Online Appendix to "A Theory of Strategic Uncertainty and Cultural Diversity"

Willemien Kets Alvaro Sandroni

June 18, 2020

I Strategic uncertainty

In the main text, we argue that culturally diverse societies face more strategic uncertainty than culturally homogeneous societies. This appendix makes that claim precise. Strategic uncertainty is high if players are not very informative, i.e., if posterior beliefs are close to the prior. In the current setting, this is equivalent to the variance in impulses being high.

To show this, we calculate the variance in impulses as a function of diversity. Fix Q_{in}, Q_{out} and recall that Q_{in}, Q_{out} are functions of culture strength q and cultural distance d. Suppose the level of diversity is β so that the minority and majority shares are β and $\tilde{\beta} = 1 - \alpha$, respectively. For a player $j \in N$ who belongs to the minority group, the expected proportion of players who have the same impulse as he does is $Q^{\min}(\beta; q, d) := \tilde{\beta} Q_{out} + \beta Q_{in}$. Likewise, for a player j who belongs to the majority group, the expected proportion of players who have the same impulse is $Q^{\mathrm{maj}}(\beta; q, d) := \tilde{\beta} Q_{in} + \beta Q_{out}$. Since $\tilde{\beta} \geq \frac{1}{2}$ and $Q_{in} > Q_{out} > \frac{1}{2}$, $Q^{\mathrm{maj}}(\beta; q, d) \geq Q^{\min}(\beta; q, d) > \frac{1}{2}$ (with strict inequality if $\beta < \frac{1}{2}$). Then, the degree of strategic uncertainty that a player in the minority and the majority face is given by

$$\begin{aligned} \mathcal{V}^{\min}(\beta; q, d) &:= Q^{\min}(\beta; q, d)\,(1 - Q^{\min}(\beta; q, d)); \\ \mathcal{V}^{\mathrm{maj}}(\beta; q, d) &:= Q^{\mathrm{maj}}(\beta; q, d)\,(1 - Q^{\mathrm{maj}}(\beta; q, d)), \end{aligned}$$

respectively. So, the majority faces less strategic uncertainty than the minority (i.e., $\mathcal{V}^{\mathrm{maj}}(\beta; q, d) < \mathcal{V}^{\min}(\beta; q, d)$). We can also define aggregate strategic uncertainty $\mathcal{V}$ for the society by

$$\mathcal{V}(\beta; q, d) := \tilde{\beta} \mathcal{V}^{\mathrm{maj}}(\beta; q, d) + \beta \mathcal{V}^{\min}(\beta; q, d).$$

As $\mathcal{V}(\beta; q, d)$ is increasing in diversity β, players face less strategic uncertainty in culturally homogeneous societies than in culturally diverse ones. Also, there is less uncertainty in societies with a strong culture ($\mathcal{V}(\beta; q, d)$ decreases with q) and that the difference between homogeneous and diverse societies is larger when the culture is strong and when the cultural distance between groups is large. That is, if $\beta' > \beta$, then $\mathcal{V}(\beta; q, d) - \mathcal{V}(\beta'; q, d)$ increases with q and with d.

II Comparison with correlated equilibrium

This appendix compares the set of all introspective equilibria for a given game (across all societies) to the set of correlated equilibria. We focus on linear games with identical preferences. Relative to correlated equilibrium, introspective equilibrium has considerable cutting power. A first observation is that *for any linear game, the set of introspective equilibria (across all societies) is always a strict subset of the class of correlated equilibria.*[1] To make this claim precise, we can identity each society (characterized by β, q, d) with the impulse distribution it generates (Section 2.3). Write Δ for the class of impulse distributions that are associated with some society. To be able to compare the set of introspective equilibria (profiles of mappings from impulses to actions) to correlated equilibria (distributions over action profiles), we consider the distributions over action profiles induced by introspective equilibrium. That is, for $\rho \in [0,1]$ and $\mu \in \Delta$, let $\Sigma_\mu(\rho)$ be the set of distributions over action profiles induced by some introspective equilibrium for the society described by μ and risk parameter ρ. By Proposition A.2, $\Sigma_\mu(\rho)$ has at least one element; and for generic values of ρ, it has precisely one element. For $\rho \in [0,1]$, let $\Sigma(\rho) = \bigcup_{\mu\in\Delta} \Sigma_\mu(\rho)$. With some abuse of terminology, we refer to $\Sigma(\rho)$ as the set of introspective equilibria (across all $\mu \in \Delta$) for risk parameter ρ. Let $\mathcal{C}(\rho)$ be the set of correlated equilibria for risk parameter ρ.[2] Then, the following claim, which is a corollary of Lemma A.5, shows that introspective equilibrium can always rule out certain behaviors that are consistent with correlated equilibrium:

Corollary II.1. [The Cutting Power of Introspective Equilibrium (I)] *For any $\rho \in [0,1]$, the set $\Sigma(\rho)$ of all introspective equilibria (for some society) is a strict subset of the set $\mathcal{C}(\rho)$ of all correlated equilibria.*

Proof. Let $\rho \in [0,1]$. Then, there is a correlated equilibrium in which all players choose H as well as a correlated equilibrium in which all players choose L (this follows because both are pure Nash equilibria). If $\rho < \frac{1}{2}$, then, by Lemma A.5, for every $\mu \in \Delta$, there is no introspective equilibrium in $\Sigma_\mu(\rho)$ such that all players choose L. Likewise, if $\rho > \frac{1}{2}$, then, by Lemma A.5, for every $\mu \in \Delta$, there is no introspective equilibrium in $\Sigma_\mu(\rho)$ such that all players choose H. Finally, if $\rho = \frac{1}{2}$, then, by Lemma A.5, for every $\mu \in \Delta$, there is no introspective equilibrium in $\Sigma_\mu(\rho)$ such that all players choose L, and no introspective equilibrium in $\Sigma_\mu(\rho)$ such that all players choose H. □

A second observation is that in some limiting cases, the set of introspective equilibria (across all societies) collapses to a singleton, as the following corollary of Lemma A.5 demonstrates.[3]

[1] We thus restrict attention to impulse distributions that are associated with some society (i.e., impulse distributions characterized by β, q, d). Without any restrictions on the class of impulse distributions, any correlated equilibrium is an introspective equilibrium for some impulse distribution. This follows from the revelation principle: fix a correlated equilibrium and take the impulse distribution to be the distribution over action profiles generated by the correlated equilibrium. Then the game has a unique introspective equilibrium in which all players follow their impulse, and this introspective equilibrium coincides with the original correlated equilibrium. See Myerson (1994) for a version of the revelation principle for complete-information game and a discussion in the context of correlated equilibrium.

[2] Note that specifying the risk parameter ρ is sufficient to pin down the incentive constraints: any two linear games with the same risk parameter have the same set of correlated equilibria.

[3] The limit of a collection of sets is the set-theoretic limit.

Corollary II.2. [The Cutting Power of Introspective Equilibrium (II)] *As ρ goes to 0, 1, or $\frac{1}{2}$, the set of introspective equilibria (across all societies) converges to a singleton:*

(a) As $\rho \to 0$, the set of introspective equilibria (across all societies) converges to the unique strategy profile where all players choose the high action regardless of their impulse;

(b) As $\rho \to 1$, the set of introspective equilibria (across all societies) converges to the unique strategy profile where all players choose the low action regardless of their impulse;

(c) As $\rho \to \frac{1}{2}$, the set of introspective equilibria (across all societies) converges to the unique strategy profile in which all players follow their impulse.

Again, the proof follows directly from Lemma A.5. So, in the limit that the risk parameter goes to 0, 1, or $\frac{1}{2}$, the set of introspective equilibria (across all societies) collapses to a singleton, and the limiting introspective equilibrium is independent of sociocultural factors. By contrast, the set of correlated equilibria does not converge to a singleton when the risk parameter goes to 0, 1, or $\frac{1}{2}$. Instead, it is a continuum (except in trivial cases). To see this, note that for any $\rho \in [0, 1]$, the set of correlated equilibria contains at least the strict Nash equilibria as well as the nonstrict pure Nash equilibrium in which a proportion ρ of players chooses H; the claim now follows by noting that, except in knife-edge cases, at least two of these Nash equilibria have different payoff profiles, and the set of correlated equilibrium payoff profiles includes the convex hull of Nash equilibrium payoff profiles.

III Experimental evidence

This appendix discusses the testable implications of the results in Section 3 in more detail and relates them to experimental evidence. We focus on linear games with identical preferences ($\rho_j = \rho$ for all j) because these games have been the focus of much of the experimental literature, though our predictions extend more generally.

A first prediction is that *the proportion of players who choose the high action increases as the risk parameter falls.* This follows from a straightforward argument based on the proof of Lemma A.5 (Kets et al., 2019). In particular, introspective equilibrium selects one of the pure Nash equilibria if and only if one of the actions stands out in terms of payoffs (i.e., ρ sufficiently close to 0 or 1). On the other hand, if there is limited asymmetry between the actions in terms of payoffs, both actions are chosen with positive probability and behavior is not consistent with Nash equilibrium. There is considerable experimental evidence for this. For two-player coordination games, Mehta et al. (1994), Straub (1995) and Schmidt et al. (2003), among many others, show that for intermediate values of the risk parameter, behavior is not consistent with Nash equilibrium: players coordinate at a higher rate than in mixed Nash equilibrium, but at a lower rate than in pure Nash equilibrium.[4] Another direct implication is that there can be inefficient lock-in: Players may coordinate on a Pareto-dominated

[4] We are not aware of any experimental studies that study games with extreme values for the risk parameter. This could be a selection effect: if the interest is in testing competing hypotheses, there is no reason to select games for which there is an obvious way to play so that all theories make the same prediction; see Schmidt et al. (2003, p. 285) for comments along these lines.

equilibrium. This prediction has received ample experimental support for a range of coordination games (see, e.g., Van Huyck et al., 1990; Cooper et al., 1990, 1992; Straub, 1995).

A second testable implication is that for intermediate values of the risk parameter (i.e., ρ close to $\frac{1}{2}$), behavior is strongly influenced by situational factors (i.e., the cultural salience of actions), and that behavioral consistency improves when strategic uncertainty decreases. Experimental support for the influence of contextual factors comes from a variety of sources. First, there is extensive evidence that past experience influences strategic behavior even when there are no incentives to build reputation or signal intentions (e.g., Schmidt et al., 2003). To the extent that history shapes impulses, this is consistent with our results. A second type of evidence for this prediction involves the (cultural) saliency of alternatives. Evidence suggests that when the payoff structure of the game provides little guidance (i.e., ρ close to $\frac{1}{2}$) and one action is (culturally) salient, then players have a pre-reflective inclination to select the salient alternative (e.g., Mehta et al., 1994, p. 659). This means that the coordination rate increases when one of the actions is significantly more salient than others, in line with our model. A third source of evidence that contextual cues influence behavior comes from individual variation in perspective-taking ability. An individual with superior perspective-taking abilities presumably has a highly informative signal about other players' impulses and will thus be better at coordinating. Curry and Jones Chesters (2012) show that in the pure coordination games of Mehta et al. (1994), subjects with superior perspective-taking ability (as measured by a self-report questionnaire) have a higher probability of coordinating when matched against the population, consistent with our theory.

A third prediction is that it is easier for people to anticipate the actions of members of their own group and that people who belong to the same group are more likely to have the same impulse. This is in line with the experimental evidence of Jackson and Xing (2014), who contrast the behavior of subjects residing in India versus the U.S. in a battle-of-the-sexes game. They find that subjects are better able to predict the actions of members of their own group. Moreover, the two groups differ in the actions that they take. To the extent that actions are a function of impulses, these findings support our assumption that players from the same group are more likely to have the same impulse and that players find it easier to anticipate the impulses of members of their own group. Consistent with our predictions, Jackson and Xing find that subjects are more successful at coordinating when they are matched with a member of their own group.

Existing equilibrium selection methods cannot account for these findings. For example, in the context of coordination games, payoff dominance selects the same Nash equilibrium independent of the risk parameter, as do team reasoning theories (Sugden, 1993). Risk dominance makes the stark prediction that players coordinate on the efficient action (with probability 1) whenever the risk parameter is less than $\frac{1}{2}$, while they coordinate on the inefficient action whenever the risk parameter is greater than $\frac{1}{2}$. So, risk dominance cannot explain why there can be miscoordination when there is limited asymmetry among the actions, as in the work of Mehta et al. (1994) and others.[5] Since the risk-dominant Nash equilibrium is selected by global games methods (Carlsson and van Damme, 1993), evolutionary models (Young, 1993; Kandori et al., 1993), and quantal response equilibrium (McKelvey and Palfrey, 1995), these methods cannot explain the observed behavior either.[6] This also

[5] Mixed Nash equilibrium can also not account for the observed behavior: The coordination rate in Mehta et al.'s (1994) and related experiments lies strictly between that in pure and mixed Nash equilibrium.

[6] The noisy introspection model of Goeree and Holt (2004) predicts non-Nash behavior in at least some coordination games. However, it is unclear how predictions vary with payoffs and thus whether the model can

holds for other concepts. Most notably, Crawford and Haller (1990), in their study of how players use asymmetries in the game to coordinate, derive the stark prediction that players coordinate (with probability 1) whenever there is *some* asymmetry between actions, no matter how small. By predicting that coordination succeeds only if there is sufficient asymmetry between the actions, our model provides a more nuanced and arguably more realistic view than existing concepts. And while some existing methods, such as risk dominance, the global games selection, and certain learning models, can account for inefficient lock-in, a novel prediction not captured by existing models is that the gap between individual incentives and socially optimal behavior is smaller when there is more strategic uncertainty in the sense that societies that experience more strategic uncertainty can avoid inefficient lock-in for a larger range of payoff parameters.

IV Details for applications

This appendix shows that our applications of linear games satisfies the conditions for Proposition 3.4 (assuming identical preferences). That is, we show that for each application, the social welfare function $W(m;\rho)$ is quadratic in m with its minimum $\underline{m}$ increasing in ρ, or, equivalently, that $W(1;\rho)-W(0;\rho)$ decreases with m.

We start with the example in Section 3.1 and related models (Sections 4.1 and 4.3). These are linear games, with the risk parameter for each player $j \in N$ equal to $\rho_j = \frac{1}{2} + \frac{\lambda}{2(1-\lambda)}(1-2\tau_j)$ (where $\lambda = \frac{1}{2}$ in Section 3.1). If $\rho_j = \rho$ for all j, then the social welfare function is quadratic with minimum $\underline{m} = \rho$.

The infinitely repeated game in Section 4.2 is a linear game with identical preferences. To show that it satisfies the relevant conditions, we show that this is true for general (2×2) coordination games:

	H	L
H	u_{HH}	u_{HL}
L	u_{LH}	u_{LL}

with $u_{HH} > u_{LH}$, $u_{LL} > u_{HL}$, and $u_{HH} \geq u_{LL}$. The infinitely repeated game is then a special case (with, e.g., $u_{\text{cc}} = u_{HH}$). Coordination games satisfy the conditions in Proposition 3.4 if $2u_{LL} > u_{LH} + u_{HL}$ (which holds if cooperation is efficient, i.e., $2u_{\text{cc}} > u_{\text{cd}} + u_{\text{dc}}$) and we are considering, e.g., increasing u_{LL} to u_{HH} keeping the other payoff parameters fixed.

V Omitted proofs

V.1 Proof of Lemma A.1

At level 0, each player follows his impulse. So, the level-0 strategies are anonymous and we can write $\sigma^0 : S \times \mathcal{U} \times \mathcal{G} \to S$ for the level-0 strategy profile. Since a player's level-0 action depends only on

reproduce the observed comparative statics.

his impulse, the level-0 strategy is jointly measurable. Hence, players' expected payoff) is well-defined.

For $k > 0$, suppose that the level-$(k-1)$ strategies are anonymous, so that we can denote the profile by $\sigma^{k-1} : S \times \mathcal{U} \times \mathcal{G} \to S$, and suppose that σ^{k-1} is jointly measurable. Then, players' expected payoff is well-defined. To show that the level-k strategies are anonymous and measurable, notice that the mapping from triples $(I, G, \boldsymbol{u}) \in S \times \mathcal{G} \times \mathcal{U}$ to the associated (interim) expected payoff $U(s, (m_{s'}(\sigma))_{s' \in S}; I, G, \boldsymbol{u})$ is jointly measurable. It then follows from the Measurable Maximum Theorem (Aliprantis and Border, 2006, Thm. 18.19) that the best-response correspondence ψ^k that maps each triple $(I, \boldsymbol{u}, G)$ into its set of best responses is nonempty and jointly measurable and admits a measurable selector. Because the action set is finite, this implies that the level-k strategy is jointly measurable and anonymous. □

V.2 Proof of Proposition A.2 (cntd)

We prove existence of introspective equilibrium for linear games with heterogeneous preferences where the distribution $F(\rho_j)$ has mean $\mu \geq \frac{1}{2}$. As for the case $\mu \leq \frac{1}{2}$, we prove the result under slightly weaker assumptions than in the main text: rather than assuming that $f(\rho_j)$ is unimodal and symmetric, we require that the density $f(\rho_j)$ satisfies

$$f(\tfrac{1}{2} + x) \geq f(\tfrac{1}{2} + y) \qquad \forall x, y \text{ s.t. } y \geq x \geq 0; \tag{V.1}$$

$$f(\tfrac{1}{2} - x) \geq f(\tfrac{1}{2} + x) \qquad \forall x \geq 0. \tag{V.2}$$

Again, these conditions are satisfied when $f(\rho_j)$ is unimodal and symmetric (with mean $\mu \geq \frac{1}{2}$) but they are strictly weaker. As before, we can rewrite the expressions for ρ^k_{HA} and ρ^k_{HB} as

$$\begin{aligned}
\rho^k_{HA} &= \tilde{\beta}(Q_{in} - \tilde{Q}_{in})F(\rho^{k-1}_{HA}) + \tilde{\beta}\tilde{Q}_{in}\big[F(\rho^{k-1}_{HA}) + F(\rho^{k-1}_{LA})\big] + \\
&\quad \beta(Q_{out} - \tilde{Q}_{out})F(\rho^{k-1}_{HB}) + \beta\tilde{Q}_{out}\big[F(\rho^{k-1}_{HB}) + F(\rho^{k-1}_{LB})\big]; \\
\rho^k_{HB} &= \tilde{\beta}(Q_{out} - \tilde{Q}_{out})F(\rho^{k-1}_{HA}) + \tilde{\beta}\tilde{Q}_{out}\big[F(\rho^{k-1}_{HA}) + F(\rho^{k-1}_{LA})\big] + \\
&\quad \beta(Q_{in} - \tilde{Q}_{in})F(\rho^{k-1}_{LB}) + \beta\tilde{Q}_{in}\big[F(\rho^{k-1}_{HB}) + F(\rho^{k-1}_{LB})\big];
\end{aligned}$$

respectively. So, by a similar argument as before, it suffices to prove that $\{\rho^k_{HA}\}_k$, $\{\rho^k_{HB}\}_k$, and $\{\bar{\rho}^k\}_k$ converge. This follows from the following analogues of Lemmas A.3–A.4:

Lemma V.1. *Suppose $f(\rho_j)$ has mean $\mu \geq \frac{1}{2}$ and satisfies (V.1)–(V.2), and fix a group $G \in \{A, B\}$ and $k > 0$. If $\bar{\rho}^k \leq \bar{\rho}^{k-1} \leq \frac{1}{2}$ and $\rho^k_{LG} \leq \rho^{k-1}_{LG}$, then $F(\rho^k_{HG}) + F(\rho^k_{LG}) \leq F(\rho^{k-1}_{HG}) + F(\rho^{k-1}_{LG})$.*

Proof. For concreteness, take $G = A$. If $\rho^k_{LA} \geq \rho^{k-1}_{LA}$, then the result follows immediately from the fact that $F(\rho_j)$ is increasing. So suppose that $\rho^k_{LA} < \rho^{k-1}_{LA}$. Define $\Delta^k := \rho^k_{LA} - \rho^{k-1}_{LA}$. By a similar argument as before,

$$0 < \Delta^k \leq \rho^{k-1}_{HA} - \rho^k_{HA}$$

and

$$\begin{aligned}
&\big[F(\rho^k_{HA}) + F(\rho^k_{LA})\big] - \big[F(\rho^{k-1}_{HA}) + F(\rho^{k-1}_{LA})\big] \\
&\geq \int_0^{\Delta^k} \Big[f(\bar{\rho}^k + (\rho^k_{HA} - \bar{\rho}^k + u)) - f(\bar{\rho}^{k-1} - (\rho^k_{HA} - \bar{\rho}^k + u)\Big] du.
\end{aligned}$$

The result then follows by noting that, under (V.1)–(V.2) (and $\mu \geq \frac{1}{2}$), for any $\bar{\rho} \leq \frac{1}{2}$ and $x \geq 0$, $f(\bar{\rho} + x) \geq f(\bar{\rho} - x)$.

To prove the result for $k = 1$, it suffices to show that for $\bar{\rho} \leq \frac{1}{2}$ and $x \geq 0$, $F(\bar{\rho}+x)+F(\bar{\rho}-x) \leq 1$. But this follows from a similar argument as before, using that $f(\bar{\rho}-y) \leq f(\bar{\rho}+y)$ for $y \geq 0$ and $\bar{\rho} \leq \frac{1}{2}$. □

Lemma V.2. *Suppose $f(\rho_j)$ has mean $\mu \geq \frac{1}{2}$ and satisfies (V.1)–(V.2). Then, for all $k > 0$, $\bar{\rho}^k \leq \bar{\rho}^{k-1} \leq \frac{1}{2}$, $\rho^k_{HA} \leq \rho^{k-1}_{HA}$ and $\rho^k_{HB} \leq \rho^{k-1}_{HB}$.*

The proof is analogous to that of Lemma A.4 and thus omitted. It now follows immediately that the sequences $\{\rho^k_{HA}\}_k$, $\{\rho^k_{HB}\}_k$, and $\{\bar{\rho}^k\}_k$ converge: As before, by Lemma V.2, each sequence is bounded and monotone. Hence, there exist $\bar{\rho} \in (0, \frac{1}{2})$, $\rho_{HA} \in [\bar{\rho}, 1)$, and $\rho_{HB} \in [\bar{\rho}, 1)$ such that $\bar{\rho}^k \downarrow \bar{\rho}$, $\rho^k_{HA} \downarrow \rho_{HA}$, and $\rho^k_{HB} \downarrow \rho_{HB}$. □

V.3 Proof of Lemma A.6

Throughout, we will use that for all I, G, and n, $\rho^{n,1}_{IG} = \rho^1_{IG}$, where $\rho^{n,k}_{IG}$ are the level-k cutoffs under $F^n(\rho_j)$ and ρ^1_{IG} is the conditional expectation for type (I, G) defined in the proof of Lemma A.5. We write $\rho^{(n)}_{IG}$ for the limits $\lim_{k\to\infty} \rho^{n.k}_{IG}$ that describe the introspective equilibria. (These limits exist by Proposition A.2.)

We start by considering generic values for ρ, i.e., $\rho \notin \{\rho^*_{HA}, \rho^*_{HB}, \rho^*_{LB}, \rho^*_{LA}\}$, where ρ^*_{IG} is the cutoff for players from group G with impulse I in introspective equilibrium when players have identical preferences (Lemma A.5). We need to consider five cases, with each case corresponding to one of the five cases in Lemma A.5 (Figure 4). Because the cutoffs in Lemma A.5 are symmetric (in $\rho = \frac{1}{2}$), we can group them into three cases:

Case 1: $\rho < \rho^*_{LA}$ or $\rho > \rho^*_{HA}$ First consider the case $\rho < \rho^*_{LA}$, i.e., $\rho < \min\{\rho^1_{LB}, \rho^2_{LA}\}$ (where ρ^k_{IG} is the level-k cutoff for players from group G with impulse I in a game with identical preferences; cf.Lemma A.5). Then, under σ_ρ, all players choose H in introspective equilibrium (Lemma A.5(a)). We discuss the case where $\rho^1_{LB} < \rho^2_{LA}$ (i.e., $\beta > \beta^*$); the proof for other cases is analogous and hence omitted. Let $\rho < \rho^1_{LB}$. Because $\rho^2_{LA} > \rho^1_{LB}$, by continuity, for $\zeta \in (0, 1)$ sufficiently small, $(1-\zeta)\rho^2_{LA} > \rho^1_{LB}$. Fix some $\zeta \in (0, 1)$ for which this holds. Because $\rho < \rho^1_{LB}$, there is N_ζ such that for $n > N_\zeta$, $F^n(\rho^1_{LB}) > 1 - \zeta$. Fix $n > N_\zeta$. Then, by Lemma A.4, for all $k \geq 2$, $F(\rho^{n,k}_{LB}) > 1 - \zeta$. Using that $\rho^{n,k}_{HA} \geq \rho^{n,k}_{HB} \geq \rho^{n,k}_{LB}$, we also have $F(\rho^{n,k}_{HA}), F(\rho^{n.k}_{HB}) > 1-\zeta$. Then, using Lemma A.4 again, for $k \geq 2$,

$$\rho^{n,k}_{LA} > (1 - \tilde{\beta}Q_{in})(1 - \zeta) = (1 - \zeta)\rho^2_{LA} > \rho^1_{LB}.$$

Hence, for all I, G, $F^n(\rho^{(n)}_{IG}) > 1-\zeta$. Since we can always choose $\zeta < \varepsilon$, we thus have $F^n(\rho^{(n)}_{IG}) > 1-\varepsilon$ for all I, G whenever $n > N_\zeta$. It is now immediate that for the game with distribution $F^n(\rho_j)$ for $n > N_\zeta$, the proportion of players choosing H is greater than $1-\varepsilon$ in introspective equilibrium in every state. For example, in state $(\theta_A, \theta_B) = (L, L)$, the proportion of players choosing H in introspective equilibrium under $F^n(\rho_j)$ is

$$\tilde{\beta} \cdot [\tilde{q}F^n(\rho^{(n)}_{HA}) + qF^n(\rho^{(n)}_{LA})] + \beta \cdot [\tilde{q}F^{(n)}(\rho^{(n)}_{HB}) + qF^n(\rho^{(n)}_{LB})] >$$
$$\tilde{\beta} \cdot [\tilde{q}(1-\varepsilon) + q(1-\varepsilon)] + \beta \cdot [\tilde{q}(1-\varepsilon) + q(1-\varepsilon)] = 1-\varepsilon.$$

The proof for the case $\rho > \rho^*_{HA}$ is analogous and thus omitted.

Case 2: $\rho \in (\rho^*_{LB}, \rho^*_{HB})$. Under σ_ρ, all players choose the action they expect to be culturally salient (Lemma A.5(c)). We start with the case $\rho \leq \frac{1}{2}$. We claim that there exist $\tilde{\rho}_{LB} < \rho, \tilde{\rho}_{HB} > \rho$ and $\tilde{N}$ such that for $n > \tilde{N}$,

$$\rho^{(n)}_{LB} \leq \tilde{\rho}_{LB}; \qquad \rho^{(n)}_{HB} \geq \tilde{\rho}_{HB}. \tag{V.3}$$

This proves the claim: Because $\rho \in (\tilde{\rho}_{LB}, \tilde{\rho}_{HB})$, for every $\varepsilon > 0$, there is $\tilde{N}_\varepsilon$ such that for $n > \tilde{N}_\varepsilon$, $F^n(\tilde{\rho}_{HB}) - F^n(\tilde{\rho}_{LB}) > 1 - \varepsilon$. Then, if we take $N_\varepsilon := \max\{\tilde{N}, \tilde{N}_\varepsilon\}$, by a similar argument as before, for $n > N_\varepsilon$, $F^n(\rho^{(n)}_{HB}) - F^n(\rho^{(n)}_{LB}) > 1 - \varepsilon$. But then the proportion of players playing according to σ_ρ under $\sigma^{(n)}$ is greater than $1 - \varepsilon$: For example, in state $(\theta_A, \theta_B) = (L, L)$, the proportion of players choosing the action they expect to be culturally salient under $F^n(\rho_j)$ is

$$\tilde{\beta} \cdot [\tilde{q} F^n(\rho^{(n)}_{HA}) + q(1 - F^n(\rho^{(n)}_{LA}))] + \beta \cdot [\tilde{q} F^n(\rho^{(n)}_{HB}) + q(1 - F^n(\rho^{(n)}_{LB}))] >$$
$$\tilde{\beta} \cdot [\tilde{q}(1 - \varepsilon) + q(1 - \varepsilon)] + \beta \cdot [\tilde{q}(1 - \varepsilon) + q(1 - \varepsilon)] = 1 - \varepsilon.$$

It remains to prove (V.3). We use that $\rho^*_{LB} = \rho^1_{LB}$ and $\rho^*_{HB} = \rho^1_{HB}$. For $\zeta \in (0, 1)$, define

$$\tilde{\rho}_{LB} := (1 - \zeta)\, \rho^1_{LB} + \zeta; \qquad \tilde{\rho}_{HB} := (1 - \zeta)\, \rho^1_{HB}.$$

Note that $\tilde{\rho}_{LB} > \rho^1_{LB}$ and $\tilde{\rho}_{HB} < \rho^1_{HB}$. Fix $\zeta \in (0, 1)$ such that $\rho \in (\tilde{\rho}_{LB}, \tilde{\rho}_{HB})$. Then, there is N_ζ such that for $n > N_\zeta$, $F^n(\tilde{\rho}_{HB}) - F^n(\tilde{\rho}_{LB}) > 1 - \zeta$. Fix $n > N_\zeta$. We show that for every $k > 0$,

$$\rho^{n,k}_{LB} < \tilde{\rho}_{LB}; \qquad \rho^{n,k}_{HB} > \tilde{\rho}_{HB},$$

which proves (V.3). We prove the claim by induction. For $k = 1$, the claim follows directly because $\tilde{\rho}_{LB} > \rho^1_{LB} = \rho^{n,1}_{LB}$ and $\tilde{\rho}_{HB} < \rho^1_{HB} = \rho^{n,1}_{HB}$. Hence, $F^n(\rho^{n,1}_{HB}) - F^n(\rho^{n,1}_{LB}) > 1 - \zeta$. For $k > 1$, suppose that $\rho^{n,k-1}_{LB} < \tilde{\rho}_{LB}$ and $\rho^{n,k-1}_{HB} > \tilde{\rho}_{HB}$. Then, $F^n(\rho^{n,k-1}_{HB}) - F^n(\rho^{n,k-1}_{LB}) > 1 - \zeta$. Using that

$$\tilde{\rho}_{LB} = (1 - \zeta)\, (\tilde{\beta} \tilde{Q}_{out} + \beta \tilde{Q}_{in}) + \zeta$$

and rewriting the relevant expressions, we obtain

$$\tilde{\rho}_{LB} - \rho^{n,k}_{LB} = \tilde{\beta} \tilde{Q}_{out}(1 - F^n(\rho^{n,k-1}_{HA})) + \beta \tilde{Q}_{in}(1 - F^n(\rho^{n,k-1}_{HB})) +$$
$$\tilde{\beta} Q_{out}(\zeta - F^n(\rho^{n,k-1}_{LA}) + \beta Q_{in}(\zeta - F^n(\rho^{n,k-1}_{LB}) > 0;$$

and

$$\rho^{n,k}_{HB} - \tilde{\rho}_{LB} = \tilde{\beta} Q_{out}(F^n(\rho^{n,k-1}_{HA}) - F^n(\rho^{n,k-1}_{LA}) - (1 - \zeta)) +$$
$$\beta Q_{in}(F^n(\rho^{n,k-1}_{HB}) - F^n(\rho^{n,k-1}_{LB}) - (1 - \zeta)) + \tilde{\beta} F^n(\rho^{n,k-1}_{LA} + \beta F^n(\rho^{n,k-1}_{LB}) > 0.$$

The rest of the proof now follows because we can always take $\zeta < \varepsilon$ (and set $\tilde{N} = N_\zeta$). The proof for the case $\rho \geq \frac{1}{2}$ is analogous and thus omitted.

Case 3: $\rho \in (\rho_L A^*, \rho^*_{LB})$ or $\rho \in (\rho^*_{HB}, \rho^*_{HA})$ Under σ_ρ, majority players choose the action they expect to be culturally salient while minority players choose a fixed action (H if $\rho \in (\rho_L A^*, \rho^*_{LB})$ and L if $\rho \in (\rho^*_{HB}, \rho^*_{HA})$; Lemma A.5(b), (d)). We start with the case $\rho \in (\rho_L A^*, \rho^*_{LB})$. We use that $\rho^*_{LA} = \rho^1_{LA}$ and $\rho^*_{LB} = \rho^1_{LB}$. We claim that there exist $\tilde{\rho}_{LA} < \rho, \tilde{\rho}_{LB} > \rho$ and $\tilde{N}$ such that for $n > \tilde{N}$,

$$\rho^{(n)}_{LA} \leq \tilde{\rho}_{LA}; \qquad \rho^{(n)}_{LB} \geq \tilde{\rho}_{LB}.$$

Again, this proves the claim. The existence of $\tilde{\rho}_{LB}$ with the desired properties is immediate: Because $\rho < \rho^{n,1}_{LB} = \rho^1_{LB}$ for all n and because $\rho^{n,k}_{LB}$ increases with k (by Lemma A.4), we can set $\tilde{\rho}_{LB} := \rho^1_{LB}$. The existence of $\tilde{\rho}_{LA}$ follows by a similar argument as in Case 2: There is $\zeta \in (0,1)$ sufficiently small such that $(1-\zeta)\cdot\rho^2_{LA} + \zeta < \rho$. Fix $\zeta \in (0,1)$ for which this holds and define

$$\tilde{\rho}_{LA} := (1-\zeta)\cdot\rho^2_{LA} + \zeta.$$

Because $\rho \in (\tilde{\rho}_{LA}, \tilde{\rho}_{LB})$, there is N_ζ such that for $n > N_\zeta$, $F^n(\tilde{\rho}_{LB}) - F^n(\tilde{\rho}_{LA}) > 1 - \zeta$. Fix $n > N_\zeta$. We show that for every $k > 0$,

$$\rho^{n,k}_{LA} \leq \tilde{\rho}_{LA}.$$

As before, we prove the claim by induction. For $k = 1$, we have $\rho^{n,1}_{LA} = \rho^1_{LA} < \rho^2_{LA} < \tilde{\rho}_{LA}$. For $k > 1$, suppose $\rho^{n,k-1}_{LA} < \tilde{\rho}_{LA}$ so that $F^n(\rho^{n,k-1}_{LA}) < \zeta$. Rewriting the relevant expressions yields

$$\tilde{\rho}_{LA} - \rho^{n,k}_{LA} = \tilde{\beta}\tilde{Q}_{in}(1 - F^n(\rho^{n,k-1}_{HA})) + \beta\tilde{Q}_{out}(1 - F^n(\rho^{n,k-1}_{HB})) + \\ \beta Q_{out}(1 - F^n(\rho^{n,k-1}_{LB}) + \tilde{\beta}Q_{in}(\zeta - F^n(\rho^{n,k-1}_{LA}) > 0.$$

Again, the result follows because we can always take $\zeta < \varepsilon$. The proof for the case $\rho \in (\rho^1_{HB}, \rho^2_{HA})$ is analogous and therefore omitted.

It remains to consider the nongeneric cases $\rho \in \{\rho^*_{HA}, \rho^*_{HB}, \rho^*_{LB}, \rho^*_{LA}\}$. We can again group the different cases together thanks to the symmetry of the cutoffs around $\frac{1}{2}$.

Case 4: $\rho = \rho^*_{LB}$ or $\rho = \rho^*_{HB}$ for $\beta < \beta^*$ We prove the result for $\rho = \rho^*_{LB}$; the proof for the case $\rho = \rho^*_{HB}$ is again similar and thus omitted. Suppose $\rho = \rho^*_{LB}$ and $\beta < \beta^*$. It is easy to verify that $\rho^{n,2}_{LB} > \rho^*_{LB}$ for all n. It then follows from Lemma A.4, $\rho^{(n)}_{LB} > \rho^*_{LB}$ for all n. As before, for $\zeta > 0$ sufficiently small,

$$\tilde{\rho}_{LA} := (1-\zeta)\rho^2_{LA} + \zeta$$

satisfies $\rho > \tilde{\rho}_{LA} > \rho^2_{LA}$. Fix $\zeta > 0$ such that this holds. Then, by a similar argument as before, there is N^1_ζ such that for all $n > N^1_\zeta$, we have $\rho^{(n)}_{LA} \leq \tilde{\rho}_{LA}$. Likewise, for $\eta > 0$ sufficiently small,

$$\tilde{\rho}_{HB} := (1-\eta)\rho^1_{HB}$$

satisfies $\rho < \tilde{\rho}_{HB}$. Fix $\eta > 0$ for which this holds. Again, by a similar argument as before, there is N^2_η such that for all $n > N^2_\eta$, we have $\rho^{(n)}_{HB} \geq \tilde{\rho}_{HB}$. It now follows that for every $\xi > 0$, there is N^3_ξ such that for $n > N^3_\xi$, we have $1 - F^n(\rho^{(n)}_{HB}) < \xi$ and $F^n(\rho^{(n)}_{LA}) < \xi$. For $\xi > 0$, let $n > N^3_\xi$. Then, the conditional expectation of the share m of players who choose H in introspective equilibrium for a player with impulse L from group B is greater than $\rho^1_{LB}(1-\xi) + \frac{1}{2}\tilde{\beta}Q_{in}$ and thus

$$\rho^{(n)}_{LB} > \rho^1_{LB}(1-\xi) + \tfrac{1}{2}\tilde{\beta}Q_{in},$$

Fix $\xi > 0$ such that $\rho^1_{LB}(1-\xi) + \frac{1}{2}\tilde{\beta}Q_{in} > \rho^1_{LB} = \rho$ (which holds for $\xi > 0$ sufficiently small). Define

$$\tilde{\rho}_{LB} := \rho^1_{LB}(1-\xi) + \tfrac{1}{2}\tilde{\beta}Q_{in}.$$

Then, the claim follows because for any $\varepsilon > 0$, we can choose $\zeta, \eta, \xi \in (0, \varepsilon)$; then, for n sufficiently large, the share of minority players who choose H (regardless of their impulse) and the share of majority players who choose the action they expect to be culturally salient is at least $1 - \varepsilon$.

Case 5: $\rho = \rho^*_{LB}$ or $\rho = \rho^*_{HB}$ for $\beta \geq \beta^*$ Notice that $\rho^*_{LA} = \rho^*_{LB}$ and $\rho^*_{HA} = \rho^*_{HB}$ for $\beta \geq \beta^*$ (Lemma A.5 and Figure 4) so this also covers the case $\rho = \rho^*_{LA}$ or $\rho = \rho^*_{HA}$ for $\beta \geq \beta^*$. We prove the result for $\rho = \rho^*_{LB}$; the proof for the case $\rho = \rho^*_{HB}$ is again similar and thus omitted. Suppose $\rho = \rho^*_{LB}$ and $\beta \geq \beta^*$. For $\eta > 0$ sufficiently small,

$$(1-\eta)\rho^1_{LA} + \tfrac{1}{2}\rho^1_{HA} > \rho^1_{LB} + \eta. \tag{V.4}$$

Fix $\eta > 0$ for which this holds, and define

$$\tilde{\rho}_{LA} := (1-\eta)\rho^1_{LA} + \tfrac{1}{2}\rho^1_{HA}.$$

We claim that $\rho^{(n)}_{LA} > \tilde{\rho}_{LA}$ for n sufficiently large. This proves the result: By taking $\eta < \varepsilon$, we find that, for n sufficiently large, the share of players who choose H regardless of their impulse is at least $1 - \varepsilon$. So, it remains to prove the claim. As before, there is N_η such that $F^n(\tilde{\rho}_{LA}) > 1 - \eta$. Fix $n > N_\eta$. Then, the conditional expectation at level 2 of the share of players choosing H for a player from group A with impulse L is

$$\tilde{\beta}\tilde{Q}_{in}F^n(\rho^1_{HA}) + \beta\tilde{Q}_{out}F^n(\rho^1_{HB}) + \beta Q_{out}F^n(\rho^1_{LB}) + \tilde{\beta}Q_{in}F^n(\rho^1_{LA}) > (1-\eta)\rho^1_{LA} + \tfrac{1}{2}\rho^1_{HA}.$$

Hence, by (V.4), $\rho^{n,2}_{LA} > \tilde{\rho}_{LA}$. By Lemma A.4, $\rho^{(n)}_{LA} > \tilde{\rho}_{LA}$.

Case 6: $\rho = \rho^*_{LA}$ or $\rho = \rho^*_{HA}$ for $\beta < \beta^*$ The proof for this case is a bit more involved, because unlike for the other cases, the equilibrium cutoff (i.e., ρ^*_{HA} or ρ^*_{LA}) for the game with identical preferences is not equal to the level-1 cutoff (i.e., ρ^1_{HA} or ρ^1_{LA}); hence, the fact that $\rho^{n,1}_{IG} = \rho^1_{IG}$ for all I, G and n (which we used extensively in the proof of other cases) is of little use here. To overcome this, this part of the proof utilizes that the distributions F^n have full support and satisfy condition (A.9). We prove the result for $\rho = \rho^*_{LA}$; the proof for the case $\rho = \rho^*_{HA}$ is again similar and thus omitted. Suppose $\rho = \rho^*_{LA}$ and $\beta < \beta^*$. First note that if we show that, for n sufficiently large, $\rho^{n,k}_{LA} > \rho^2_{LA}$ for some $k \geq 2$, then we are done. To see this, suppose that there exist N and $k \geq 2$ such that for $n > N$, $\rho^{n,k}_{LA} > \rho^2_{LA}$. Then, by symmetry, for $n > N$, $F^n(\rho^{n,k}_{LA}) > \frac{1}{2}$; and thus, by a similar argument as before, for every $\eta > 0$, there is N_η such that for $n > N_\eta$, $\rho^{n,k+1}_{LA} > (1-\eta)\rho^2_{LA} + \frac{1}{2}\tilde{\beta}Q_{in} = 1 - \eta - \left(\frac{1}{2} - \eta\right)\tilde{\beta}Q_{in}$. For $\eta > 0$ sufficiently small, the right-hand side of this inequality is greater than ρ^2_{LA}; fix $\eta > 0$ such that this holds. Then, by Lemma A.4, for every $\zeta > 0$, there is N_ζ such that for $n > N_\zeta$, $F^n(\rho^{(n)}_{LA}) > 1 - \zeta$. This proves the result: By taking $\zeta < \varepsilon$, we find that, for n sufficiently large, the share of players who choose H regardless of their impulse is at least $1 - \varepsilon$. It remains to show that, for n sufficiently large, $\rho^{n,k}_{LA} > \rho^2_{LA}$ for some $k \geq 2$. If $\rho^{n,2}_{LA} > \rho^2_{LA}$ for n sufficiently large, then this holds trivially. So suppose this is not the case. It is easy to verify that, for every $\eta > 0$, there is N_η such that for $n > N_\eta$, $\rho^{n,2}_{LA} > (1-\eta)\,\rho^2_{LA}$. Fix $\eta > 0$ such that $\rho - (1-\eta)\rho^2_{LA} < \rho^1_{LB} - \rho$ and fix $n > N_\eta$. Since the distribution $F^n(\rho_j)$ has full support, we can write

$$\rho^{n,3}_{LA} = \rho^2_{LA} + F^n(\rho^{n,2}_{LA}) \Bigg[\tilde{\beta} Q_{in} - \tilde{\beta}\tilde{Q}_{in} \left(\frac{1 - F^n(\rho^{n,2}_{HA})}{F^n(\rho^{n,2}_{LA})} \right) - \\ \beta \tilde{Q}_{out} \left(\frac{1 - F^n(\rho^{n,2}_{HB})}{F^n(\rho^{n,2}_{LA})} \right) - \beta Q_{out} \left(\frac{1 - F^n(\rho^{n,2}_{LB})}{F^n(\rho^{n,2}_{LA})} \right) \Bigg]. \tag{V.5}$$

Notice that $\rho^{n,2}_{LB} \geq \rho^1_{LB}$ (Lemma A.4) and that $\rho^{n,2}_{HA} \geq \rho^{n,2}_{HB} \geq \rho^{n,2}_{LB}$. By construction, $\rho - \rho^{n,2}_{LA} < \rho - (1-\eta)\rho^2_{LA} < \rho^1_{LB} - \rho$. But then, by (A.9), $\rho^{n,3}_{LA} > \rho^2_{LA}$ for n sufficiently large. □

V.4 Proof of Lemma A.9

We define the Bellman equations for level $k = 1$. Given a posterior belief $\mu \in [0,1]$ that the other player has an impulse to cooperate in the current period, the value function at level $k = 1$ is

$$V(\mu) = \max\Big\{ \mu\left[(1-\delta)\,u_{\text{dc}} + \delta\,V(0)\right] + (1-\mu)\left[(1-\delta)\,u_{\text{dd}} + \delta\,V(0)\right], \\ \mu\left[(1-\delta)\,u_{\text{cc}} + \delta\,V(1)\right] + (1-\mu)\left[(1-\delta)\,u_{\text{cd}} + \delta\,V(0)\right]\Big\}.$$

This yields

$$V(0) = u_{\text{dd}}; \qquad V^1(1) = \begin{cases} u_{\text{cc}} & \text{if } \delta \geq \delta_{GT}; \\ (1-\delta)\,u_{\text{dc}} + \delta\,u_{\text{dd}} & \text{otherwise;} \end{cases}$$

where

$$\delta_{GT} := \frac{u_{\text{dc}} - u_{\text{cc}}}{u_{\text{dc}} - u_{\text{dd}})}$$

is the standard grim-trigger threshold. For $\mu \in (0,1)$,

$$V(\mu) = \begin{cases} \mu\left[(1-\delta)\,u_{\text{cc}} + \delta\,u_{\text{cc}}\right] + (1-\mu)\left[(1-\delta)\,u_{\text{cd}} + \delta\,u_{\text{dd}}\right] & \text{if } \delta \geq \delta_\mu; \\ \mu\left[(1-\delta)\,u_{\text{dc}} + \delta\,u_{\text{dd}}\right] + (1-\mu)\left[(1-\delta)\,u_{\text{dd}} + \delta\,u_{\text{dd}}\right] & \text{otherwise;} \end{cases}$$

where

$$\delta_\mu := \frac{\mu(u_{\text{dc}} - u_{\text{cc}}) + (1-\mu)\,(u_{\text{dd}} - u_{\text{cd}})}{\mu\,(u_{\text{dc}} - u_{\text{dd}}) + (1-\mu)\,(u_{\text{dd}} - u_{\text{cd}})},$$

which is decreasing in μ and is strictly greater than the grim-trigger threshold δ_{GT} whenever $\mu < 1$.

Then, if we denote by μ^0_{IG} the posterior belief for a player with impulse I from group G that the other player has an impulse to cooperate at $t = 0$, then the discounted sum of expected payoffs to players with impulse I from group G at level 1 is $V(\mu^0_{IG})$. It is then easy to check that, at level 1, a player with impulse I from group G chooses the grim trigger strategy if $\delta \geq \delta_{\mu^0_{IG}}$ and defect in every period if $\delta \leq \delta_{\mu^0_{IG}}$. Likewise, if μ^0_{IG} is the posterior belief for a player with impulse I from group G that the other player has an impulse to choose H in the reduced game, then, at level 1, a player with impulse I from group G chooses the grim trigger strategy if $\delta \geq \delta_{\mu^0_{IG}}$ and defect in every period if $\delta \leq \delta_{\mu^0_{IG}}$. For $k > 1$, suppose, inductively that at level $k-1$, a player with impulse I from group G chooses the grim trigger strategy if $\delta \geq \delta_{\mu^{k-2}_{IG}}$ and defect in every period if $\delta \leq \delta_{\mu^{k-2}_{IG}}$, where μ^{k-2}_{IG} is the player's posterior that the other player chooses the grim trigger strategy at level $k-2$. Then, if we denote by μ_{IG}^{k-1} the posterior belief for a player with impulse I from group G that the other player chooses the grim trigger strategy at level $k-1$, the discounted sum of expected payoffs to players with impulse I from group G at level k is $V(\mu_{IG}^{k-1})$. Then, it is again easy to check that, at level k, a player with impulse I from group G chooses the grim trigger strategy if $\delta \geq \delta_{\mu_{IG}^{k-1}}$ and defect in every period if $\delta \leq \delta_{\mu_{IG}^{k}}$. Likewise, if μ_{IG}^{k-1} is the posterior belief for a player with impulse I from group G that the other player chooses H at level $k-1$ in the reduced game, then, at level k, a player with impulse I from group G chooses the grim trigger strategy if $\delta \geq \delta_{\mu_{IG}^{k-1}}$ and defect in every period if $\delta \leq \delta_{\mu_{IG}^{k-1}}$. By a similar argument as in the proof of Lemma A.5, the introspective process converges within finitely many steps. Hence, in any introspective equilibrium of the repeated game, players either choose the grim trigger strategy or they defect in every period; moreover, they choose the grim trigger strategy if and only if they choose H in the introspective equilibrium of the reduced game. □

References

Aliprantis, C. and K. Border (2006). *Infinite Dimensional Analysis: A Hitchhiker's Guide* (3rd ed.). Springer.

Carlsson, H. and E. van Damme (1993). Global games and equilibrium selection. *Econometrica 61*, 989–1018.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1990). Selection criteria in coordination games: Some experimental results. *American Economic Review 80*, 218–233.

Cooper, R., D. V. DeJong, R. Forsythe, and T. W. Ross (1992). Communication in coordination games. *Quarterly Journal of Economics 107*, 739–771.

Crawford, V. P. and H. Haller (1990). Learning how to cooperate: Optimal play in repeated coordination games. *Econometrica 58*, 571–595.

Curry, O. and M. Jones Chesters (2012). Putting ourselves in the other fellow's shoes: The role of 'theory of mind' in solving coordination problems. *Journal of Cognition and Culture 12*, 147–159.

Goeree, J. K. and C. A. Holt (2004). A model of noisy introspection. *Games and Economic Behavior 46*, 365–382.

Jackson, M. O. and Y. Xing (2014). Culture-dependent strategies in coordination games. *Proceedings of the National Academy of Sciences 111*, 10889–10896.

Kandori, M., G. J. Mailath, and R. Rob (1993). Learning, mutation, and long run equilibria in games. *Econometrica 61*, 29–56.

Kets, W., W. Kager, and A. Sandroni (2019). The value of a coordination game. Working paper.

McKelvey, R. D. and T. R. Palfrey (1995). Quantal response equilibria for normal form games. *Games and Economic Behavior 10*, 6–38.

Mehta, J., C. Starmer, and R. Sugden (1994). The nature of salience: An experimental investigation of pure coordination games. *American Economic Review 84*, 658–673.

Myerson, R. B. (1994). Communication, correlated equilibria and incentive compatibility. Volume 2 of *Handbook of Game Theory with Economic Applications*, Chapter 24, pp. 827–847. Elsevier.

Schmidt, D., R. Shupp, J. M. Walker, and E. Ostrom (2003). Playing safe in coordination games: The roles of risk dominance, payoff dominance, and history of play. *Games and Economic Behavior 42*, 281–299.

Straub, P. G. (1995). Risk dominance and coordination failures in static games. *Quarterly Review of Economics and Finance 35*, 339–363.

Sugden, R. (1993). Thinking as a team: Toward an explanation of nonselfish behavior. *Social Philosophy and Policy 10*, 69–89.

Van Huyck, J. B., R. C. Battalio, and R. O. Beil (1990). Tacit coordination games, strategic uncertainty, and coordination failure. *American Economic Review 80*, 234–248.

Young, H. P. (1993). The evolution of conventions. *Econometrica 61*, 57–84.