $j \neq i$; $\sigma_i \neq A$ or $\sigma'_i = A$ or $\tilde{\sigma}_i = A$.

Let $\tilde{\sigma}' = \tilde{\sigma}$, so that $V_A(\tilde{\sigma}) \subseteq V_A(\tilde{\sigma}')$ and $V_{\neg A}(\sigma') \subseteq V_A(\tilde{\sigma}')$ are satisfied. Further, note that $c_{\{i\}}(\tilde{\sigma}, \tilde{\sigma}') = 0$, so that $c_{\{i\}}(\sigma, \sigma') \geq c_{\{i\}}(\tilde{\sigma}, \tilde{\sigma}')$.

**Case 3.** $\sigma_j = \sigma'_j$ for all $j \neq i$; $\sigma_i = A$ and $\sigma'_i \neq A$ and $\tilde{\sigma}_i \neq A$.

This case must be checked for the behavioral rule under consideration.

In summary, to check asymmetry towards $A$ for individual behavioral rules, we only need to consider cases in which player $i$ switches from $A$ at $\sigma$ to another strategy at $\sigma'$, and from a strategy other than $A$ at $\tilde{\sigma}$ to $A$ at $\tilde{\sigma}'$.

### I.3 Example - coordination with more than two strategies

For all $j \in V$, $j \neq i$, let $u_{ij}$ be given by the table in Figure 1. Let $i$ follow a rule such that $c_{\{i\}}(\sigma, \sigma) = 0$ and $c_{\{i\}}(\sigma, (\sigma'_i, \sigma_{-i})) = U_i(\sigma)$ for $\sigma'_i \neq \sigma_i$. The reader will observe that this is a condition dependent rule as considered in Section 5. To show asymmetry of $c_{\{i\}}$ towards $A$, we check **Case 3** from Section I.2.

Consider $\sigma$, $\sigma'$, $\tilde{\sigma}$ such that $\sigma_i = A$, $\sigma'_i \neq A$, $\tilde{\sigma}_i \neq A$ and $V_{\neg A}(\sigma) \subseteq V_A(\tilde{\sigma})$. From $\sigma$, a switch by player $i$ from $A$ to some other strategy has a cost of $U_i(\sigma) = a(|V_A(\sigma)| - 1)$. That is, $c_{\{i\}}(\sigma, \sigma') = a(|V_A(\sigma)| - 1)$. Let $\tilde{\sigma}'$ differ from $\tilde{\sigma}$ only in that $\tilde{\sigma}'_i = A$. Note that $V_A(\tilde{\sigma}) \subseteq V_A(\tilde{\sigma}')$ and $V_{\neg A}(\sigma') \subseteq V_A(\tilde{\sigma}')$ as required by Definition 13. From $\tilde{\sigma}$, the cost of player $i$ switching to $A$ depends on his strategy at $\tilde{\sigma}$. Specifically, if $\tilde{\sigma}_i = B_k$, then $c_{\{i\}}(\tilde{\sigma}, \tilde{\sigma}') = U_i(\tilde{\sigma}) = b_k(|V_{B_k}(\tilde{\sigma})| - 1)$ . We have

$$
\text{(I.1)} \qquad c_{\{i\}}(\tilde{\sigma}, \tilde{\sigma}') = b_k(|V_{B_k}(\tilde{\sigma})| - 1) \underbrace{\leq}_{\substack{\text{by } b_k \leq b_1 \\ \text{and } V_{B_k}(\cdot) \subseteq V_{\neg A}(\cdot)}} b_1(|V_{\neg A}(\tilde{\sigma})| - 1)
$$

$$
\underbrace{\leq}_{\substack{\text{as } V_{\neg A}(\sigma) \subseteq V_A(\tilde{\sigma}) \\ \Rightarrow V_{\neg A}(\tilde{\sigma}) \subseteq V_A(\sigma)}} b_1(|V_A(\sigma)| - 1) \underbrace{\leq}_{\text{by } b_1 \leq a} a(|V_A(\sigma)| - 1) = c_{\{i\}}(\sigma, \sigma').
$$

Therefore, the conditions of Definition 13 are satisfied and $c_i$ is asymmetric towards $A$.

## References


Alger, I. and Weibull, J. W. (2013). Homo moralis–preference evolution under incomplete information and assortative matching. *Econometrica*, 81(6):2269–2302.

Alger, I. and Weibull, J. W. (2016). Evolution and Kantian morality. *Games and Economic Behavior*, 98:56–67.

Bergstrom, T. C. (1995). On the evolution of altruistic ethical rules for siblings. *American Economic Review*, pages 58–81.

Dokumaci, E. and Sandholm, W. H. (2011). Large deviations and multinomial probit choice. *Journal of Economic Theory*, 146(5):2151–2158.

Nax, H. H. and Rigos, A. (2016). Assortativity evolving from social dilemmas. *Journal of Theoretical Biology*, 395:194–203.

Newton, J. (2012). Coalitional stochastic stability. *Games and Economic Behavior*, 75(2):842–54.

Newton, J. (2017). The preferences of homo moralis are unstable under evolving assortativity. *International Journal of Game Theory*, 46(2):583–589.

Norman, T. W. (2009). Rapid evolution under inertia. *Games and Economic Behavior*, 66(2):865–879.

Peski, M. (2010). Generalized risk-dominance and asymmetric dynamics. *Journal of Economic Theory*, 145(1):216 – 248.

Sawa, R. (2014). Coalitional stochastic stability in games, networks and markets. *Games and Economic Behavior*, 88:90–111.

Wu, J. (2017). Political institutions and the evolution of character traits. *Games and Economic Behavior*, 106:260–276.