What Situation Is This? Shared Frames and Collective Performance

Strategy Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

What Situation Is This? Shared Frames and Collective Performance

Robert Gibbons, Marco LiCalzi, Massimo Warglien

informs
http://pubsonline.informs.org/journal/stsc

STRATEGY SCIENCE
Vol. 6, No. 2, June 2021, pp. 124–140
ISSN 2333-2050 (print), ISSN 2333-2077 (online)

What Situation Is This? Shared Frames and Collective Performance

Robert Gibbons,[a] Marco LiCalzi,[b] Massimo Warglien[b]

[a] Massachusetts Institute of Technology, Cambridge, Massachusetts 02142; [b] Università Ca' Foscari Venezia, 30121 Venice, Italy

Contact: rgibbons@mit.edu, https://orcid.org/0000-0001-6765-3039 (RG); licalzi@unive.it, https://orcid.org/0000-0003-4718-708X (ML); warglien@unive.it, https://orcid.org/0000-0002-0615-4938 (MW)

Received: September 30, 2019
Revised: July 8, 2020; October 11, 2020
Accepted: October 15, 2020
Published Online in Articles in Advance: February 11, 2021

https://doi.org/10.1287/stsc.2020.0120

Abstract. We study agents who distill the complex world around them using cognitive frames. We assume that agents share the same frame and analyze how the frame affects their collective performance. In one-shot and repeated interactions, the frame causes agents to be either better or worse off than if they could perceive the environment in full detail: it creates a fog of cooperation or a fog of conflict. In repeated interactions, the frame is as important as agents' patience in determining the outcome: for a fixed discount factor, when all agents choose what they perceive as their best play, there remain significant performance differences induced by different frames. A low-performing team conducting a site visit to observe a high-performing team will be mystified, sometimes observing different actions than they expected or being given unexpected reasons for the actions they expected. Finally, we distinguish between incremental versus radical changes in frames, and we develop a model of category formation to analyze challenges faced by a leader who seeks to improve the agents' collective performance.

History: This paper has been accepted for the *Strategy Science* Special Issue on Reinvigorating Research on Organizational Culture and Its Link to Strategy.

Funding: This project has received funding from the European Union's Horizon 2020 Research and Innovation programme [Grant Agreement 732942], from MIT Sloan's Program on Innovation in Markets and Organizations, and from the Change, Organizational Plasticity, and Evolution (COPE) project of the Danish Council for Independent Research [Contract 12-133546].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/stsc.2020.0120.

Keywords: categorization • cognitive frame • culture • leadership • performance

1. Introduction

This paper studies how shared cognition can create a link between organizational culture and organizational performance. We analyze how different cultures carrying different cognitive representations support stable differences in performance, and we explore challenges faced by a leader who seeks to change an organization's frame in order to improve performance.

In modeling organizational culture as shared cognition, we follow a long tradition. Pettigrew (1979) describes organizational culture as the "system of terms, forms, categories and images [that] interprets a people's own situation to themselves" (p. 574). Similarly, Schein (2010) argues that organizational culture creates "mindsets and frames of reference … [that are] invisible and to a considerable degree unconscious" (p. 14). And a more specific tradition supports our particular modeling approach, viewing shared categorization as a building block of culture. For example, Patterson (2014) asserts that "the basis of all cultural knowledge is our capacity to categorize" (p. 8), and Denzau and North (1994) discuss "a culturally provided set of categories" (p. 5).

Our model assumes that an organization's members share a *cognitive frame* that distills the complex world around them into a finite number of categories we call *situations*. Consistent with much of the literature in cognitive science, agents in our model are unaware that they distill the world through the categories of such frames: they know only (i) the set of situations that could arise and, when an underlying state of the world is realized, (ii) which of these situations has arisen. In this sense, agents in our model ask themselves, What situation is this? (Goffman 1974, March and Olsen 1983) and reach answers determined by their frame.

We close the model by assuming not only that organizational culture determines how agents see the world but also that agents act rationally given what they perceive. The agents' shared frame thus determines a mapping from situations to their optimal actions. We obtain the following results.

First, in a one-shot interaction, the coarse representation induced by the categories of a frame results in a unique equilibrium that can either decrease or increase the parties' payoffs, compared with having full information about the environment. We say that an organization's shared frame may induce either a "fog of conflict" or a "fog of cooperation." This initial result is consistent with the argument that differences in organizational performance may stem from differences in cognitive frames.

Second, in a repeated interaction, standard arguments from repeated games allow the parties to increase their payoffs above the one-shot level if they are sufficiently patient. We focus on the opposite comparative static: fix the parties' discount factor and analyze how their frame affects their *highest equilibrium payoffs* in a repeated interaction. Holding discounting constant, there are again frames under which the parties' payoffs are higher (or lower) than under full information. Importantly, this is not a standard result about multiple equilibria, as follows.

Kreps (1990) proposed long ago that different equilibria in a repeated game might correspond to different corporate cultures (shared understandings of "how we do things around here") associated with different performance levels across plants and firms. Although highly suggestive, there is a concern with modeling performance differences as resulting from different equilibria in a given game: low performers know that better equilibria exist, and yet the model gives these parties no way to try to reach a better equilibrium and offers no rationale for why moving to a better equilibrium might be difficult. Our model formalizes one such difficulty: low performers are playing the best equilibrium they can perceive; reaching a better equilibrium would require changing the parties' frame.

Whether in the unique equilibrium of our static analysis or the best equilibrium in our repeated analysis, a unifying feature is that a difference in frames can cause parties to (a) ascribe a state of the world to different situations and even (b) see different actions as optimal in situations that they describe equivalently. As a result, if we imagine low performers visiting a high-performing organization, the low performers may see their hosts achieving higher performance in ways that the low performers cannot understand how to imitate. This inimitability is necessary if an organizational culture is to create competitive advantage (Barney 1986).

Building on these classic contributions by Kreps and by Barney, our third set of results concerns the consequences and mechanisms of attempting to change cognitive frames. Regarding the consequences, we distinguish between *incremental change*, when the boundaries of situations change but the parties' optimal actions in given situations do not, and *radical change*, where both the boundaries and the optimal actions vary. We show that if revising the perceived boundaries of situations is more rapid than adjusting perceived optimal actions, radical change can induce either worse-before-better or better-before-worse performance paths (Repenning and Sterman 2002).

Regarding the mechanisms, we follow an established literature in psychology (Medin and Schaffer 1978, Nosofsky 1986) and develop a model of category formation based on the exemplars stored in agents' memories. We then explore how a range of actions by a leader may change the organization's frame, and we uncover some trade-offs associated with attempting such changes.

Finally, we offer an initial sketch of parties with different frames interacting with each other. We distinguish between *incremental discord*, when the parties apply the same rules of behavior to what they perceive, and *radical discord*, where even their rules of behavior are different. We explore how, after discord arises, the parties may enter into a dialogue and coordinate their actions using sincere communication.

In summary, we see our model as exploring an integrated account of aspects of organizational culture, performance, and leadership that have heretofore been considered separately (and typically not in formal models). We defer a detailed review of relevant literature to Section 7 after the development of our model, thereby facilitating the comparison with its precursors and alternatives.

2. The Model

Interactions that involve either pure common interest or pure conflict are archetypes of social and economic life, and a rich repertoire of cultural and linguistic resources is available to represent these poles. Yet many interactions mix collaborative and conflictual motives. In contexts of organizational interest—including team production, labor relations, strategic alliances, and interactions along the supply chain, among many others—agents must interpret the combination of collaborative and conflictual motives before deciding how to behave.

The polar cases are cognitively simple: agents' interests are perfectly correlated, either positively or negatively, making it easy for agents to process such games. But interactions that mix collaborative and conflictual motives are more difficult to apprehend. As Schelling (1960) remarks, "The difficulty is in finding a sufficiently rich name for the mixed game in which there is both conflict and mutual dependence. . . . [For example,] in the common-interest game we can refer to [the players] as "partners" and in the pure-conflict game as "opponents" or "adversaries"; but the mixed relation that is involved in wars, strikes, negotiations, and so forth, requires a more ambivalent term" (p. 89). More recently, this middle ground has begun to be populated with terms such as "co-opetition" (Brandenburger and Nalebuff 1996) and "frenemy."

In this section, we develop a simple model of how frames shape the interpretation of Schelling's mixed-motive interactions. Our model combines three features. First, we consider a space of games rather than a single game; second, agents have a coarse representation of their environment and act according to it; and third, they perceive the coexistence of motives as a blend of the common-interest and pure-conflict archetypes.

2.1. A Space of Games

Consider two archetypical games: a common-interest game (CI) on the left side of Figure 1 and a zero-sum game (ZS) on the right, with r in $[0,1]$.

These games represent basic situations of collaboration and conflict: for any r in $(0,1)$, they have dominant strategies; that is, there are unequivocal motivations to cooperate (H) or compete (L), respectively.

These two archetypes provide building blocks for more complex interactions. For example, as we will see in what follows, a prisoner's dilemma is a mixed-motive game that combines the two archetypes. More generally, we analyze a blended interaction involving two aspects: (1) the *reward* r from cooperation in the common-interest game and $1-r$ from defection in the zero-sum game and (2) the *prominence* p in $[0,1]$ for the common-interest component and $1-p$ for the zero-sum component.[1]

Figure 2 shows a typical game $G(r,p)$ generated from this blending of the archetypes in Figure 1.

The nature (and attractiveness) of a blended interaction depends on both the prominence p of the common-interest archetype and the payoffs that each archetype contributes to the blend. For example, suppose $p = ½$ (e.g., the two archetypes are equally likely). Then, for $r > ½$, the blended game in Figure 2 is a common-interest game, but for $r < ½$, the blended game is a prisoners' dilemma. More generally, for $r + p > 1$, the blended game has common interests (and H is the dominant strategy); for $r + p < 1$, it is a prisoners' dilemma (where L is the dominant strategy). In other words, r and p jointly determine whether the cooperative or competitive motive prevails.

We assume that r and p are independently and uniformly distributed on $[0,1]$, so the agents face a bidimensional *space of games* $\mathcal{G} = [0,1]^2$, depicted in Figure 3.

This space of games captures a wide range of interactions. As just one example, in relations between suppliers and assemblers of complex components, a component may blend standard and dedicated elements, where the former are associated with a competitive interaction but the latter with common interests (p), and the value added by dedicated elements (r) also matters in interpreting the blended interaction and deciding how to play.

We use this space of games to model interactions between agents who have limited ability to discriminate among the blended games in $\mathcal{G}$. We conclude this section by briefly considering the case where each party can discriminate any game g from $\mathcal{G}$ and then play the appropriate dominant strategy for g, strictly as a benchmark for comparing our model's results. Then each party's expected payoff is 1/6; see Proposition A.1 in the online appendix, where we have collected propositions and proofs.

2.2. Coarse Perception

We now impose our assumption that the parties have limited ability to discriminate among games in $\mathcal{G}$. In particular, we henceforth assume that each dimension of the space $\mathcal{G}$—the reward r in $[0,1]$ and the prominence p in $[0,1]$—is too rich to allow either party to perceive all its elements as distinct. Instead, each agent apprehends each dimension using a finite partition. For simplicity, we work with binary categorizations, defined by the thresholds $\hat{r}$ and $\hat{p}$. Thus, an agent categorizes r as high (h) if $r > \hat{r}$ and low (ℓ) if $r < \hat{r}$; similarly, p is high (h) if $p > \hat{p}$ and low (ℓ) if $p < \hat{p}$.

An agent with binary categorizations for r and p perceives four cells, as depicted in Figure 4.

We call each of the four cells a *situation*. A cell bundles together many games, all of which are perceived by a party as instances of the same situation. That is, when an agent faces a game from $\mathcal{G}$ and wonders what kind of situation she is in, only four answers come to her mind. For example, the upper right cell S_1 corresponds to the situation where both r and p are perceived as h. That is, S_1 involves both high reward and high prominence, close to the common-interest archetype with $r = p = 1$. The other three situations have analogous interpretations.

Figure 1. (Color online) A Common-Interest Game (Left) and a Zero-Sum Game (Right)

	H	L
H	r, r	$0, 0$
L	$0, 0$	$-r, -r$

CI

	H	L
H	$0, 0$	$-(1-r), (1-r)$
L	$(1-r), -(1-r)$	$0, 0$

ZS

Published in Strategy Science on February 11, 2021 as DOI: 10.1287/stsc.2020.0120.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 2. The Payoff Matrix for a Generic Game $G(r, p)$

	H	L
H	pr, pr	$-(1-p)(1-r), (1-p)(1-r)$
L	$(1-p)(1-r), -(1-p)(1-r)$	$-pr, -pr$

The *frame* of an agent is the collection of the situations that she perceives, identified by the threshold pair $(\hat{r}, \hat{p})$. We assume that an agent is unaware that she is framing. The model-builder, not the agent, knows that the agent (a) categorizes games and (b) does so via the threshold pair $(\hat{r}, \hat{p})$.

Until Section 6, we assume that two interacting parties share the same frame $(\hat{r}, \hat{p})$, with $0 < \hat{p}, \hat{r} < 1$. This is how we model (admittedly, quite starkly) the idea that these two parties have been shaped by the same organizational culture. A more realistic assumption might be that individuals' frames are more highly correlated within organizations than between organizations, but not necessarily perfectly correlated for individuals within a given organization. We offer an initial sketch of this alternative case in Section 6. Until then, we assume that two parties from a given organization share the same frame: in each of the four situations associated with the frame, the two parties perceive a single 2×2 symmetric game with payoffs equal to the expected payoffs from all the games ascribed to that situation. In this sense, agents' strategic understanding of the space of games $\mathcal{G}$ is coarsened into the four situations $S_1, S_2, S_3,$ and S_4 in Figure 4.

We label the upper right and bottom left situations S_1 and S_3 *consonant*, because their descriptors r and p are both high or both low: the reward r and the prominence p of cooperation are aligned. As we will see, each consonant situation has an unambiguous interpretation, with clear implications for agents' actions. By contrast, we say that the two situations S_2 and S_4 are *dissonant* because their descriptors are misaligned: one is high and the other is low. As we will see, dissonant situations have ambivalent interpretations whose resolution may diverge under (even slightly) different frames, producing different implications for agents' actions.

3. One-Shot Interaction

This section considers a one-shot interaction between two parties under a shared frame $(\hat{r}, \hat{p})$. We assume that their behavior is rational conditional on their frame: once they have interpreted a given situation, they are rational players within their interpreted world.

The expected payoffs to the first party (rescaled by a factor of 4) for each of the four situations perceived under the frame $(\hat{r}, \hat{p})$ are shown in Figure 5.

Conditional on the frame, the agents have correct beliefs about the distribution of payoffs in each situation; see the notion of *interpreted signal* in Hong and Page (2009). After playing a perceived situation, the parties receive the payoffs associated with the actual game $G(r, p)$ that was drawn and ascribe the difference between expected payoff and realized payoff to noise.

Rational behavior in the two consonant situations is unequivocal. Figure 5 shows that situation S_1 is always perceived as a CI game under any frame $(\hat{r}, \hat{p})$, so H (cooperate) is the dominant strategy. Similarly, situation S_3 is always perceived as a PD game, so L (defect) is the dominant strategy. Regardless of the frame, the rational behavior is to play H in S_1 and L in S_3. Intuitively, the two consonant situations S_1 and S_3 are adjacent to the common-interest and zero-sum archetypes, respectively: their interpretation (and the resulting behavior) matches their close proximity to an archetype.

Assuming that the frame satisfies $\hat{r} + \hat{p} \neq 1$, there is also a unique dominant strategy for the dissonant situations S_2 and S_4. This is characterized in the next proposition, which is a corollary of Proposition A.4 in the online appendix.

Proposition 1. *The unique dominant strategy for S_2 and S_4 is H if $\hat{r} + \hat{p} > 1$, and it is L if $\hat{r} + \hat{p} < 1$.*

Unlike consonant situations, the dominant strategy in dissonant situations depends on the frame.

Figure 3. (Color online) The Space $\mathcal{G}$ of Games, with Dominant Strategy H or L

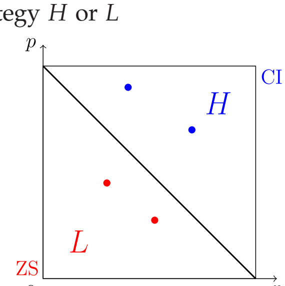
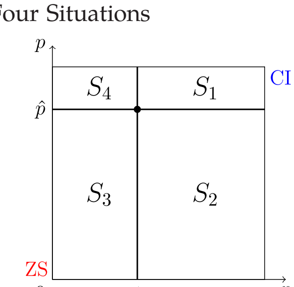


Figure 4. (Color online) A Categorization of the Game Space $\mathcal{G}$ into Four Situations



Published in Strategy Science on February 11, 2021 as DOI: 10.1287/stsc.2020.0120.
This article has not been copyedited or formatted. The final version may differ from this version.

Figure 5. Perceived Payoffs for the Row Player in the Four Situations Under the Frame $(\hat{r}, \hat{p})$

S_4

	H	L
H	$\hat{r}(1+\hat{p})$	$-(2-\hat{r})(1-\hat{p})$
L	$(2-\hat{r})(1-\hat{p})$	$-\hat{r}(1+\hat{p})$

S_1

	H	L
H	$(1+\hat{r})(1+\hat{p})$	$-(1-\hat{r})(1-\hat{p})$
L	$(1-\hat{r})(1-\hat{p})$	$-(1+\hat{r})(1+\hat{p})$

S_3

	H	L
H	$\hat{r}\hat{p}$	$-(2-\hat{r})(2-\hat{p})$
L	$(2-\hat{r})(2-\hat{p})$	$-\hat{r}\hat{p}$

S_2

	H	L
H	$(1+\hat{r})\hat{p}$	$-(1-\hat{r})(2-\hat{p})$
L	$(1-\hat{r})(2-\hat{p})$	$-(1+\hat{r})\hat{p}$

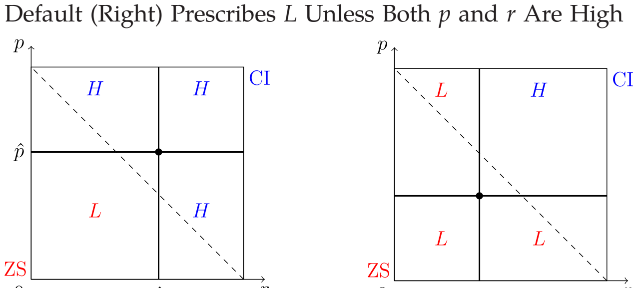
Intuitively, the descriptors are misaligned because a dissonant situation exhibits Schelling's mixed motives: the agent's frame resolves the ambivalent interpretation in favor of one strategy or the other.

Combining the dominant strategies over the four situations, we find two rules of behavior, shown in Figure 6. The first rule, depicted on the left, is optimal if $\hat{r} + \hat{p} > 1$: play H in any situation except S_3, and then play L; we call this rule *cooperation by default* because it prescribes playing H unless both r and p are low. The second, shown on the right, is optimal if $\hat{r} + \hat{p} < 1$: play H only in S_1 and otherwise play L; we call this rule *defection by default* because it prescribes playing L unless both r and p are high.

These rules of behavior will be useful in our exposition. Note that "default" refers to the situations, not to their probability of occurring. That is, cooperation by default prescribes H unless the exception (S_3) occurs, and defection by default prescribes L unless the exception (S_1) occurs. The overall probabilities with which H and L are played depend on further details beyond these rules of behavior.

Because the frame is shared and payoffs are symmetric, the parties will play the same strategy in a given situation. If $\hat{r} + \hat{p} > 1$, they will cooperate by default, playing (H, H) in all situations except (L, L) in S_3; if $\hat{r} + \hat{p} < 1$, they will defect by default, playing (L, L) in all situations except (H, H) in S_1. In sum, different frames can induce different strategy profiles when parties encounter dissonant situations.

Figure 6. (Color online) Cooperation by Default (Left) Prescribes H Unless Both p and r Are Low; Defection by Default (Right) Prescribes L Unless Both p and r Are High



Having computed optimal strategies, we next analyze how the parties' expected payoffs depend on the thresholds $(\hat{r}, \hat{p})$ of their shared frame. First, payoffs change continuously in $(\hat{r}, \hat{p})$ if the variation in thresholds does not change the parties' rule of behavior; second, if the rule of behavior switches, then there is a discontinuous change in payoffs.

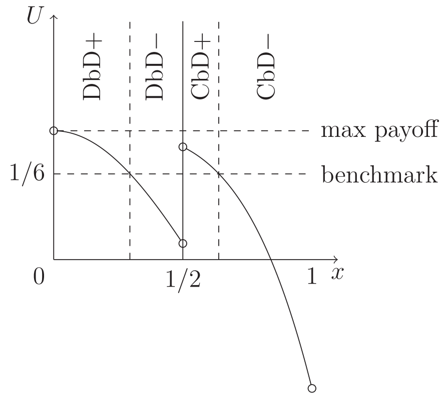
Proposition A.5 in the online appendix gives the expected payoff to each party as a function of $\hat{r}$ and $\hat{p}$. As an example, suppose $\hat{r} = \hat{p} = x$ so that a change in x makes both thresholds shift in lockstep. The parties choose cooperation by default (denoted by CbD) for $x > 1/2$ and defection by default (DbD) for $x < 1/2$. Figure 7 shows the payoff to each party as a function of x.

Within each default-rule region, payoffs continuously decrease in x. On the other hand, moving x rightward across 1/2 implies an abrupt increase in payoffs, as the parties switch from defection by default to cooperation by default. Nonetheless, depending on x, the former rule may outperform the latter.

Figure 7 also shows that framing games as situations can either help or hurt the parties' payoffs, relative to the benchmark case where each game is perceived as distinct: the benchmark payoff of 1/6 cuts across the payoff curve. Intuitively, one may think of the frame as creating a fog that confounds different games into a single situation, forcing a party to deal with all such games in one way. Depending on the frame, the result is either a *fog of conflict* (marked as a minus sign), under which agents achieve lower expected payoffs than they would under full information, or a *fog of cooperation* (marked as a plus sign), under which expected payoffs are higher. Note that either fog can occur under either rule of behavior, so frames evidently do more than determine rules of behavior.

Beyond the special case of $\hat{r} = \hat{p} = x$ shown in Figure 7, we can identify which frames generate which kind of fog. See the online appendix, where Proposition A.6 states a formal characterization followed by a visual summary. The main message is that a tiny

Figure 7. Payoffs as a Function of x When the Frame Is $(\hat{r}, \hat{p}) = (x, x)$



Note. The labels "DbD" and "CbD" identify the dominant strategy; the modifier "+/−" denotes payoffs higher or lower than the benchmark.

change in the threshold(s) that causes a switch in the rule of behavior yields an abrupt change in payoffs. This discontinuity motivates part of our discussion about changing frames in Section 5.

As one way to summarize this static model, imagine two parties who share a low-performing frame visiting two other parties who share a high-performing frame. All parties perceive situations in terms of their own frames, and the low performers observe the actions chosen by the high performers. For now, we simply consider what the low performers will see and what they might then infer; we defer discussions of (a) attempts to change frames and (b) interactions between parties with different frames until Sections 5 and 6, respectively.

Consider the discontinuity at $x = 1/2$ in Figure 7, and suppose that the low- and high-performing frames have $x_\ell < 1/2 < x_h$, with x_ℓ and x_h close but on different sides of 1/2. The high-performing frame supports cooperation by default, as in the left panel of Figure 6, whereas the low-performing frame supports defection by default, as in the right panel of Figure 6. The low performers see dissonant situations as PD games and hence expect (L, L), whereas the high performers see dissonant situations as CI games and so choose (H, H). The low performers will be mystified by the visit: for example, when they see PD and expect (L, L), they observe their hosts frequently playing (H, H) (namely, in the games when both sets of parties perceive dissonant situations). Further confusion, albeit of an opposite kind, occurs over those few games that low performers see as the consonant CI situation but high performers perceive as the consonant PD situation: when these games occur, the low performers expect (H, H) but observe their hosts playing (L, L).

If the low and high performers discuss what they saw or why the high performers acted as they did, the low performers will occasionally discover that the high performers saw different situations, and they will frequently discover that the high performers perceived the same situation but considered different actions to be optimal. Neither discovery would necessarily make them aware that anyone perceives the world coarsely (not to mention differently so). In short, the difference in cognitive frames may be an inimitable source of competitive advantage (Barney 1986).

To summarize this section, we see our static model as a small but novel contribution toward understanding widespread evidence of differences in cooperation.[2] It is common to interpret such differences in cooperation as arising from differences in preferences; our model provides a complementary explanation based on differences in cognition—specifically, differences in interpretation.[3] Although we do not expect our simple model to capture this wide range of empirical evidence, we believe that cognition (and especially interpretation) can offer a promising explanatory approach.

4. Repeated Interaction

Having constructed a model where shared frames shape behavior in static situations, we next consider the case of infinitely repeated interactions. Under any frame, the consonant situation S_3 is perceived as a PD. Furthermore, if $\hat{r} + \hat{p} < 1$, then the dissonant situations S_2 and S_4 are also perceived as PDs. In a repeated interaction, familiar logic might allow the parties to cooperate in some or all of these PDs, even if they would defect in a one-shot interaction.

We analyze such opportunities for long-term cooperation using a multiperiod model where in each period the stage game is randomly drawn from the space $\mathcal{G}$ of games and perceived as one of four situations under the shared frame $(\hat{r}, \hat{p})$. As in the static model, given their frame, the parties have correct beliefs: before a game is drawn in a given period, the parties expect to face situation S_1 with probability $(1 - \hat{r})(1 - \hat{p})$, situation S_2 with probability $(1 - \hat{r})\hat{p}$, situation S_3 with probability $\hat{r}\hat{p}$, and situation S_4 with probability $\hat{r}(1 - \hat{p})$. We assume that the parties have the same discount factor $\delta < 1$, and we rescale their discounted payoffs by a factor $(1 - \delta)$ to make them comparable to the one-shot payoffs.

We consider subgame-perfect equilibria where an unexpected defection (i.e., playing L when H was expected in a PD situation) triggers Nash reversion thereafter (i.e., defection in all future PD situations and cooperation in all future CI situations).

There are two cases of interest. The first is *full cooperation*, when agents play (H, H) across all situations. The second case is *improved cooperation*, when the static model leads to defection by default but the repeated interaction can support cooperation by default; that is, in the repeated game, players switch from defection to cooperation when facing dissonant situations but not when facing situation S_3.

Recall that the rule of behavior in the static model is cooperation by default if $\hat{r} + \hat{p} > 1$ and defection by default if $\hat{r} + \hat{p} < 1$. As previously mentioned, we reduce the number of parameters by assuming $\hat{r} = \hat{p} = x$; then the two rules obtain for $x > 1/2$ and $x < 1/2$. When $x > 1/2$, the static model leads to cooperation by default, and we study when the repeated interaction may support full cooperation. When $x < 1/2$, the static model leads to defection by default, and we study when the repeated interaction may support either full cooperation or improved cooperation.

Consider $x > 1/2$: the only situation perceived as a PD is S_3. The frame generated by x has three effects. First, because S_3 occurs with probability $\hat{r}\hat{p}$, and we assume $\hat{r} = \hat{p} = x$, in each future round the probability that the PD situation occurs is x^2: the greater the x is, the larger the PD situation looms. Second, when the PD situation does occur and the other party is expected to cooperate, the temptation to play defection (L) instead of cooperation (H) is decreasing in x. Finally, the threat of a long-term payoff loss from Nash reversion after defection is increasing in x.

These effects of shared cognition on the perceived frequency of PD situations and on the perceived relative strength of temptation versus punishment all influence the viability of long-term cooperation. Nevertheless, the familiar intuition that a sufficiently high δ supports full cooperation survives: if

$$\delta \geq \frac{2-2x}{2-2x+x^4},$$

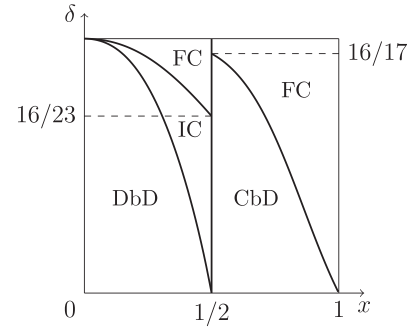
then there is a Nash-reversion equilibrium where the parties play H in situation S_3; see Proposition A.7 in the online appendix. This is illustrated in Figure 8 for $x > 1/2$: given δ and x, either the parties can sustain full cooperation (FC) across all situations or they cooperate by default (CbD), which, given $x > 1/2$, is the rule from the static game. In particular, for $\delta \geq 16/17$, sustaining full cooperation is possible for any value $x > 1/2$.

Consider now $x < 1/2$, when the static model yields defection by default. Proposition A.8 in the online appendix demonstrates a richer result. Let

$$\overline{\delta}_1(x) = \frac{2-2x}{2-2x+2x^2-x^4} \quad \text{and}$$
$$\overline{\delta}_2(x) = \frac{1-2x}{1-2x+2x^2-2x^4}.$$

In a repeated interaction, Nash reversion may be used to support full cooperation across all situations if $\delta \geq \overline{\delta}_1(x)$ and to support improved cooperation (IC) if $\delta \geq \overline{\delta}_2(x)$. Clearly, full cooperation is harder to achieve because $\overline{\delta}_1(x) \geq \overline{\delta}_2(x)$ for all $x < 1/2$. If neither inequality holds, the parties are stuck with defection by default as in the static model.

Figure 8. Best Feasible Cooperation in the Repeated Interaction



Note. The labels "DbD," "CbD," "IC," and "FC" identify the strategy profiles yielding the highest payoffs under Nash reversion for a frame x and a discount factor δ.

Our analysis reiterates the familiar theme that repetition and patience may allow the parties to achieve higher payoffs than in the static model. The novel point here is that performance differences may arise from differences in shared frames, even when all parties share the same discount factor and are playing the best repeated-interaction equilibrium they can, given how they perceive the space of games.

This novel point is illustrated most vividly if we fix a discount factor $16/23 < \delta < 16/17$. Then Figure 8 shows that as x progresses from 0 to 1, the best outcome that parties can sustain in a repeated interaction changes from defection by default to improved cooperation (i.e., cooperation by default), to full cooperation, then back to cooperation by default, and to full cooperation again—all for the same discount factor.

For comparison, consider the benchmark case where the parties can distinguish all the games in $\mathcal{G}$. Proposition A.9 in the online appendix shows that in the benchmark case, Nash reversion supports full cooperation if $\delta \geq 12/13$. Because this value of δ is between 16/23 and 16/17, in repeated interaction there are values of x for which coarse perception creates a fog of cooperation (where full cooperation is feasible under framing but not without) as well as values of x for which it creates a fog of conflict (where full cooperation is not feasible under framing but is without).

In short, even with a shared discount factor, differences in frames can cause parties to disagree about the best equilibrium feasible in the repeated game. For example, in Figure 8, consider a discount factor between 16/23 and 16/17 and two values of x below 1/2—one value of x such that those parties see improved cooperation as the best feasible equilibrium in the repeated game and another (larger) value of x such that those parties see full cooperation as feasible. In this example, all parties believe that there is an equilibrium in the repeated game that outperforms spot play, but the former think that (H, H) cannot be sustained in situation S_3, whereas the latter think that it can. The low performers might diagnose this disagreement about equilibrium strategies as a disagreement about the probabilities of (or the payoffs in) the situation S_3. Thus, as in our static model, disagreement about equilibrium in the repeated interaction might *not* cause the parties to imagine that they perceive the world coarsely (not to mention differently so).

We viewed our static model as a small step toward understanding a broad set of findings concerning widespread evidence of differences in cooperation. By contrast, we see our repeated model as a larger step toward a much more specific goal: moving beyond the intriguing suggestion by Kreps (1990) that different equilibria in a repeated game might correspond to different corporate cultures. There is a concern with modeling performance differences where low performers know that better equilibria exist, and yet the model gives these parties no way to try to reach a better equilibrium and offers no rationale for why moving to a better equilibrium might be difficult. Our model formalizes one such difficulty: low performers are playing the best equilibrium they can perceive; reaching a better equilibrium would require changing the parties' frame, to which we now turn.

5. Changing Frames

Our basic model assumes that, within a given organization or group, frames are (a) fixed and (b) shared. This section relaxes the first assumption and explores some difficulties in changing frames to improve performance. Section 6 relaxes the second and sketches how parties with different frames might resolve their discord.

Within this section, we first consider the consequences of frame change. In particular, in Section 5.1 we distinguish between *incremental* versus *radical* change in the frame: the former modifies only the boundaries of situations, whereas the latter changes also the optimal actions in some situation. We then build on this distinction to illustrate some risks of attempting change, including "worse-before-better" dynamics (Repenning and Sterman 2002).

Section 5.2 turns to mechanisms behind frame change. We offer a basic model for the formation and evolution of a frame, inspired by the psychological literature on categorization, and we then analyze its implications for a leader who seeks to manage her followers' frame.

In most of this section's discussion of frame change we focus on the long run, after the parties have accomplished not only (i) appraisal (i.e., new thresholds are in place) but also (ii) evaluation (i.e., beliefs about payoffs in new situations are in place). Both appraisal and evaluation could take time, and both deserve their own analyses, but we cannot conduct those here.

5.1. Consequences of Frame Change

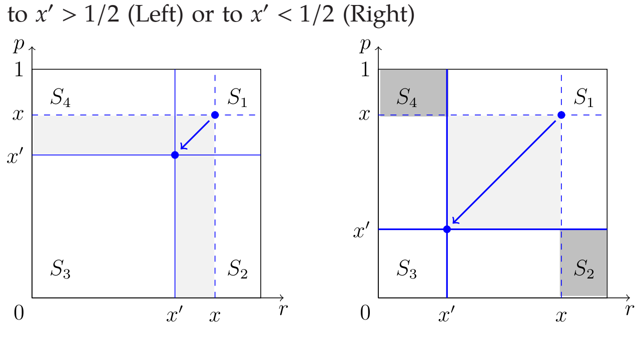
We begin by distinguishing between two kinds of frame change: incremental versus radical. Imagine that the current thresholds for the frame are $\hat{p} = \hat{r} = x$. Suppose that the initial common threshold $x > 1/2$ shifts down to $x' < x$. We distinguish two cases: (a) $x' > 1/2$ and (b) $x' < 1/2$.

The case where $x' > 1/2$ is shown in the left panel of Figure 9. When the parties' frame changes, they recategorize some games as different situations.

Because $x' < x$, the probabilities that the parties perceive S_1, S_2, and S_4 increase, and the probability that they perceive S_3 decreases. On the other hand, because $x' > 1/2$, the rational rule of behavior does not change: it remains cooperation by default, and S_1, S_2, and S_4 are still played cooperatively. In short, the parties' behavior changes only because the parties recategorize some games from S_3 to S_1, S_2, or S_4 and thus switch behavior from L to H; these games correspond to the light gray area in the left panel of Figure 9.

The case where $x' < 1/2$ is shown in the right panel of Figure 9. As before, the parties recategorize some games from S_3 to S_1 and switch their behaviors in these games from L to H; see the light gray area in the right panel of Figure 9. Notably, there is now a second source of change in behavior: the parties used to play cooperation by default but, when $x' < 1/2$, they switch to defection by default and change actions (from H to L) in the dissonant situations S_2 and S_4. Hence, games ascribed to dissonant situations *both before and after the change in frame* are now played differently; see the dark gray areas in the right panel of Figure 9.

Figure 9. (Color online) The Changes in Perceived Situations After Lowering Thresholds from $\hat{p} = \hat{r} = x > 1/2$ to $x' > 1/2$ (Left) or to $x' < 1/2$ (Right)



Clearly, any shift in the frame's boundaries changes what states of the world parties ascribe to particular situations (i.e., the size of the frame's cells). More important, after they learn the payoffs associated with these new situations, the frame change may also affect which actions they choose in specific situations (their rule of behavior). We call a change in frame *incremental* when the agents' rule of behavior does not change, and we call it *radical* when it does.

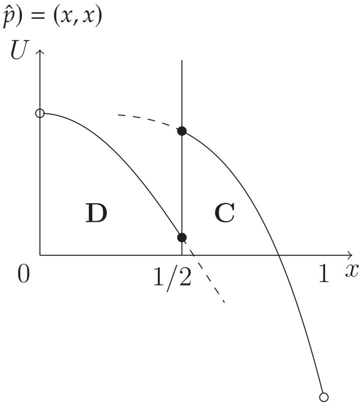
We now imagine a *leader* who seeks to change a group's frame to improve its performance. The group members (i.e., the "parties" in Sections 3 and 4, who share a given frame) are referred to as the *followers*. We assume that the followers' frame changes in the same way at the same time for both followers. We also assume that the leader knows the whole model, including that the followers perceive the space of games as situations generated by a frame $(\hat{r}, \hat{p})$.

Returning to the example from earlier in this section where $\hat{r} = \hat{p} = x$, suppose that the leader may lower or raise the threshold x from its initial value. Figure 10 shows the expected payoff U to each follower, analogous to Figure 7; see Proposition A.5 for details.

Suppose that the leader lacks full control of the followers' new frame.[4] For example, assume that the leader controls the direction of frame change—either W (shift x westward) or E (shift x eastward)—but not its magnitude.

Suppose that the initial threshold is $x > 1/2$, so the followers initially use cooperation by default. The action E is dominated by staying put. The action W, on the other hand, is risky: a small shift to a new threshold $x' > 1/2$ increases payoffs, but a larger shift to a new threshold just below 1/2 discontinuously decreases payoffs. Thus, under cooperation by default, attempts at *incremental change* (i.e., a mild reduction of the threshold) are worthwhile, unless the risk of a radical change—with followers switching to defection by default—is too high.

Figure 10. Follower's Payoff as a Function of x When the Frame Is $(\hat{r}, \hat{p}) = (x, x)$



Alternatively, suppose that the initial threshold is $x < 1/2$, so the followers initially use defection by default. The action W now increases payoffs through incremental change, whereas the action E may lead to a radical change: if the threshold crosses the 1/2 barrier, then payoffs increase substantially, but if the threshold moves right without crossing the barrier, then payoffs are worse than before.

A second risk of frame change concerns the speed of change. As noted previously, suppose the followers go through two steps after the leader's intervention: (i) appraisal (whether and how much thresholds change) and (ii) evaluation (how long it takes before followers update their beliefs about payoffs in a reconfigured situation). Here, we suppose that (i) occurs instantly but there is delay in (ii). That is, the followers exhibit inertia (because of a delay in updating beliefs about payoffs) and hence stick to their previous rule of behavior for a while.

Under incremental change, the original rule of behavior is still optimal, so there are no delayed effects on behavior (even if evaluation is slow, as postulated in this example). But suppose that $x < 1/2$ and the leader achieves a radical change to $x' > 1/2$. The followers were using defection by default, so after a radical change crossing 1/2 from the left, the payoff stays on the lower dashed curve until the followers complete the evaluation step (ii), after which behavior changes and the payoff jumps up to the higher solid curve—at the new threshold $x' > 1/2$. In short, cognitive inertia in the evaluation of a radical change in frame may cause a transient decline in performance before producing its positive effects: this dynamic is called an *implementation dip* by Fullan (2001) and "worse-before-better" by Repenning and Sterman (2002, p. 279).

Conversely, suppose $x > 1/2$, and the leader intends incremental change to the left. If the frame changes too far, becoming a radical change to $x' < 1/2$ (and close to 1/2), the change would enjoy an initial success before its ultimate failure: transient payoffs would be on the upper dashed curve, but long-run payoffs would be on the lower solid curve.

In much of the business strategy and organization literature, radical change is conceived as "long jumps" in some space of organizational features (Levinthal 1997, Roberts 2004). In this view, the costs and risks of change stem from the need to reach distant points by small steps. Our notion of radical change is different, because it refers to a switch in behavior rules that engenders a discontinuity in performance. In our model, when an organization is close to the point of discontinuity, even a small step may cause radical change; when this occurs, the asynchronous update of frames and behavior rules can generate an implementation dip.

Published in Strategy Science on February 11, 2021 as DOI: 10.1287/stsc.2020.0120.
This article has not been copyedited or formatted. The final version may differ from this version.

5.2. Managing Frames

Schein (2010) defines the culture of a group as "a pattern of shared basic assumptions" (p. 18) and argues that the essence of leadership is creating and managing this culture. We take the frame of organization members as such a shared basic assumption and now consider how a leader might attempt to change an organization's frame.

First, we enrich our basic model by providing a mechanism for the formation and evolution of a frame, inspired by the psychological literature on categorization. Then we analyze some trade-offs faced by a leader who seeks to change her followers' frame.

There is an established literature on categorization in the cognitive sciences, with a variety of formal models that describe or predict how human subjects organize their sensory experience into categories; see Pothos and Wills (2011). Two dominant approaches to categorization are prototype theory (Rosch 1973, Osherson and Smith 1981) and exemplar theory (Medin and Schaffer 1978, Nosofsky 1986). The first postulates that there is some central element (the prototype) for each cluster of similar objects; a novel stimulus is attributed to the category associated with the closest prototype. The second stipulates that each category is associated with some exemplars stored in memory rather than by an abstract summary representation; a novel stimulus is attributed to the category that maximizes the stimulus's overall similarity with the category's set of exemplars. The huge literature comparing these two (and other related) approaches has produced mixed evidence, depending on the fine details of the specific applications. We use a mixed approach that is simpler to present.

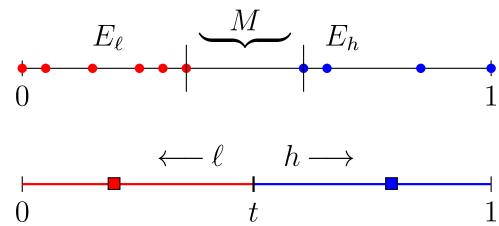
We impose two simplifications. First, the categorization is deterministic: a stimulus is uniquely assigned to a category. Second, all the exemplars lie on the main diagonal; that is, for each exemplar $e = (r, p)$, we have $r = p = x$ for some x in $(0, 1)$. Abusing notation, we write x to denote both the exemplar and its coordinates.

There are only two categories, ℓ (low) and h (high), with *exemplar sets* E_ℓ and E_h. Each exemplar set contains its archetype: the zero-sum game $(0, 0)$ is in E_ℓ, and the common-interest game $(1, 1)$ is in E_h. The cardinalities of the exemplar sets $n_\ell \geq 1$ and $n_h \geq 1$ may be different. We assume $\max E_\ell < \min E_h$ so that there is a *middle ground* M separating E_ℓ from E_h. Intuitively, the middle ground is where the tug-of-war between the exemplars for ℓ and for h may be usually summarized into a threshold $\hat{x}$. The case where $n_\ell = 6$ and $n_h = 4$ is shown in the top panel of Figure 11.

For the two exemplar sets E_ℓ and E_h, we compute the *average values* $\bar{e}_\ell$ and $\bar{e}_h$, depicted as squares in the bottom panel of Figure 11.

We use $\bar{e}_\ell$ and $\bar{e}_h$ as prototypes for the two categories ℓ and h: a novel stimulus x is categorized as low (ℓ) if it is closer to $\bar{e}_\ell$ and as high (h) if it is closer to $\bar{e}_h$. This construction is equivalent to using the threshold $t = (1/2)\bar{e}_\ell + (1/2)\bar{e}_h$ and categorizing x as ℓ if $x < t$ and as h if $x > t$. This yields a partition for ℓ and h into adjacent intervals, as shown previously.

Figure 11. (Color online) Graphical Representation of Exemplar Sets E_ℓ and E_h



Putting back the diagonal in the unit square, the frame with thresholds $\hat{r} = \hat{p} = t$ is shown in Figure 12.

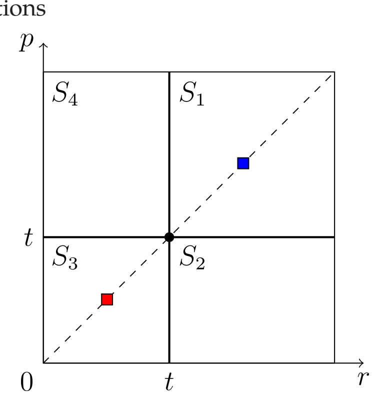
We assume that the leader knows the whole model, including how exemplars stored in (or removed from) the organization's memory affect the followers' frame. We imagine the leader attempting to change the followers' frame by taking actions such as using specific language, or extolling certain behaviors, or telling particular stories.

We illustrate four considerations for a leader, each through its own simple vignette. The goal is to present simple examples; each vignette might be subject to a deeper analysis. The leader seeks to maximize the sum of the followers' static payoffs: up to an irrelevant constant, the leader's payoff is $V(\hat{x}) = 1 - 2\hat{x}^4$ if $\hat{x} > ½$ and $1 - 4\hat{x}^2 + 2\hat{x}^4$ if $\hat{x} < ½$.

5.2.1. Changing Exemplars.

An obvious way to shift category boundaries is to add exemplars. For example, the acquisition of Lotus by IBM marked a key moment in chairman and chief executive officer Lou Gertstner's culture change at IBM, shifting the boundaries of accepted sources of new technology to include acquisitions (Gerstner 2002).

Figure 12. (Color online) A Categorization with Four Situations



In addition to changing frames by adding exemplars, in our model, frames can also change by deleting exemplars—"unlearning," as Hedberg (1981, p. 3) calls it. Note that deletion is constrained by the existing exemplars in organizational memory. If the leader needs to fine-tune the frame to the threshold $\hat{t}$, none of the available deletions may get it right, whereas $\hat{t}$ may be feasible with the right addition(s) of exemplars.

5.2.2. The Weight of Memory. The size n_i of an exemplar set E_i dictates the degree of resistance to changing the frame: when n_i is large, the addition or deletion of individual exemplars has a smaller effect. An organizational memory that stores a larger number of exemplars has a dampening effect on the addition (or deletion) of an exemplar. Consider an organization spread over two sites. Assume both sites have the same average values $\bar{e}_\ell, \bar{e}_h$ for the exemplar sets and hence the same threshold t and the same frame. But suppose the first is a "brownfield" with a deep memory (higher values for n_i) and the second is a "greenfield" with a shallow memory (lower values for n_i). Then the frame at the first site is more resistant to change. If the leader attempts a frame shift for the whole organization by adding a new exemplar, each site will process it with respect to its own memory, leading to different new frames at the two sites.

5.2.3. Fictitious Exemplars. Organizations learn not only from direct experience but also from various forms of vicarious learning (Levitt and March 1988), including stories (Selznick 1957). When actual occurrences cannot serve as useful exemplars, a leader may attempt frame change by telling stories as fictitious exemplars. Compared with an actual exemplar, stories may have vagaries of interpretation that make their effects more difficult to predict (Boje 1991). Whether a story is worth telling or not will depend on trade-offs of the potential advantages of a new threshold versus the risk that the followers interpret the story differently from the leader's intention. Section 5.2.3 in the online appendix reports an example illustrating this point.

There are cases, especially when the organizational memory is deep, in which the leader may need to add an extreme exemplar to achieve desired change. One way is to introduce an *outlandish exemplar* describing a case that is very expressive but possibly infeasible in practice. Allegories, analogies, and parables are constructs that need not be factually possible but may effectively promote a different viewpoint.

Our model can be extended to accommodate outlandish exemplars. Recall the assumption x in $[0, 1]$. Adding an actual exemplar from $[0, 1]$ to E_h can move the threshold up at most to $t' = (n_h \bar{e}_h + 1)/(n_h + 1)$. By contrast, suppose an outlandish exemplar for E_h is associated with $x = 1 + \alpha$, where $\alpha > 0$ is the fictitious excess over the highest feasible exemplar. If the leader succeeds in adding x to E_h, then the value of the new threshold is $t' + \alpha/(n_h + 1)$. Note that the amount by which the fictitious excess α changes the threshold is mediated by the depth n_h of the organizational memory.

The followers may interpret an outlandish exemplar with fictitious excess α as inspirational as intended by the leader, or they might reject it as preposterous. One could model these possibilities by assuming that the probability $\Pr(\alpha)$ that an outlandish exemplar $x = 1 + \alpha$ is assimilated into the organizational culture is a decreasing function of α. There would then be a trade-off between the potential impact of an outlandish exemplar and the risk of its rejection.

5.2.4. Acting Now or Later. If the leader needs to wait for an actual occurrence to use it as an exemplar, she might face a real-option problem. Using the current (candidate) exemplar for incremental change may not only improve organizational performance but also increase organizational inertia, both by making organizational memory deeper and by shrinking the set of future exemplars that can achieve radical change. For example, in Figure 10, if the current threshold x is just below 1/2, adding an exemplar that moves the threshold slightly left to $x' < x$ improves performance immediately but also increases the cardinality n_ℓ of one exemplar set and increases the distance that the threshold must travel to achieve radical change (to $x'' > 1/2$). Put more evocatively, because "inertia of organizational capabilities is the source of the value of real options" (Kogut and Kulatilaka 2001, p. 746), a leader may prefer to pass on an incremental improvement now and preserve the opportunity to trigger radical change later. Waiting to act may be an investment in strategic flexibility. Section 5.2.4 in the online appendix provides an example.

6. Discord

We have thus far explored the consequences of assuming that parties from the same organization share the same frame. The general case where agents have different frames is outside the scope of this paper, but this section takes a first step toward analyzing how parties with different frames might react to the discovery that they do not view the world the same way and attempt to resolve their discord.

As mentioned previously, the parties are not aware that they are using frames to categorize situations. In the previous sections, the parties share the same frame and hence always take the same action. In this section, however, the parties have different frames, so their actions might be miscoordinated.

We assume that if the parties' actions are miscoordinated, they enter into a sincere dialogue. For example, they might ask each other, "What situation did you see?" Or they might ask, "What action to you think is optimal in that situation?" We explore the idea that after miscoordination, the parties agree that in future they will truthfully report to each other what situation each perceives before either chooses an action.

Imagine that the current thresholds for the frames of the two parties are $\hat{p}_1 = \hat{r}_1 = x_1$ and $\hat{p}_2 = \hat{r}_2 = x_2$, with $x_1 < x_2$. We distinguish two cases: (a) $1/2 < x_1$ or $x_2 < 1/2$ and (b) $x_1 < 1/2 < x_2$. In the first case, the agents use the same behavior rule; in the second case, they use different behavior rules.

The first case (for $x_1 > 1/2$) is shown in Figure 13, where the two panels show the different individual frames. Before miscoordination occurs, because their frames have thresholds above 1/2, both parties use cooperation by default. However, because their individual frames are different, there are games that the parties play differently. We say that a discord is *incremental* when the parties are using the same behavior rule, but their rules are supported by different frames.

The left panel in Figure 14 depicts the refinement of the four-cell individual frames into a nine-cell *joint categorization*, where each cell shows the strategy profile supported by the two individual frames—before the parties encounter miscoordination. Each cell in an individual frame is identified by two binary pieces of information: high/low r and high/low p. In comparison, each cell in the joint categorization is identified by four binary pieces of information: high/low r and high/low p for either agent.

There are three cells where parties will miscoordinate (colored in light gray) and play *HL*. Corresponding to the light gray area, the probability of miscoordination is $x_2^2 - x_1^2$: if two parties have slightly different frames (and x_1 is close to x_2), it may take a long time before they miscoordinate on *HL*.

Once miscoordination takes place and the parties debrief, they find out that both of them were using cooperation by default. Miscoordination has occurred

Figure 13. (Color online) Two Frames Inducing the Same Behavior Rule

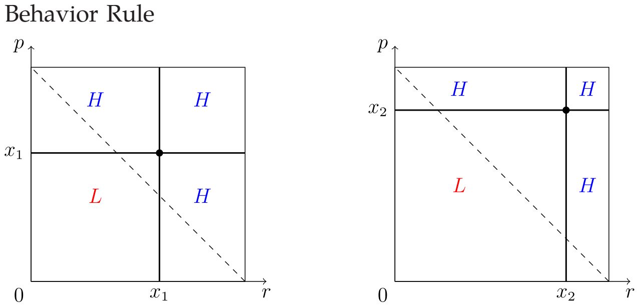
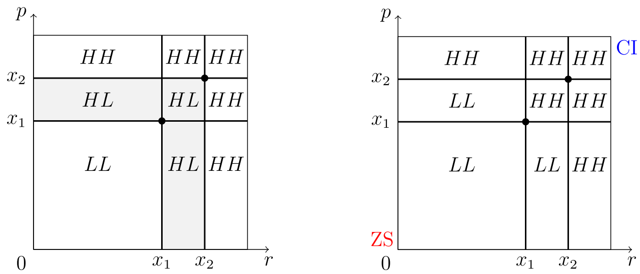


Figure 14. (Color online) Incremental Discord (Left) and Its Resolution Under Truthful Communication (Right)



because the first party saw high dimensions but the second party saw low dimensions. Being unaware of their cognitive mechanisms, they cannot elicit the source of their incremental discord. But if they exchange truthful information about their interpretations *before* the next play, they can pool their interpretations and use the joint categorization shown in the left panel of Figure 14.

We analyze the steady state after the parties learn their (expected) payoffs for each possible action in each cell of the joint categorization. At that point, we assume that agents are myopic optimisers, who decide how to play their current situation independently of future interactions, and that they resolve their discord by conditioning their individual strategies on the joint categorization. Using Proposition A.4 in the online appendix, it turns out that the dominant strategy for either agent is to play *H unless* at least three of the four signals are low, as shown in the right panel of Figure 14. We call this rule *cooperation by consensus*. In this terminology, an incremental discord over cooperation by default is resolved by moving to cooperation by consensus.

So far, we have analyzed the subcase $1/2 < x_1 < x_2$. The other subcase $x_1 < x_2 < 1/2$ is similar, with incremental discord between two parties now using defection by default. When the parties rely on the joint categorization, the dominant strategy for either agent is to play *L unless* at least three of the four signals are high. An incremental discord over defection by default is thus resolved by moving to defection by consensus.

A different case occurs for $x_1 < 1/2 < x_2$—that is, when the frame of the first agent supports defection by default and the frame of the second supports cooperation by default. We say that a discord is *radical* when the parties are using distinct behavior rules, supported by different frames. One might expect only incremental discord when two parties have substantial shared experience in a single organization, and so miscoordination might be infrequent. By contrast, immediately after a merger or when a new boss or peer or subordinate is hired, there might be higher probability of radical discord, as we now analyze.

Figure 15 shows the three cells where parties miscoordinate under radical discord: they play *HL* in one of them (colored in light gray) and *LH* in two of them (colored in dark gray).

The probability of miscoordination is $(x_2 - x_1)^2 + 2x_1(1 - x_2) \geq 1/4$: miscoordination is quite likely to occur. After it takes place and the parties debrief, they find out that they are using different behavior rules. If they resolve discord using the same process as mentioned previously, they are led once again to cooperation by consensus (if $x_1 > 1 - x_2$, as in Figure 15) or defection by consensus (if $x_1 < 1 - x_2$). Intuitively, the type of consensus is driven by the agent who has a threshold closer to 1/2 and hence less extreme categories.

To summarize this initial analysis of parties with different frames, incremental discord between two agents who play cooperation (respectively, defection) by default is always resolved by moving to cooperation (respectively, defection) by consensus. Radical discord, instead, is resolved by moving to cooperation or defection by consensus, depending on which agent has less extreme categories.

7. Literature Review

This paper links categorization to organizational culture and performance. Taken singly, there are huge literatures on each of these three topics; any attempt to summarize them would be outside the scope of this paper. Instead, we proceed in three steps. First, we address the links between culture and performance. Then we discuss culture and categorization. Finally, we address issues of organizational change and the role of cognitive frames. We know of no other work linking all these topics in a unified approach. We close our review with a short foray into the game-theoretic literature.

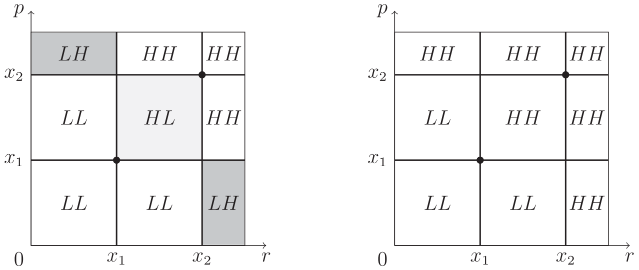
7.1. Organizational Culture and Performance

A stream of work in the management literature emphasizes the positive contribution of strong cultures to organizational performance; see Sørensen (2002) for a thoughtful review. Chatman and O'Reilly (2016) summarize recent evidence on how organizational culture affects performance and articulate the notion of culture strength and content. Our emphasis on culture as shared frames echoes the notion of strong culture and shows how its content may have diverging effects on organizational performance.

In a prescient paper, Barney (1986) analyzes the attributes an organizational culture must have to generate performance advantage—chiefly, inimitability, an issue we address in Sections 4 and 5. Weber and Camerer (2003) offer experimental evidence that cultural conventions are hard to reproduce and that cultural misalignment has performance implications.

Leibenstein (1982) conjectures—through informal use of prisoners' dilemma language—that underperforming enterprises might be stuck in defect-defect, whereas superior performers might have learned to play cooperate-cooperate. Kreps (1990) provides illustrative repeated-game models, highlighting gaps in the theory to be filled. Gibbons and Henderson (2013) connect Leibenstein and Kreps back to Barney by emphasizing that repeated-game models entail not just the familiar credibility problem (should you believe the promise being made?) but also an equally important clarity problem (is there a shared understanding of the promise being made?).

Figure 15. Radical Discord (Left) and Its Resolution Under Truthful Communication (Right)



7.2. Culture and Categorization

As early as 1952, anthropology had over 160 different definitions of culture (Kroeber and Kluckhohn 1952), but a clear definition of the relationships between culture and cognition emerged only later (D'Andrade 1995, Bender et al. 2010). It was not until the 1990s that cultural analysis broadly acknowledged cognitive science at its roots (Zerubavel 1991, 1997; Sperber 1996; DiMaggio 1997). This increasing emphasis on the shared cognitive aspects of culture echoes Geertz's (1973, p. 12) pithy "culture is public because meaning is."

As discussed in the introduction, some management scholars have noted the connection between cognition and culture since the onset of studies on organizational culture. And there is work in economics and political science emphasizing that shared mental models can be held by individuals with common backgrounds or experiences (Denzau and North 1994). Aoki (2001) explores how shifts in equilibria are associated with changes in the parties' "common cognitive representations" (p. 235). More recently, Hoff and Stiglitz (2016) have called for economic analyses to consider "cultural mental models [such as] concepts, categories, social identities, [and] narratives" (p. 26).

Our choice to focus on categories as basic cognitive entities is not arbitrary. Zerubavel (1991, pp. 1 and 3) argues that "the way we cut up the world clearly affects the way we organize our everyday life. . . . The way we draw lines varies considerably from one society to another as well as across historical periods within the same society." Recently, Hannan et al. (2019) have argued that the analysis of categories can provide guiding principles for cultural analysis.

7.3. Culture, Frames, and Organizational Change

In the broad literature on organizational change, culture and frames are recurrent themes (Burke 2017). However, in most cases, they are associated to inertial forces, fostering stability and triggering defensive resistance (Argyris 1985). More recent attempts to reconsider how frames can play an active role in promoting change (Kaplan 2008, Kellogg 2011) have cast frames as cognitive/political resources for aligning new organizational coalitions supporting change.

Using a top-down approach, empirical case studies of macroscale organizational change have focused on the role of corporate leadership as the key driver for the change of shared frames (Schein 1985); see, for example, Goodstein and Burke (1991) and Kotter and Heskett (1992) on British Airways, Gerstner (2002) on IBM, and Fiss and Zajac (2006) on the German corporate system. Across this literature, change is associated to agents who deliberately use discourse, stories, exemplar experiences, and incentives to modify the frames that shape how organizational actors perceive the world.

It is a recurrent theme in the management literature that organizational changes do not lie on a continuum but are better captured by a distinction between two different types of change: incremental versus radical (Greenwood and Hinings 1996), convergence versus reorientation (Tushman and Romanelli 1985), continuous versus discontinuous (Weick and Quinn 1999). Radical change is usually defined as the simultaneous change of several key domains of organizational activity (Gersick 1991, Romanelli and Tushman 1994), possibly necessitated by complementarities (Milgrom and Roberts 1995). The notion that radical change requires a cognitive discontinuity—a restructuring of frames—has been suggested in a number of contributions (Barr et al. 1992, Weick and Quinn 1999, Gavetti 2012, Werner and Cornelissen 2014). Our approach explicitly connects cognition and behavior: radical change is defined by a discontinuous change of behavioral patterns (the mapping of situations to actions) associated to a shift in frame boundaries. We are not aware that this perspective has been developed in the literature on culture, frames, and change.

7.4. Categorization in Game Theory

Categorization has appeared in the game-theoretic literature since Jehiel (2005), who considers single games where each player partitions the opponents' moves into categories. We focus on the case where the categorization spans many different games. Heller and Winter (2016) assume that agents simultaneously decide their own categorizations over games, committing to play the same strategy over the same category; in our model, instead, agents are unaware that they are framing. Samuelson (2001) and Mengel (2012) study alternative processes for the categorization of games that, different from our Section 5.2, are not inspired by empirical evidence. Bednar and Page (2007) demonstrate how different rules of behavior may spontaneously emerge when different games are bundled in the same category.

8. Conclusions

This paper offers a new perspective on how organizational culture might be a strategic resource generating persistent performance differences across firms. We build a theoretical framework that combines organizational and cognitive approaches through the assumption that an organization's members perceive the environment through a shared frame.

We provide several results. First, changes in the cognitive frame may induce differences in collective performance. Second, in a repeated interaction, the frame is as important as agents' patience in achieving cooperation. Third, changes in the frame may have starkly different consequences for performance, depending on how they affect the mapping of perceived situations to actions (incremental versus radical change). We show that radical change may create worse-before-better dynamics. Fourth, we also consider the formation of categories, exploring how a leader may act to change the followers' frame, including the timing of change and the effects of organizational memory and direct versus indirect experience. Finally, we show how parties with different frames may resolve their discord using sincere communication.

In future work, we intend to develop our framework in various directions. One deals with modeling the effects of communication on frames, considering how messages affect the salience of the dimensions over which situations are categorized. We hope to explore the role of language in leadership and organizational culture change. Another concerns cognitive misalignment, when parties understand that frames are imperfectly shared. We are interested in how the parties' efforts may lead to a repair or a collapse of their cooperation.

More broadly, we hope that over time our approach will shed new light on established but puzzling phenomena, as well as suggest new questions. In particular, we see three interesting areas where our approach might contribute: relational contracts, culture as strategy, and leadership.

First, there is a rich and growing literature on relational contracts (see Malcomson (2013) for a survey), but all the models we know are in equilibrium from the beginning. As a result, learning can produce delight or disappointment but never true surprise. We hope our future analysis of misaligned frames begins to capture such surprise. Another consequence of equilibrium models is that there is never any need to discuss strategies or intentions before the relationship begins. Because essentially no relationship in the real world begins without up-front discussions, it would seem useful for the theory to catch up with this fact, and we again hope that our analysis of misalignment will move in this direction.

Second, about culture as strategy, Barney's (1986) insight that culture must be inimitable if it is to create competitive advantage usually makes culture indescribable and/or taken for granted. By contrast, in our model, parties who share a frame have no problem talking—to themselves or to others—about their rule of behavior (i.e., their mapping from situations to actions), but parties with other frames will disagree at least about when different situations have been realized (as in Section 3) or whether proposed repeated-game strategies are equilibria (as in Section 4). The fact that parties are unaware of their framing prevents them from talking about their frames, even if they can communicate their strategies. We hope our future work on language and leadership provide an underpinning for Barney's inimitability.

Finally, the leader in Section 5 has a superior understanding of followers' framing. There are other models that imagine the leader knowing more than the followers do (e.g., in "leading by example"; Hermalin 1998). We hope to explore the case of a leader who has some (necessarily) private information about an idea of her devising, such as her strategic intent, but lacks the language to fully share it with her followers.

Acknowledgments

The authors thank Andreas Blume, Vincent Crawford, Emir Kamenica, Katherine Kellogg, Anton Kolotilin, Margaret Meyer, Wanda Orlikowski, Alessandro Pavan, Andrea Prat, Phil Reny, Joel Sobel, Marco Tolotti, and Catherine Turco, as well as seminar audiences at the Universitat Autònoma de Barcelona, Carnegie Mellon University, Columbia University, the European Summer Symposium in Economic Theory 2017, the Massachusetts Institute of Technology, the Max-Planck-Institut für Mathematik in den Naturwissenschaften, London Business School, the University of Padua, the Universitat Pompeu Fabra, the University of Siena, the Stanford Graduate School of Business, Vanderbilt University, and Washington University in St. Louis, for helpful comments.

Endnotes

[1] *Prominence* is the weight attributed by the agent to the common-interest archetype when perceiving a blended interaction. More concretely, our analysis can be seen as studying games where agents believe that (after they move) nature chooses the common-interest game with probability p and the zero-sum game with probability $1 - p$.

[2] Some of the field evidence points to differences in cooperation during evolution (Boyd and Richerson 2009) and among cultures (Henrich et al. 2005), communities (Ostrom 1990), firms (Leibenstein 1982), organizations (Schein 1985), and teams (Cole 1991).

[3] Experiments show that cultural frames cause individuals to perceive situations as "cooperative" or "competitive" (Keller and Loewenstein 2011) and how inducing different frames affects cooperation levels (Pruitt 1970, Liberman et al. 2004, Ellingsen et al. 2012).

[4] March (1981, p. 563) reminds us that "organizations are continually changing, [. . .], but change within them cannot ordinarily be arbitrarily controlled."

References

Aoki M (2001) *Toward a Comparative Institutional Analysis* (MIT Press, Cambridge, MA).

Argyris C (1985) *Strategy, Change and Defensive Routines* (Pitman, Boston).

Barney JB (1986) Organizational culture: Can it be a source of sustained competitive advantage? *Acad. Management Rev.* 11(3): 656–665.

Barr PS, Stimpert JL, Huff AS (1992) Cognitive change, strategic action, and organizational renewal. *Strategic Management J.* 13(S1):15–36.

Bednar J, Page S (2007) Can game(s) theory explain culture?: The emergence of cultural behavior within multiple games. *Rationality Soc.* 19(1):65–97.

Bender A, Hutchins E, Medin D (2010) Anthropology in cognitive science. *Topics Cognitive Sci.* 2(3):374–385.

Boje DM (1991) The storytelling organization: A study of story performance in an office-supply firm. *Admin. Sci. Quart.* 36(1): 106–126.

Boyd R, Richerson PJ (2009) Culture and the evolution of human cooperation. *Philos. Trans. Roy. Soc. London B Biol. Sci.* 364(1533): 3281–3288.

Brandenburger A, Nalebuff BJ (1996) *Co-opetition* (Currency Doubleday, New York).

Burke WW (2017) *Organization Change: Theory and Practice*, 5th ed. (Sage, Thousand Oaks, CA).

Chatman JA, O'Reilly CA (2016) Paradigm lost: Reinvigorating the study of organizational culture. *Res. Organ. Behav.* 36: 199–224.

Cole RE (1991) *Strategies for Learning: Small-Group Activities in American, Japanese, and Swedish Industry* (University of California Press, Berkeley).

D'Andrade RG (1995) *The Development of Cognitive Anthropology* (Cambridge University Press, Cambridge, UK).

Denzau A, North D (1994) Shared mental models: Ideologies and institutions. *Kyklos* 47(1):3–31.

DiMaggio P (1997) Culture and cognition. *Annual Rev. Sociol.* 23: 263–287.

Ellingsen T, Johannesson M, Mollerstromm J, Munkhammar S (2012) Social framing effects: Preferences or beliefs? *Games Econom. Behav.* 76(1):117–130.

Fiss PC, Zajac EJ (2006) The symbolic management of strategic change: Sensegiving via framing and decoupling. *Acad. Management J.* 49(6):1173–1193.

Fullan M (2001) *Leading in a Culture of Change* (Jossey-Bass, San Francisco).

Gavetti G (2012) Toward a behavioral theory of strategy. *Organ. Sci.* 23(1):267–285.

Geertz C (1973) Thick description: Toward an interpretive theory of culture. *The Interpretation of Cultures: Selected Essays* (Basic Books, New York), 3–30.

Gersick CJ (1991) Revolutionary change theories: A multilevel exploration of the punctuated equilibrium paradigm. *Acad. Management Rev.* 16(1):10–36.

Gerstner LV Jr (2002) *Who Says Elephants Can't Dance? Leading a Great Enterprise Through Dramatic Change* (HarperCollins, New York).

Gibbons R, Henderson R (2013) What do managers do?: Exploring persistent performance differences among seemingly similar enterprises. Gibbons R, Roberts J, eds. *The Handbook of Organizational Economics.* (Princeton University Press, Princeton, NJ),680–731.

Goffman E (1974) *Frame Analysis: An Essay on the Organization of Experience* (Harvard University Press, Cambridge, MA).

Goodstein LD, Burke WW (1991) Creating successful organization change. *Organ. Dynam.* 19(4):5–17.

Greenwood R, Hinings CR (1996) Understanding radical organizational change: Bringing together the old and the new institutionalism. *Acad. Management Rev.* 21(4):1022–1054.

Hannan MT, Le Mens G, Hsu G, Kovács B, Negro G, Pólos L, Pontikes E, Sharkey AJ (2019) *Concepts and Categories: Foundations for Sociological and Cultural Analysis* (Columbia University Press, New York).

Hedberg B (1981) How organizations learn and unlearn. Nystrom PC, Starbuck WH, eds. *Handbook of Organizational Design* (Oxford University Press, Oxford, UK), 1–27.

Heller Y, Winter E (2016) Rule rationality. *Internat. Econom. Rev.* 57:997–1026.

Henrich J, Boyd R, Bowles S, Camerer C, Fehr E, Gintis H, McElreath R, et al. (2005) "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. *Behav. Brain Sci.* 28(6):795–815.

Hermalin BE (1998) Toward an economic theory of leadership: Leading by example. *Amer. Econom. Rev.* 88(5):1188–1206.

Hoff K, Stiglitz JE (2016) Striving for balance in economics: Toward a theory of the social determination of behavior. *J. Econom. Behav. Organ.* 126(Part B):25–57.

Hong L, Page S (2009) Interpreted and generated signals. *J. Econom. Theory* 144(5):2174–2196.

Jehiel P (2005) Analogy-based expectation equilibrium. *J. Econom. Theory* 123(2):81–104.

Kaplan S (2008) Framing contests: Strategy making under uncertainty. *Organ. Sci.* 19(5):729–752.

Keller J, Loewenstein J (2011) The cultural category of cooperation: A cultural consensus model analysis for China and the United States. *Organ. Sci.* 22(2):299–319.

Kellogg KC (2011) Hot lights and cold steel: Cultural and political toolkits for practice change in surgery. *Organ. Sci.* 22(2):482–502.

Kogut B, Kulatilaka N (2001) Capabilities as real options. *Organ. Sci.* 12(6):744–758.

Kotter JP, Heskett JL (1992) *Corporate Culture and Performance* (Free Press, New York).

Kreps DM (1990) Corporate culture and economic theory. Alt JE, Shepsle KA, eds. *Perspectives on Positive Political Economy* (Cambridge University Press, Cambridge, UK), 90–143.

Kroeber AL, Kluckhohn C (1952) *Culture: A Critical Review of Concepts and Definitions*, Papers of the Peabody Museum of American Archaeology and Ethnology (Peabody Museum, Cambridge, MA).

Leibenstein H (1982) The prisoners' dilemma in the invisible hand: An analysis of intrafirm productivity. *Amer. Econom. Rev.* 72(2): 92–97.

Levinthal DA (1997) Adaptation on rugged landscapes. *Management Sci.* 43(7):934–950.

Levitt B, March JG (1988) Organizational learning. *Annual Rev. Sociol.* 14:319–338.

Liberman V, Samuels SM, Ross L (2004) The name of the game: Predictive power of reputations vs. situational labels in determining prisoner's dilemma game moves. *Personality Soc. Psych. Bull.* 30(9):1175–1185.

Malcomson JS (2013) Relational incentive contracts. Gibbons R, Roberts J, eds. *The Handbook of Organizational Economics* (Princeton University Press, Princeton, NJ), 1014–1065.

March JG (1981) Footnotes to organizational change. *Admin. Sci. Quart.* 26(4):563–577.

March JG, Olsen JS (1983) The new institutionalism: Organizational factors in political life. *Amer. Political Sci. Rev.* 78(3):734–749.

Medin DL, Schaffer MM (1978) Context theory of classification learning. *Psych. Rev.* 85(3):207–238.

Mengel F (2012) Learning across games. *Games Econom. Behav.* 74(2): 601–619.

Milgrom P, Roberts J (1995) Complementarities and fit strategy, structure, and organizational change in manufacturing. *J. Account. Econom.* 19(2–3):179–208.

Nosofsky RM (1986) Attention, similarity, and the identification-categorization relationship. *J. Experiment. Psych. General* 115(1): 39–57.

Osherson DN, Smith EE (1981) On the adequacy of prototype theory as a theory of concepts. *Cognition* 9(1):35–58.

Ostrom E (1990) *Governing the Commons: The Evolution of Institutions for Collective Action* (Cambridge University Press, New York).

Patterson O (2014) Making sense of culture. *Annual Rev. Sociol.* 40:1–30.

Pettigrew AM (1979) On studying organizational cultures. *Admin. Sci. Quart.* 24(4):570–581.

Pothos EM, Wills AJ, eds. (2011) Formal Approaches in Categorization (Cambridge University Press, Cambridge, UK).

Pruitt DG (1970) Motivational processes in the decomposed prisoner's dilemma game. *J. Personality Soc. Psych.* 14(3):227–238.

Repenning NP, Sterman JD (2002) Capability traps and self-confirming attribution errors in the dynamics of process improvement. *Admin. Sci. Quart.* 47(2):265–295.

Roberts J (2004) *The Modern Firm: Organizational Design for Performance and Growth* (Oxford University Press, Oxford, UK).

Romanelli E, Tushman ML (1994) Organizational transformation as punctuated equilibrium: An empirical test. *Acad. Management J.* 37(5):1141–1166.

Rosch EH (1973) Natural categories. *Cognitive Psych.* 4(3):328–350.

Samuelson L (2001) Analogies, adaptation, and anomalies. *J. Econom. Theory* 97(2):320–366.

Schein EH (2010) *Organizational Culture and Leadership*, 4th ed. (Jossey-Bass, San Francisco). [Orig. pub. 1985.]

Schelling TC (1960) *The Strategy of Conflict* (Harvard University Press, Cambridge, MA).

Selznick P (1957) *Leadership in Administration: A Sociological Interpretation* (Harper & Row, New York).

Sørensen JB (2002) The strength of corporate culture and the reliability of firm performance. *Admin. Sci. Quart.* 47(1):70–91.

Sperber D (1996) *Explaining Culture: A Naturalistic Approach* (Blackwell, Oxford, UK).

Tushman ML, Romanelli E (1985) Organizational evolution: A metamorphosis model of convergence and reorientation. *Res. Organ. Behav.* 7:171–222.

Weber RA, Camerer CF (2003) Cultural conflict and merger failure: An experimental approach. *Management Sci.* 49(4):400–415.

Weick KE, Quinn RE (1999) Organizational change and development. *Annual Rev. Psych.* 50:361–386.

Werner MD, Cornelissen JP (2014) Framing the change: Switching and blending frames and their role in instigating institutional change. *Organ. Stud.* 35(10):1449–1472.

Zerubavel E (1991) *The Fine Line: Making Distinctions in Everyday Life* (Free Press, New York).

Zerubavel E (1997) *Social Mindscapes: An Invitation to Cognitive Sociology* (Harvard University Press, Cambridge, MA).

Robert Gibbons is the Sloan Distinguished Professor of Management, Sloan School of Management, and a professor of organizational economics, Department of Economics, Massachusetts Institute of Technology. His main research and teaching interests are in organizational economics. He received a PhD in decision sciences from the Graduate School of Business of Stanford University.

Marco LiCalzi is a professor of mathematical methods for economics at the Department of Management, Università Ca' Foscari Venezia. His main research and teaching interests are in game theory. He received a PhD in decision sciences from the Graduate School of Business of Stanford University.

Massimo Warglien is a professor of management at the Department of Management, Università Ca' Foscari Venezia. His research and teaching are in the areas of organization theory, language and interaction, behavioral decision making, and cross-cultural and multispecies studies.