

# Birds of a Feather . . . Enforce Social Norms? Interactions Among Culture, Norms, and Strategy

Hongyi Li  
Eric Van den Steen

Working Paper 20-045



# Birds of a Feather . . . Enforce Social Norms? Interactions Among Culture, Norms, and Strategy

Hongyi Li  
University of New South Wales

Eric Van den Steen  
Harvard Business School

**Working Paper 20-045**

Copyright © 2019 by Hongyi Li and Eric Van den Steen

Working papers are in draft form. This working paper is distributed for purposes of comment and discussion only. It may not be reproduced without permission of the copyright holder. Copies of working papers are available from the author.

Funding for this research was provided in part by Harvard Business School.

# Birds of a Feather . . . Enforce Social Norms? Interactions among Culture, Norms, and Strategy

Hongyi Li and Eric Van den Steen \*

October 1, 2019

## Abstract

This paper analyzes how shared beliefs and preferences (or values) cause the emergence of social norms; why people may enforce norms that go *against* their own beliefs and preferences/values; and how this may cause a disconnect to develop between the organization's norms and its underlying beliefs and preferences. We further show, among other things, that such social norms are more likely in attractive organizations, for behaviors that have modest personal consequences, and in organizations where employees depend on others' choices to a moderate degree. We finally discuss how these mechanisms help our understanding of culture change and of the interaction between culture and strategy. We argue that culture is not only an input to strategy, but also a substitute and a potential competitor.

## 1 Introduction

Both norms and 'shared beliefs and preferences' (or values) are important facets of corporate culture: shared beliefs and preferences/values are at the heart of most definitions of corporate culture (Schein 1985) whereas norms help make culture into a control mechanism (O'Reilly and Chatman 1996). Understanding the connections between these two - especially how one may lead to another - is vital not only to understand the full effects of culture and to determine how culture is best defined but also, it turns out, to understand the relationship between culture and strategy. The purpose of this paper is to carefully study these links and their implications by developing micro-foundations for culture, norms, and their interactions with each other and with strategy.

At the heart of this paper is the result that shared beliefs and preferences cause the emergence of social norms and may cause people to enforce norms *even if they go against* their own beliefs and preferences. This result, which has a wide range of implications, turns out to be a fairly direct consequence of the combination of the strong forces towards shared beliefs and preferences (Schein 1985, Chatman 1991, Van den Steen 2010b) with the fact that beliefs and preferences are often unobservable and must therefore be inferred from people's behavior. This combination generates a pressure for people to act *as if* they hold the 'right' beliefs and preferences, in order to fit in. This results in what is essentially a social norm - with people taking 'appropriate' actions that conflict with their own beliefs and preferences in order to avoid social sanctions. Moreover,

---

\*UNSW Business School (hongyi@hongyi.li) and Harvard Business School (evandensteen@hbs.edu). This work benefited from many discussions with, and inspiration from, Robert Akerlof, Ramon Casadesus-Masanell, John Roberts, and especially Bob Gibbons. We thank Catherine Barrera for research assistance on an earlier incarnation of this project. Jenny Chatman and Sameer Srivastava gave some instrumental encouragement.

when enforcement is observable, people may enforce norms that go *against* their own beliefs and preferences – again in order to act as if they have the right beliefs. In this theory, social norm enforcement and sanctions are thus *not* (just) about the actions themselves but often more about what the norm violation revealed about the person’s beliefs or preferences. By following and enforcing norms, people show that they are of the ‘right’ type – or at least not (too much) of the wrong type. We develop this argument in qualitative terms but also illustrate it with a very simple model of a setting where norms develop and where people will end up enforcing norms that they don’t agree with.<sup>1</sup>

We then argue that this emergence of norms from shared beliefs and preferences has important implications. First of all, it implies that norms act as a lever on shared beliefs and preferences: norms make it look ‘as if’ more people share the beliefs/preferences than actually do. This ‘leverage’ effect not only affects the costs and benefits of shared beliefs/preferences; it also raises empirical issues for how to measure culture; it provides an explanation for how cultures can sometimes change quite suddenly; and it clarifies how people can move between seemingly conflicting cultures. This furthermore resolves an apparent paradox in the ‘culture as control view’ between the premise that norms come from shared values and the fact that norms only control when values are not shared. It shows, in particular, that for norms to act as a means of control, the underlying beliefs and preferences should be *partially* shared – not fully shared. Second, and of particular importance in this context, norms also strengthen the role of culture as a control mechanism – with important implications for strategy, to which we return below. Third, in a reverse-direction effect, norms may also induce selection and thus create more shared beliefs and preferences. Finally, we also explore how this mechanism may cause a disconnect to develop between an organization’s norms and its underlying beliefs and values, with important implications for culture and culture change.

Apart from these implications for the nature and functioning of culture, the theory also has implications for norms. We show, for example, that norms are more likely in attractive organizations and along dimensions of behavior where norm adherence is less costly. We also show that norms are more likely in settings where people depend on others’ choices to an intermediate degree. This last result suggests that the task environment – in particular how much people depend on others’ choices – can be used to affect the emergence or extinction of norms.<sup>2</sup>

The paper finally discusses what the theory implies for culture change, for definitions of culture, and especially for the interaction between culture and (business) strategy. In terms of the definition of culture, the paper discusses the explicit inclusion of ‘norms’ in the definition – for which the paper finds arguments in both directions – and the role of ‘*perceived* shared beliefs and preferences.’ For the relationship between culture and strategy, the emergence of norms has the fundamental implication that culture is not just an input to strategy, as it is traditionally considered, but also a potential ‘*competitor*’ – given that it is a direct substitute for strategy’s core function as a guide for decisions.<sup>3</sup> This may explain why ‘culture versus strategy’ is so high on the agenda for managers.

As this paper sits squarely at the intersection of a number of very different fields, we start the paper with a very brief overview of earlier work (in economics) on ‘culture as shared beliefs and preferences.’ As part of this introductory overview, we also discuss how different definitions of culture relate, which raises questions on the exact meaning of ‘values’ and whether culture

---

<sup>1</sup>Note that this is intended as a semi-formal paper. Whereas many of the arguments in the paper are illustrated with the model and with formally proved results, the analysis does not provide a full formal proof of all arguments. Such completely formal approach requires a different type of paper and goes beyond the current purpose.

<sup>2</sup>The paper does not claim that all norms are of this type. Only that this is a logical mechanism for norms to emerge and that such norms should obey these comparative statics.

<sup>3</sup>The focus is here on business strategy or organization strategy, not on global or corporate strategy.

definitions do in fact specify necessary and sufficient conditions (as they should). At the end of the paper, we return to the definition of culture and what this paper may have added.

## 1.1 Literature

This paper is connected to three quite different literatures. The literature closest to this paper is obviously the work in organizational economics on ‘culture as shared beliefs and preferences’, such as Van den Steen (2010a,b) and Li (2016) and its antecedents, such as Crémer (1993) and Lazear (1995) – and ultimately, obviously, Schein (1985). Because this work will be discussed in more detail in Section 2, we will discuss here some more distant connections. In particular, it is important to observe that there are in fact two quite different approaches to corporate culture in the economics literature. Apart from the view of ‘culture as shared beliefs and preferences’ that this paper uses, there is also an approach – starting with Kreps (1990) and including work such as Carrillo and Gromb (1999), Weber and Camerer (2003), and Gibbons and Henderson (2012) – that essentially defines culture as equilibrium selection: in the presence of multiple equilibria – such as driving on the right hand side versus on the left hand side – culture is about creating a common expectation as to which equilibrium will be played, which then makes it optimal for everyone to adhere to that equilibrium. Note that this is also about ‘shared beliefs’ but now of a very specific kind: shared beliefs about the equilibrium that is being played. This relates closely to the idea of ‘norms as conventions’ which is quite different from ‘social norms’: norms as conventions exist only if there are multiple equilibria and are self-enforcing (as driving on the wrong side of the road is a bad idea), whereas ‘social norms’ are independent of the multiplicity of equilibria and typically rely on sanctions (possibly internalized). Whereas ‘culture as conventions’ is a very interesting perspective, it is somewhat disconnected from the perspective in this paper (although Van den Steen (2010a) conjectured that shared beliefs and preferences play a role in equilibrium selection).<sup>4</sup>

A second literature this paper connects to is obviously the organizational behavior literature on culture and norms. The very definition of culture as ‘shared beliefs and preferences’ was intended to capture Schein’s (1985) definition. Moreover, the early work in the ‘shared beliefs and preferences’ literature formalized some of Schein’s and others’ ideas before going beyond it. Schein (1985), however, does not include norms in his definition of culture and barely touches it in his seminal work. The current paper grew in part out of an effort to try to reconcile O’Reilly and Chatman’s (1996) important ideas of ‘culture as control’ with ‘culture as shared beliefs and preferences,’ especially in the context of strategy (where this is critical). Whereas O’Reilly and Chatman (1996) take values and social norms essentially as given, see norms as the direct translation of values into behavior, and see ‘the distinction between them [as] one of emphasis’, this paper instead studies in detail the connection between beliefs/preferences on the one hand and norms on the other – in particular, how one emerges from the other. We believe that this paper is the first to study culture from this particular angle. Interestingly, Chatman and O’Reilly’s (2016) three dimensions of culture – strength, consensus, and content – emerge very naturally and very clearly from the theory and model in this paper, suggesting that this link may be relevant to their theory.

---

<sup>4</sup>There is also a literature that is orthogonal to this. In particular, Guiso, Sapienza, and Zingales (2006) study the effect of *personal* values, as determined by your *national culture*. Their definition of culture, as a consequence, does not include any reference to the ‘*sharedness*,’ which is *the* key discriminant of the corporate culture literature. Similarly, Guiso, Sapienza, and Zingales (2015) study in fact the effect of how employees perceive top management – in particular, whether they believe that management keeps its word or that it is honest and ethical – independent of whether this is related to some shared values. This is both a different use and interpretation of ‘culture’ and a very different focus in terms of research.

Within the wider sociology and psychology literature on social norms, the work that comes closest to this paper is probably the ‘social identity’ theory of social norms such as Hogg and Turner (1987). This work, however, only relates to a proposed secondary variation of the model – the variation on ‘need for identity’ – rather than to the paper’s main theory. Relative to that literature, the secondary variation on this paper’s model would be more detailed and more formal about the exact mechanisms. Other perspectives on norms focus on efficiency (Coleman 1990) or on socialization (Parsons 1951) and are therefore far removed from this work.

Finally, this paper is obviously also connected to the strategy literature on the connection between culture and (business) strategy.<sup>5</sup> The academic literature on this topic is surprisingly small.<sup>6</sup> Most of this literature, exemplified by Barney (1986), has approached culture as an input to strategy, like a resource or an asset, and then focused on the question how culture can create a competitive advantage. We will discuss that perspective further in Section 5 and explain how this paper suggests a very different perspective. Culture also arises in more practitioner-oriented work (for example, Groysberg et al. 2018 or Levine 2019) and textbook chapters on business strategy implementation (for example, Grant 2016 and Saloner et al. 2001) but these discussions do not go into detail on why exactly culture is such a challenge for strategy, resulting in a fairly general warning to consider culture when developing strategy or to make sure that the strategy and culture fit. The purpose of this paper, on the other hand, is to help develop a richer perspective based on a more detailed understanding of the mechanisms underlying the interaction between culture and strategy.

Overall, the contribution of this paper is to provide a detailed and formal perspective on how shared beliefs and preferences leads to norms and how that furthers our understanding of the interaction between culture and (business) strategy and related questions.

## 2 Culture as Shared Beliefs and Preferences

### 2.1 Short Overview

This paper starts from the definition of culture as ‘shared beliefs and preferences’ (Van den Steen 2010a). This simple and transparent definition (which encompasses shared values) turns out to be quite powerful in terms of implications and predictions. Whereas this definition was originally proposed to capture Schein’s definition in formal-economic terms, it also differs from Schein’s, as we will discuss.<sup>7</sup>

One important advantage of such a simple definition is that it makes it easy to analyze and understand its nature, properties, and implications.<sup>8</sup> Van den Steen (2010a) used a formal model to show how such ‘culture as shared beliefs and preferences’ leads to more delegation, less monitoring, higher satisfaction, higher motivation, faster coordination, less influence activities, and more communication, but also to less experimentation and less information collection. The reason why culture has such a wide range of performance effects is because shared beliefs and preferences resolve agency problems (at the source), given that agency problems find their origin in conflicting beliefs, values, and objectives. Van den Steen (2010b) then studied the *origins* of such culture and

---

<sup>5</sup>The global strategy literature also has work on culture as an environmental factor, opportunity, or constraint, such as Ghemawat (2001) or Siegel, Licht, and Schwartz (2013), but this is quite removed from this paper.

<sup>6</sup>The 128-page strategy reading list for PhD students compiled by a group of faculty in the ‘Strategy Research Initiative’ mentioned the word *culture* just once, as a potential asset (SRI 2009).

<sup>7</sup>The definition built on earlier work by Crémer (1993) and Lazear (1995), as discussed in Van den Steen (2010a).

<sup>8</sup>It also makes it easy to explain culture to people who are new to it.

formally showed that ‘shared beliefs and preferences’ have many of the key properties attributed to culture by Schein (1985): they develop through screening, self-sorting, and joint learning; are particularly influenced by the CEO’s beliefs and preferences; and persist despite turnover. The paper also made some new predictions, e.g., that such culture will be stronger among important employees, in older and more successful firms, and when employees make important decisions (as in professional firms). Li (2016) showed formally how fast growth may lead to fragmentation of such culture, that the fragmentation may persist even after growth ends, and that it is particularly likely with growth driven by acquisitions.<sup>9</sup> Epstein and Lindner-Pomerantz (2017) investigate under what conditions outlier-cultures can survive, whereas Prasad and Tomaino (2019) show how the strength of culture interacts with the availability of resources to determine economic outcomes. On the empirical side, Campbell (2012) and Abernethy, Dekker, and Schulz (2015) provide very interesting evidence on the effects of the employee selection aspect of such culture and how that determines outcomes.

## 2.2 Discussion of the Definition

Overall, the definition turns out to be quite powerful: despite being very simple, it generates lots of implications – both previously associated with culture and new. But the definition was originally criticized for being different from, e.g., Schein’s. These very differences, however, have also proved to deliver considerable benefits. Discussing some of these differences therefore gives useful insight into the conception of culture.

A first important difference is that proposed definitions such as Schein’s (1985) sometimes include ingredients like the *origin* of the beliefs/values/preferences. Schein (1985), for example, includes in his famous proposed definition that the shared assumptions must have been ‘learned [by the group] as it solved problems’ and ‘taught to new members.’ But a true definition gives necessary and sufficient conditions: any ‘culture’ must satisfy the definition (‘necessary’) and anything that satisfies the definition must be a ‘culture’ (‘sufficient’). Schein’s proposed definition would thus imply that shared assumptions that became shared by screening or self-selection would *not* constitute a culture, which contradicts most of the culture literature and Schein’s own work. So it seems that Schein (1985) provided essentially a *description* of a very common and appealing case rather than an actual definition, as is common in the management literature. (Van den Steen (2018) showed, for example, that the most common proposed definitions of ‘strategy’ fail to give ‘necessary and sufficient’ conditions.) That issue does not, of course, diminish the seminal importance of Schein’s work, but it is important for evaluating definitions.<sup>10</sup>

A second difference between this definition and many definitions in the culture literature is that we define culture here in terms of ‘preferences’ instead of ‘values,’ which differ in ways that go to the heart of culture.<sup>11</sup> Preferences are just likes, dislikes, and attitudes, with varying emotions. Values, on the other hand – used in the plural – are preferences that have an additional moral dimension and a general sense of desirability. The Collins Dictionary, for example, defines ‘values’

---

<sup>9</sup>Whereas the paper formulates the model and results in terms of shared knowledge, it can also be interpreted – and was originally intended – as being about shared beliefs or preferences.

<sup>10</sup>A careful evaluation of ‘necessary and sufficient’ also often highlights needs for clarification. For example, a definition that says that culture consists of ‘shared values and shared norms’ must be clear on whether you need *both* values *and* norms or whether one or the other is sufficient; whether the norms are allowed to contradict the values; who evaluates whether something is a value or not; etc.

<sup>11</sup>Note that the stated definition in Van den Steen (2010a) is also in terms of ‘values’ but that paper used ‘values’ as a synonym for preferences, simply because it was more in line with the terminology of the management literature.

(plural) as ‘moral principles and beliefs’ whereas Parks and Guay (2009) – cited by Chatman and O’Reilly (2016) – define it as ‘conceptions of the desirable.’

So is culture about morally desirable ‘values’ or may it also encompass more general and more mundane ‘beliefs and preferences’? The current paper favors ‘preferences’ (or an explicit clarification that ‘values’ can also include more mundane preferences) for two reasons.

First, only considering ‘values’ excludes some mundane but performance-relevant ‘shared beliefs and preferences’, such as a tolerance for chaos versus a desire for structure or a preference for teams versus working alone. Second, and closely related, restricting culture to *desirable* ‘values’ risks distancing it from some of the most difficult management problems – caused by negative cultures like those characterized by complacency, favoritism, or much worse.<sup>12</sup>

Of course, broadening culture to shared preferences may raise some questions. Can we have a culture of ‘liking blue,’ for example? This, however, should be decided by theory and empirics rather than restricted by the definition. It points in fact to an important understudied research question: along which dimensions are shared beliefs and preferences most likely? The current paper suggests, for example, that culture is more likely along dimensions that end up affecting others – thus more likely along ‘liking structure’ than along ‘liking blue’. Another concern could be that ‘shared beliefs and preferences’ is too broad and hurts the predictive power of the theory, but the existing research results clearly suggest otherwise.

Overall, this paper subscribes to the view that clear definitions – with unambiguous necessary and sufficient conditions – are the foundation of good research. And if a field, for some reason, (temporarily or permanently) entertains multiple definitions, then it is imperative that each paper is clear on what definition it uses and that connections are actively investigated. We therefore hope that this paper contributes to that evolving discussion on the definition of culture. We return to that issue later.

### 3 From Shared Beliefs to Norms

The core idea of this paper is that culture as ‘shared beliefs and preferences’ gives rise to social ‘norms’ with people enforcing norms even if they go against their personal beliefs and preferences. The key animating force in this mechanism is a *‘pressure for perceived similarity,’* i.e., a pressure to *look* similar to others in terms of beliefs and preferences – of which we will immediately discuss the sources. As similarity in beliefs and preferences is not directly observable, it must be inferred from the person’s actions and behavior, thus creating a pressure for people to *act* similar to others. This generates what are in effect social norms: a socially desired behavior, deviations from which are sanctioned by whatever form the pressure takes. These sanctions can be some form of punishment or can simply entail withholding benefits that are provided to those with the right beliefs and preferences.

But there is more: this mechanism also drives people to enforce norms, even norms that go against their own beliefs and preferences/values. The reason is that people who really hold the shared beliefs and preferences will enforce the norm by applying the sanctions. But since these sanctions are often observable, engaging in these enforcement actions becomes part of the observable behavior of someone with the ‘right’ beliefs and preferences, to which the pressure for

---

<sup>12</sup>Defining culture as having some degree of desirability also risks making important performance aspects of culture tautological. What would it mean to show that culture is valuable if culture is restricted to be about desirable values? Moreover, this very discussion suggests that the term ‘value’ has some ambiguity.



perceived similarity also applies. This creates a pressure to punish or sanction deviators. And *norm enforcement itself becomes part of the norm.*

So let us now return to the ‘*pressure for perceived similarity.*’ Where does this come from? There are a number of related mechanisms. We will discuss here in detail one mechanism that is particularly important for organizations and then mention more concisely two others. This first mechanism results from two forces that reinforce each other. The first force comes from the fact that in organizations, people often depend on other people’s decisions and thus want these other people to make the ‘right’ decisions. In the face of differences in beliefs, values, and objectives, making the ‘right’ decisions often means making the decisions that the focal actor would have made. The focal actor therefore wants these other people to have the same beliefs and preferences. For example, if someone believes that technical excellence is critical to success, that person will not only deeply care about technical excellence but will also like others who care about technical excellence as they help to achieve success. Such person will therefore hire others with similar beliefs, promote them, and give them resources. Not only does this lead directly to shared beliefs and preferences but it also means that people who conform are more likely to be hired, to be promoted, and to get resources to work with, thus also strengthening their position and role. An academic example is that someone who cares a lot about research typically wants colleagues who also care about research so that these colleagues put appropriate weight on research when hiring or allocating funds. We will use the term ‘functional homophily’ to refer to this preference for similar others (‘homophily’) because it affects you (‘functional’). While we will discuss below other mechanisms that make people prefer others with similar beliefs and preferences, functional homophily is a very common and strong force in organizations. Moreover, it goes beyond direct personal interactions and applies to any other person in the organization who makes decisions that affect you. But because beliefs and preferences are invisible, you can’t just select a person who is similar; instead, you try to select a person who *appears similar*. This results in a pressure for perceived similarity (to those who are in control).

The second force, which works as a lever on the first, is the fact that when people have *shared* beliefs and preferences, their attempts to influence beliefs and preferences all go in the same direction and reinforce each other. Consider, for example, the effect of functional homophily in an organization where the first two members need to select a third. If these two existing members are opposites (along the relevant dimensions) then their efforts to hire similar others will go in opposite directions and cancel each other out. If, on the other hand, the two existing members are similar along the relevant dimensions, then they agree on hiring a similar third and homogeneity begets more homogeneity. As such, homogeneity in the sense of shared beliefs and preferences acts as a *lever* on forces that lead to the selection of similar people. This second force also makes things like ‘critical mass’ important and leads to predictions about the dynamics of culture based on the current degree of sharedness.

### 3.1 Other Mechanisms

There are two other mechanisms that lead to a pressure for perceived similarity. All these mechanisms have the second force – how sharedness of beliefs and preferences acts as a lever – in common. They all differ in the first force, which was ‘functional homophily’ in the first mechanism. A first alternative mechanism to ‘functional homophily’ is obviously regular homophily, the well-known and well-documented effect that people simply *like* people who are similar to themselves and build relations with them (Huston and Levinger 1978, McPherson, Smith-Lovin, and Cook 2001). In the current context, the key implication of homophily is that if one actor wants to be liked by another,

it helps if that actor is perceived to be similar to them, i.e., it results in a ‘*pressure for perceived similarity*’. And while there is little you can do about that on visible characteristics, the situation is very different for beliefs and preferences, as before.

A second alternative mechanism is driven by people’s need for identity, in the sense of a person’s beliefs about themselves. Such identity is derived in part from group membership. I may believe, for example, that I’m a tough person because I belong to a gang that has a reputation for being tough. But the need for identity makes the members of the group care about the group’s reputation and therefore about the other group members’ actions. If someone in the supposedly-tough group cries, then that may damage the group’s image as being tough and thus hurt the members’ identity. Members are therefore pressured, through sanctions and other means, to conform to the group’s desired image, thus creating a pressure for perceived similarity to the group’s desired image. But as each group member conforms, the result is effectively a pressure for perceived similarity among them. Note, however, that a group’s desired image may differ from its true nature. Whereas this causes a misalignment between the group’s norms and its underlying shared beliefs and preferences, it still results in norms. We will discuss another – more important – source of misalignment below.

Notice that these different mechanisms can have very different comparative statics, which makes it possible to test and discriminate among them. Identity, for example, suggests stronger norms when group membership is more salient. Functional homophily is stronger when people are affected more by others’ decisions.

### **3.2 A Preference for Differences?**

There are obviously also cases where people may prefer to work with others who are *different* from themselves. People may simply prefer diversity, for example, or see a need for trying different things. This observation does not contradict the theory. Instead, it produces a few useful insights on this theory of culture and norms.

First, if there are dimensions along which people prefer to work with others who are *different* then these dimensions will simply not become part of the culture or of the norms. In as far as it is predictable along what dimensions that will be, this can give us some predictions about the dimensions along which we are more (or less) likely to see culture and norms. For example, we might hypothesize that people will generally like agreement on work practices (order versus structure, team versus individual, open fights versus consensus, ...) but that certain people will like diversity in backgrounds or experience. Second, a preference for diversity along some dimension often causes a preference for similarity along a related dimension. For example, if I like diversity of backgrounds, then I typically want to work with people who have a similar like for diversity of backgrounds. This would result in a culture with a shared preference for diversity. If I think experimentation is important, then I may want to work with someone who has different beliefs on the most likely solution but similar beliefs on the importance of experimentation. Third, some preferences for differences are not ‘stable’ in a group. If I’m lazy, I may prefer to work with someone who is not lazy, so that that other person does all the work. But that other person will likely *not* prefer to work with me because I’m lazy; instead they also want to work with someone who is not lazy. Overall, a preference for differences does not contradict the theory, but can instead be helpful to make predictions about the types of homogeneity and norms that are more likely.

### 3.3 Some Comparative Statics

Whereas Section 4 will discuss the implications of the main result – that shared beliefs and preferences lead to norms – there are some other potential insights from this theory that are worth mentioning. (Note that, whereas the results of Section 6 clearly support these insights, the illustrative nature of the model means that they do not fully prove them. The results thus come in part from qualitative reasoning.)

The following insights are illustrated explicitly by the results in Section 6:

- Norms are more likely in attractive organizations (from which exclusion is thus costly) or when strong sanctions can be credibly applied.
- Norms are more likely when the normative behavior is less costly to the participant. We would thus expect to see a lot of relatively ‘cheap’ norms (such as dress codes, behavior codes, word usage, etc.) and much fewer ‘expensive’ norms (such as houses to buy or type of partner to have).
- Norms are more likely in settings where participants’ outcomes depend to a moderate degree on others’ decisions and behavior.

This last point deserves some more explanation. Our original intuition when starting this project was that norms only emerge when there is sufficient interdependence: without dependence on others’ behavior there is no reason to enforce norms. But the analysis of Section 6 showed, at first to our surprise, that there can also be too much interdependence to sustain norms! The reason is that, when interdependence is very strong, the organization is not so attractive to people who disagree with its shared beliefs or preferences. Unafraid of exclusion, such participants see no need to adhere to norms, thus making the norms fail. But in this case the failure of norms will lead to increased turnover – as norm violators will be excluded – ultimately resulting in an organization with fully shared beliefs and preferences. This comparative static on interdependence is of particular interest as it suggests an interesting path to potentially affect the emergence of norms: forcing or loosening cooperation may affect the emergence (or extinction) of social norms. The result also implies that norms are more likely in settings where people interact a reasonable amount (so they are affected more by others’ behavior). This potentially explains why we see norms in organized settings more than in unstructured groups.

By informally analyzing variations on a model in the style of Section 6, we conjecture the following insights:

- Norms cause self-selection: people with strong ‘wrong’ beliefs/preferences will avoid joining the organization in the first place because the existence of norms imposes costs on them. First, norms force such people into a choice between two ‘evils’: either obey the norm but do something that goes against your beliefs/preferences or violate the norm and face sanctions. Second, norms also make the behavior that violates their beliefs/preferences ubiquitous, further reducing the attractiveness of joining. If joining requires any effort or expense, such people are less likely to incur that effort/expense and thus less likely to join.
- Norms are more likely for behavior that is socially desirable or at least acceptable, i.e., norms are more likely for values than for vices.
- Norms are more likely in settings where there is a common future and where that future matters more.

- Vulnerable members of the organization – those who are easy to sanction or who bear strong consequences from the sanctions – are more likely to adhere to norms. Invulnerable members are less likely to adhere.
- Members whose beliefs and preferences are common knowledge or beyond doubt are less likely to adhere to norms (that go against their preferences or beliefs), as they don't need to prove their type.
- A norm requires a critical mass of participants who are perceived to hold the norm-supporting beliefs and preferences.<sup>13</sup>

The results of Srivastava, Goldberg, Manian, and Potts (2018) are particularly interesting in this context – if we take a somewhat broad view of their interpretation – because they are in part consistent with the existing theory and in part suggest new results that seem clearly consistent with it. Their paper studies how enculturation determines success by looking at how people's email style comes to resemble that of their colleagues. If we interpret adhering to an email style as adhering to a norm, then we can see how the results compare to the theory. Fast adoption of norms – which may measure how close norms are to personal beliefs and preferences – predicts high retention, which would thus be consistent. Not adopting norms predicts getting 'eliminated' from the organization. And a decrease of adherence to norms after having adopted norms previously predicts voluntary exit, which seems to capture the effect that some feel that adherence to norms is too costly and leave or that their desire to be in the organization has decreased for other reasons and now they don't feel a need to adhere to norms any more.

### 3.4 Misalignment Between Norms and Underlying Beliefs/Preferences

Occasionally, the mechanisms discussed above can lead to a misalignment between the norms of an organization and its underlying beliefs and preferences. In fact, we will immediately explain why it is (in principle) possible that the organization enforces norms that go directly against *everyone's* shared beliefs and preferences – and even when everyone is aware of the misalignment. Whereas such an extreme case is unlikely in practice and therefore more for illustrative purposes, the real possibility of more moderate misalignment has important implications for the broader theory of culture and for practice. Some of these implications will come up in the next sections of the paper. This section focuses on the mechanism behind the misalignment.

So how can such misalignment happen? The fundamental issue is that participants follow norms because of how they *believe* others will react to their actions. At the same time, participants may over time lose track of who the people are with beliefs or preferences that really support the norm, for example because they lost track of who actually started the norms versus who just went along or because they joined after the norm was already established. If there is turnover and participants can't reliably distinguish who has beliefs that really support the norm, then at some point all the people with beliefs that gave rise to the norm may have left, but the norm is still enforced because people aren't fully aware of that turnover; and in the face of uncertainty, people may then find

---

<sup>13</sup>This is difficult to see in the model of Section 6 for two reasons. First, there are, for transparency and analytical simplicity, only two types. As a consequence, anyone of unknown type is 50% likely to have your type. This makes it possible to support norms with only a small majority of people who have been revealed. In a setting with a larger number of types, norms will only emerge when there is a substantial group with shared beliefs/preferences. Second, elimination is by simple majority vote in that model. A more onerous elimination process would require a larger critical mass.

it safer to follow and enforce the norms.<sup>14</sup> In fact, even if a person is sure that all the original ‘instigators’ have left, that person may keep following and enforcing a norm that goes against her own beliefs and preferences because she is not sure whether *others* know that fact and because she is afraid that these others may sanction her if she fails to follow or enforce the norm.<sup>15 16</sup>

### 3.5 A Note on Assumptions

It is probably useful to discuss a few assumptions implicit in the theory that are likely taken for granted by some readers but that are either non-obvious or raise questions for others. This is especially true as this work sits at the intersection of multiple literatures and in an area where these assumptions may really matter.

One such assumption is that people’s preferences do not change. Even though it is well-known that preferences might be malleable to some degree, we do not include that possibility in the analysis or in our more qualitative discussions, for a number of reasons. First, assuming fixed preferences is a conservative assumption, in the sense that it would become easier to explain the results if we dropped it (and let everyone’s beliefs and preferences just change to, say the majority’s). Second, explaining all norms by preference change does not ring true to us and conflicts with our own experience. So we still need an explanation of norms for situations where preferences do not change – and thus we need a theory like this one. Finally, we believe that the theory would hold up even if reasonable degrees and types of malleability are included, but including these assumptions goes beyond the research intended here.

A second assumption is that people are rational, including that their beliefs are (correctly) updated as new information becomes available or when new information can be inferred. While we understand the concerns about unbounded rationality, we see rationality as a approximating assumption that captures ‘logical’ behavior and that works sufficiently well in most settings. Moreover, its analytical power and its ability to develop theories ‘out of sample’ makes it a powerful tool. We believe that the test is ultimately in the pudding: does the theory that emerged from assuming rationality pass the ‘reasonability’ test in terms of what it demands from rationality? We strongly believe that this is the case here.

## 4 Implications of Norms Driven by Shared Beliefs and Preferences

Let us now turn to some further implications of the main result. We will first discuss what the result means for the *effects* of culture before turning to what they mean for the *definition* and conceptualization of culture.

---

<sup>14</sup>This could be captured in the model by introducing a vanishing probability that each participant is (unobservably) replaced with a random draw.

<sup>15</sup>Even worse, in a type of result that may be familiar from the common knowledge literature in game theory, it is possible that norms are enforced even though *everyone* in the organization is aware that the norms conflict with *everyone else’s* beliefs and preferences. The reason is again because of a ‘beliefs about beliefs’ issue: even when everyone knows that fact, each person may be unsure about others’ beliefs and, in the face of that uncertainty, may then find it safer to follow and enforce the norm rather than risk getting eliminated. For a fun introduction to these ideas, see Dewdney (1989) with the puzzle of the three philosophers.

<sup>16</sup>This situation is reminiscent of the fairly well-known ‘Four Monkeys’ scientific experiment reported by Hamel and Prahalad (1996) where monkeys learned from other monkeys not to climb a pole that had bananas at the top because these earlier monkeys had been sprayed when trying to do so. New monkeys did not try to climb the pole even after the spraying was stopped. Unfortunately, there is no evidence, from Hamel and Prahalad or other sources, that such an experiment – where monkeys get sprayed when they climb a pole with bananas – has really been conducted.

## 4.1 Norms as Levers

A first important effect of such social norms is that it is ‘as if’ the shared beliefs/preferences are more wide-spread than they actually are. In particular, because some of the people who don’t share the beliefs/preferences act as if they do, it will be – to some degree – as if they really share these beliefs/preferences. As a consequence, norms work as a lever on shared beliefs and preferences.

This matters in two broad ways. First, it matters because it affects how the cost and benefits of homogeneity, as identified by Van den Steen (2010a), will play out. This ‘norms as levers’ effect has two opposing effects on the costs and benefits of homogeneity. On the one hand, ‘norms as levers’ can potentially extend these costs and benefits to employees who act ‘as if’ they hold the shared beliefs and preferences by following the norm. For some of the effect, however, this will only be the case if the norms are enforced independent of the agency relation under consideration. For example, a principal will feel comfortable delegating to, or reducing monitoring of, an agent who may not share his or her beliefs *only if* the actions are observed by others who will enforce the norms. On the other hand, there is also a reverse effect: ‘norms as levers’ can weaken both the benefits and costs of homogeneity if the principal is unsure whether the agent will obey the norm. For example, when faced with some probability of ‘wrong’ actions and – due to the norm – unable to tell the difference between those who really share the beliefs and preferences versus those who just go along with the norms, the principal may decide to forgo delegation or reduced monitoring altogether.<sup>17</sup>

A second broad way in which ‘norms as levers’ matters is that it helps our understanding of how organizational culture really works. It explains, for example, how people can move between organizations with very different – even conflicting – cultures and nevertheless ‘fit in’ everywhere: if the person has relatively weak beliefs/preferences along the relevant dimensions, they can function in each organization by adjusting and adhering to the prevailing norms. This effect also matters directly for culture change as it implies that there will often be a (considerable) group of members of the organization who follow the norms but are very open to a different culture. Recognizing a conflict between social norms and underlying beliefs and preferences may be key to finding leverage to change organizational culture and norms. Finally, ‘norms as levers’ also has empirical implications when testing the effects of culture. As already pointed out, this can be an interesting angle to interpret work such as Goldberg, Srivastava, Manian, Monroe, and Potts (2016) or Srivastava, Goldberg, Manian, and Potts (2018) who derive measures of culture from people’s similar style in emails. In as far as the styles are reflective of norms, the ability to observe norm adoption provides some interesting angles.

## 4.2 Norms as Control

A second effect of norms is to provide an important channel for culture to function as a control mechanism, as proposed by O’Reilly and Chatman (1996). Given that control has long been viewed as one of the core functions of management and organization (e.g., Fayol 1916), the fact that culture can work as a control mechanism is obviously significant in and of itself. And whereas there exist a lot of formal control mechanisms – from direct monitoring and (subjective) evaluations to incentives – culture as a control mechanism has some distinct advantages, for example because its reach is very different and because it may come across as less openly controlling.

Norms are obviously not the only way in which culture as shared beliefs and preferences exerts control over an organization’s actions. Shaping shared beliefs and preferences serves as a direct

---

<sup>17</sup>Note that this is really caused by the existence of a norm as the type of a player would be known otherwise.

form of control – after all, people with the ‘right’ beliefs and preferences will take the ‘right’ actions even in the absence of norms. There is also an interesting relation between these two forms of control: attempts at culture change may fail to really change beliefs and preferences, yet, by influencing the emergent norms, achieve the ultimate goal of exerting control over actions.

‘Norms as control’ may seem to be just a different way to look at ‘norms as a lever on shared beliefs and preferences’ in that norms make the behavior ‘as if’ there are shared beliefs and preferences. But the two perspectives ultimately focus on very different aspects of the same broad phenomenon. ‘Norms as levers’ is all about homogeneity – and is in principle independent of what the actual actions are. ‘Norms as control’ is all about the actual actions – and is in principle independent of whether there is a degree of homogeneity or not.

The ‘leverage’ effect also clarifies one seemingly contradictory feature of culture as control: ‘norms as control’ have most bite for people who do not share the underlying values; so if norms are purely about people following their values, then norms stop effecting control. ‘Norms as control’ is thus most relevant in setting with *partially* shared beliefs and preferences.

### 4.3 Norms as Signals

The theory also implies that norms are often not necessarily about the behavior itself, but about what that behavior reveals about the person – and thus about other behaviors that the person may engage in. Norms then act as signals for underlying beliefs or preferences. For example, a norm of wearing ties is presumably not driven by a deeply held value that ‘wearing ties is important’ but, for example, by a shared belief that formal dress reflects respect for customers and that respect for customers is critical. In the model of the paper, norm enforcement is indeed largely about what norm violations reveal about the person’s preferences.

This ‘norms as signals’ observation has important implications. First, inferring underlying beliefs and preferences from norms may be difficult, as famously observed by Schein (1985). People may wear ties because of the underlying belief that it reflects respect or because of an underlying preference for ‘elegance.’ But, second, understanding the true driver is fundamental when it comes to changing norms. If some norm is about signaling a particular belief, then the norm can sometimes be changed by providing an acceptable alternative to signal that or by changing the actual belief itself. But both of these are possible only if you understand what that belief or preference really is. Third, it also raises a question whether some types of norms should be considered as just an expression of culture, rather than a real part of culture. For example, is culture about wearing ties or about showing that you care about customers? We return to that question below.

This ‘norms as signals’ also connects to a very different role of norms: norms give new participants a window into the underlying shared beliefs and preferences of the organization. With respect to the distinction between descriptive versus injunctive norms (Deutsch and Gerard 1955), note that in this theory descriptive and injunctive norms essentially coincide: people learn how to behave from observing what others do, but what they implicitly learn is the behavior that satisfies the injunctive norms; their reason for following descriptive norms is precisely because the descriptive norms are also injunctive norms.

### 4.4 Norms as Selection: From Norms back to Shared Beliefs and Preferences

Up to this point, we have looked at how shared beliefs/preferences lead to norms. But there is also an important causal chain in the reverse direction, i.e., from norms to shared beliefs/preferences. In particular, norms force people to behave ‘as if’ they hold the underlying shared beliefs and

preferences. Doing so is fine for people who do share these beliefs and preferences but is often uncomfortable for people with different beliefs and preferences. This is especially true for those who hold such conflicting beliefs and preferences strongly. Such people may leave the organization and they are more likely to do so if their own beliefs and preferences conflict more or if they hold these conflicting beliefs and preferences more strongly. It follows that exactly those people who weaken the culture most are also most likely to leave, thus reinforcing the culture in a virtuous cycle. This is thus another way in which norms end up acting as a lever for culture, but now in a dynamic sense: the norms – generated by the culture as shared beliefs/preferences – lead to self-sorting that further spreads that same culture.

This mechanism can also be used in a very different way: as a lever to change culture. We will discuss that as part of the next section.

## 4.5 Culture Change

This perspective on the relation between shared beliefs/preferences and norms has important implications for how to potentially change an organization’s culture and norms. For this discussion, it is useful to start from the observation that changing an organization’s underlying shared beliefs and preferences is challenging and may require drastic measures. There are in particular two mechanisms through which such change can happen (Schein 1985, Chatman 1991, Van den Steen 2010b). By far the most effective mechanism is turnover: replace people who hold the ‘wrong’ beliefs and preferences with people who hold the ‘right’ beliefs and preferences. Considerable turnover – both forced and voluntary – is indeed typical for settings where organizations have changed their culture. The second mechanism is to try to modify the actual beliefs and preferences that people hold. While there are some exceptional circumstances where this may happen relatively quickly,<sup>18</sup> this is almost always a slow and difficult process, if it happens at all. By and large, then, changing beliefs and preferences seems to happen mainly through turnover, both forced and voluntary, which is not only painful but also very costly – including the costs of lost experience and specific human capital, and human cost.

Building on these observations, the paper first of all implies a fundamental cautionary tale about trying to change norms without tackling the underlying shared beliefs and preferences. In particular, doing so may be futile, as the organization will likely revert to the old norms once the pressure is removed. (Personal observations suggest that such ‘reversion’ is quite common.) As will be discussed below, temporarily imposing new norms *may* lead to such a change in underlying beliefs and preferences, and then have a long-term impact, but that is far from guaranteed. So one should always be wary about the sustainability of apparent short term changes in norms.

But the paper does also suggest a number of channels that can *facilitate real culture change* (in the sense of changing shared beliefs and preferences). The first – and probably most important one – is the observation that often a lot of the people do not have strong beliefs or preferences and just go along with the prevailing norms. Such participants can often switch quite easily from one culture regime to another. This also explains why organizations can really change their culture without having 100% turnover. (Note that, as suggested by the model in Section 6, employees who adapt to a changing culture will not necessarily be perceived as having the ‘wrong’ beliefs and values, but simply as having weak beliefs and values.) To assess the likelihood of success with a desired culture change, it is therefore helpful to have some assessment of the share of people who

---

<sup>18</sup>An example would be a setting where current norms or behaviors have disastrous consequences that are not readily observed by those who follow these norms or engage in these behaviors. Pointing out these consequences may have drastic impacts on beliefs and preferences over such behaviors, and then lead to quick changes in norms.



hold the current shared beliefs and preferences strongly versus those who just go along. Note that this also predicts that employees hired after the culture change are more likely to fit with the new culture than those who remained, consistent with the empirical findings of Campbell (2012).

Second, Subsection 4.4 suggested the use of ‘norms as selection’ as a tool to facilitate culture change. In particular, temporarily forcing different norms may lead to self-sorting towards people who hold beliefs and preferences consistent with these enforced norms. But this works only under certain conditions. The issue is that unless there is a change in the underlying beliefs/preferences that drive the norms, the organization will revert to its old norms as soon as the forcing ends. If participants realize that, they have limited reason to sort: they can just live through the temporary discomfort and be back to their old ways. So a *vicious/virtuous cycle* emerges: the prospect of no change in underlying beliefs and preferences removes the incentives to self-sort and thus ensures that there will be no change. Three things are therefore important to make this work. The first is a credible prospect of a substantial period with the new norms, enforced or otherwise, for example by an organization change that creates some commitment to the new norms. The second is to encourage turnover that furthers the change in underlying beliefs and preferences; for example, by facilitating employees’ voluntary departure, by explicitly using forced turnover as a response to norm violations, and by making the new norms very visible in recruiting. The third is to be very cognizant of the vicious or virtuous cycle by creating critical mass and being proactive about managing perceptions, for example by making any progress very visible.

A third channel for culture and norm change that follows from the paper is how a change in the task environment that creates or breaks co-dependence along certain task dimensions or along certain social dimensions may be helpful to change culture and norms. In particular, such changes in the task environment may alter the forces that sustain the norms and that lead to shared beliefs and values. If, for example, employees are forced to cooperate more and work more in teams, they will start to care more about the beliefs and preferences of their teammates, setting in motion a process towards more shared beliefs and preferences and towards social norms, and the other way around.

A fourth channel is the observation that some norms may be fairly superficial expressions of deeper held beliefs and preferences. Understanding whether or when that is the case can be helpful to determine whether some norms might be easier to change than others.

The final channel rests on the observation that it is largely the *perceived* shared beliefs and preferences that matter. This is clear from the observation in Subsection 3.4 that we can have norms that contradict the actual shared beliefs and preferences: if there is a shared perception that most employees – especially powerful employees – hold a specific set of beliefs and preferences then that will sustain norms and enforcement of norms that are consistent with these perceived beliefs and preferences. But in case of such misalignment, revealing the true beliefs and preferences may drastically alter the forces that sustain norms and lead to rapid culture change. When culture change is needed, it is thus again helpful to try to discover the true versus perceived beliefs and preferences, as that may provide a powerful lever for culture change.

Overall, whereas wholesale culture change is typically a very difficult and slow process, this paper points out that 1) there are non-typical cases – that can be identified – where wholesale culture change is a lot easier and 2) sometimes changes in part of the culture, for example part of the norms, may be much easier and that may be all that matters.

## 4.6 Effects of Shared Beliefs and Preferences Themselves (beyond Norms)

All these effects of norms raise the important question whether *all* effects of shared beliefs and preferences may in fact go through norms. This question matters not only for practice – to really understand the effects of culture – but also for the theory: if all effects do go through norms, then ‘shared beliefs and preferences’ could potentially be redundant as part of the definition.

The answer, however, is unequivocally that shared beliefs and preferences matter independent of norms, as a lot of the effects of shared beliefs and preferences do not depend on norms. This is obvious from Van den Steen (2010a): that paper does not have any norms, yet shared beliefs and preferences have a wide range of effects.

## 4.7 Implications for the Definition and Conceptualization of Culture

The results of this paper also have implications for the definition of culture.

First, this section has made clear – if it wasn’t already – that norms are a critical facet of culture. So any good conception or definition of culture must carefully consider how it captures that important facet, directly or indirectly.

Second, the results of the paper offer arguments both in favor and against explicitly including norms in definitions of culture. Most obviously, the paper shows that norms may be implied by the (perceived) shared beliefs and preferences. One could therefore argue that it suffices to specify these (perceived) shared beliefs and preferences, i.e., that shared beliefs and preferences are a ‘sufficient statistic’ for the norms. In the other direction, however, the paper does not provide any evidence, theoretical or empirical, that *all* relevant norms result from this mechanism. But this raises an important issue that has been mostly implicit in the literature: the culture literature seems to have implicitly limited its interest to these norms that are driven somehow by shared beliefs and preferences (or values), which would weaken that argument. Another argument in favor of including norms in the definition is that the same set of shared beliefs and preferences can give rise to quite different norms, suggesting that shared beliefs and preferences are in fact *not* a sufficient statistic for norms.

Third, together with Van den Steen (2010a), the paper does provide a clear argument that shared beliefs and preferences have performance effects beyond norms and should thus be included separately in any definition or conception of culture. But, as discussed earlier, it does see the choice between ‘values’ and ‘preferences’ as very consequential for the relevance of culture.

Finally, this paper makes very clear that it is important to distinguish between the *actual* shared beliefs and preferences and the *perceived* shared beliefs and preferences. The analysis and results of this paper suggest in fact that, unlike the existing definitions in the literature, culture should probably be defined as the ‘*perceived* shared beliefs and preferences’ (among a group of critical size) with the norms that follow from it. It turns out that all results of, for example, Van den Steen (2010a), Van den Steen (2010b), or Li (2016) go through under that new definition (as there is no private information in these models).

The definition of culture is still an open research question. We just hope to have brought useful perspectives and questions for that discussion to the table.

## 5 Culture, Norms, and Strategy: Does Culture Eat Strategy for Breakfast?

Traditionally, strategy has thought of culture as an input – in the sense of an asset or a resource – that can help or hurt the strategy (e.g., Barney 1986). A frugal culture, for example, supports a low-cost strategy but may hurt a high-end strategy. Hence, a company needs a culture that fits its strategy, just like it needs a brand that fits its strategy. In this traditional view, culture is one of many inputs to strategy and you can't execute the strategy with the wrong culture or with the wrong brand.

This perspective misses a critical aspect of the relationship between culture and strategy, however. Executives don't typically ask 'which is more important, strategy or brand?', but they *do* ask 'which is most important, strategy or culture?' Somehow the relationship between culture and strategy thus seems to be fundamentally different than between brand and strategy. The theory in this paper suggests a perspective that is consistent with the above observation. In particular, it shows that culture is not just one of many inputs into strategy or a component of strategy execution – though it is that too; culture is additionally a direct *substitute* for strategy and often *competes* with it for control. In particular, the 'core guidance' definition of strategy as 'the smallest set of choices to optimally guide other choices' (Van den Steen 2017) makes clear that the role of strategy is to guide the company's decisions. Strategy is thus essentially a control mechanism to achieve coordination and direction. But culture is a control mechanism too, through norms and through the direct influence of shared beliefs and preferences. This makes strategy and culture as shared beliefs and preferences indeed *substitutes and potential competitors*. To return now to the motivating observation: because brand is seen as part of strategy, juxtaposing 'brand vs strategy' makes little sense<sup>19</sup>; culture, on the other hand, is seen as a potential alternative or competitor to strategy, and then the juxtaposition makes perfect sense.

Strategy unfortunately has a weak hand when it needs to compete with culture. The issue is that strategy needs credibility or reliability to be effective (Van den Steen 2017): employees will take strategy as guidance only if they believe others will follow it. This results in a vicious/virtuous cycle where people follow the strategy (only) because they believe that others will follow it, which is fragile. A conflict between culture and strategy often ends up turning it into a vicious cycle where culture wins out because cultural norms derive their credibility from stable beliefs and preferences. Culture thus interferes with the effectiveness of strategy on a very fundamental level. In this sense, culture may indeed eat an (*inappropriate*) strategy for breakfast.

Of course – as a control mechanism – strategy does have important benefits and advantages over culture. Strategy, when done well, is more under the direct control of management and much more malleable than culture; it can reach where culture can't; and it can be made much more clear and unambiguous. It follows that it is critical both to design strategy and to modify culture so as to ensure that culture becomes a carrier of strategy, rather than a competitor.

This paper, even though just a first cut at this issue, already has some practical implications on this issue. First, the earlier discussion of culture change already gives some pointers as to when it might be feasible to modify culture. For example, sometimes norms may conflict with the strategy, but the underlying beliefs and preferences do not. Caring about students may be expressed in a norm to make classes into easy entertainment or in a norm to make classes difficult and challenging. Two very different strategies can thus be consistent with the underlying beliefs and preferences, as long as these can be channeled into the right norm. Second, discussions of culture often talk about

---

<sup>19</sup>Which is more important, a car or a wheel?

a company’s culture as an aligned monolith, whereas cultures are instead complex patchworks of different – at times conflicting – beliefs and preferences and norms that follow from them. Changing one aspect of culture, when well targeted, may then be much more feasible than a wholesale culture change. Third, it is also important to note that strategy will slowly shape culture, as it will lead to self-sorting towards people who align with the strategy. So a strategy change, when actually implemented, may become easier over time as the culture adjusts.

Overall, the theory makes a strong case for the importance of developing a detailed understanding of culture and its ‘structure.’ Small conflicts, contradictions, and inconsistencies may be helpful levers for modifying culture to make it carry the right strategy. Whereas bad strategy may get eaten by culture, great strategy knows how to be carried by culture (and make culture do the hard work).

## 6 A Formal Model

This section presents a very simple model that captures the main forces and results at the heart of this paper and illustrates some of its core insights. Whereas this formal analysis provides solid support for the paper, the model is too specific – due to its limited focus – to provide a full formalization of the theory (which thus requires further research). Instead, the purpose of the model and analysis is threefold. First, the model and analysis illustrate the mechanism and make very transparent exactly how the logic works. Second, the model and analysis work as a basic check on the more qualitative argument. Third, and closely related to the second, the model also pushes the boundaries of the original argument. In this particular case, the analysis actually revealed an important omission in the original qualitative arguments that then led to new insights, as we will discuss. (The online appendix has a more formal and precise statement of the model and result and contains the proofs.)

The model captures a stylized setting of an organization with  $N$  employee positions, where we assume  $N = 5$  for simplicity.<sup>20</sup> The employee in position  $n$  needs to choose a public course of action  $a_n \in \{L, R\}$  and a private action  $b_n \in \{L, R\}$ , with a choice between Left ( $L$ ) or Right ( $R$ ). The public action is taken up-front and visible to all – like setting an official policy – whereas the private action is taken at the end and private information – as in the case of a last minute switch or some deviation from policy in the field. Employees are randomly selected from an infinite pool of applicants who differ in their preference for  $L$  versus  $R$  (or, equivalently, who disagree – in the sense of differing priors – whether  $L$  or  $R$  is the better course of action). In particular, half the pool prefers  $R$  and half prefers  $L$ . Moreover, both among those who prefer  $R$  and among those who prefer  $L$ , a proportion  $1 - \gamma$  are what we call ‘regular’ types, labeled  $r$  and  $l$ , while a proportion  $\gamma$  are ‘strong’ types, labeled  $R$  and  $L$ . We assume (for analytical reasons)  $\gamma$  to be vanishingly small.<sup>21</sup> Let  $\tau_n \in \{L, R\}$  denote the action preference of the employee in position  $n$ . Whereas these preferences are private information, everyone knows the composition of the pool.

The game consists of two parts. (For a formal timeline, see the online appendix.) In the first part, employees are hired and choose their public action, but can also try to eliminate other employees if they so prefer. They will, in fact, get (at most) three opportunities to eliminate others. Formally, after  $N$  employees are selected randomly from the pool, they simultaneously choose their public actions from  $\{L, R\}$ . After observing each others’ actions, all employees simultaneously ‘vote’

<sup>20</sup>Because part of the proof (currently) consists of brute force calculations of incentive constraints, it is helpful to keep the number of positions down. We limit for the same reason the number of elimination rounds to three.

<sup>21</sup>This  $\gamma \downarrow 0$  assumption vastly simplifies the determination of beliefs.

which of the other employees they would like to ‘eliminate’. Any employee who gets a majority of votes ( $> N/2$ ) against her is immediately eliminated and replaced by a new employee randomly selected from the pool. All replacements happen simultaneously. At the moment an employee is eliminated, all remaining employees incur a cost  $\epsilon$  for each employee eliminated. (This can be interpreted as the personal cost of turnover, e.g., training, taking over duties, or separation costs, or some empathy with those who are removed.) If no one was replaced, the game proceeds to the second part. But if anyone was replaced, then there is a second elimination vote (with immediate replacements) after all employees had the opportunity to again choose (or re-choose) their public actions. (‘Old’ employees can thus switch their earlier action choice, if they want to.) As before, the game proceeds to the second part if no one was replaced, but (now for the last time) if someone was replaced, there is a third and final elimination vote (with immediate replacements) after employees had a chance to choose or re-choose their actions in a third public action round. Now the game proceeds to the second part no matter what. (If any employee was eliminated and replaced in the last round, her public action stands.) Finally, in the second part, employees choose their private actions and their final payoffs are realized. A regular employee in position  $n$  gets utility

$$U_n = U + (VJ_{a_n=\tau_n} + vJ_{b_n=\tau_n}) + \sum_m (WJ_{a_m=\tau_n} + wJ_{b_m=\tau_n})$$

where  $J_X = 1$  if condition  $X$  is satisfied and equals  $-1$  if it is not. Here,  $U > 0$  is some benefit from being part of the organization,  $V, v > 0$  are the benefits/costs from the employee’s own choices, and  $W, w > 0$  are benefits/costs that an employee enjoys/suffers from the choices of others in the organization. We will finally assume that strong-type employees (types  $L$  and  $R$ ) choose mechanically: they always choose the action of their type  $a_n = b_n = \tau_n$  and vote against anyone who they believe prefers the opposite action.<sup>22</sup>

The key point of the analysis now is to show that this setting exhibits a very natural equilibrium, which we call the ‘standard norm equilibrium,’ in which either an  $L$ -norm or an  $R$ -norm emerges (in a sense that we will make precise below) which is upheld and enforced even by employees who prefer the opposite course of action. This result will also allow us to derive some of the comparative statics we discussed.

To explain what the equilibrium looks like, it is helpful to start with a few observations. These are unfortunately somewhat technical but are necessary for a precise explanation. Most of these are about the beliefs that employees hold about each other. First, because all choices before the end-round are public and everyone starts from the same prior beliefs, everyone will share the same belief about a particular employee’s type (with the exception, obviously, of that employee – as she knows her own type). Call this the employee’s ‘perceived type’. Second, in the equilibrium that we consider, only a few beliefs are possible as ‘perceived types’. Most importantly, because the equilibrium is in pure strategies, if the perceived type of a player puts strictly positive probability on more than one type, say on  $R$  and  $r$ , then these probabilities will be proportional to those in the prior, in this case  $(1 - \gamma, \gamma)$ , and we then say that the employee’s perceived type is  $R/r$ . (Note that, with this notation, the prior type is  $R/r/L/l$  with probabilities  $(50 - \gamma/2, \gamma/2, 50 - \gamma/2, \gamma/2)$ .) A second restriction is that if an employee’s perceived type puts positive probability on more than one type and one of these is a strong type, then it must also put positive probability on the corresponding regular type. So  $R/l$  is not possible and neither is  $R/L/l$ , but  $R/r/l$  is possible.

We also need to introduce some terminology. If an employee’s perceived type includes both  $r$  and  $l$ , for example  $R/r/l$ , we say that the employee is *unrevealed*. Note that in this case, the

---

<sup>22</sup>These choices can be endogenized with appropriate utility functions but at a considerable analytical cost.

beliefs will put essentially 50/50 weight on  $l$  and  $r$  (as  $\gamma$  is vanishing). It then follows that, in the equilibrium that we consider, an employee will be either unrevealed, or her perceived type is one of the following:  $L, l, L/l, R, r, R/r$ . Naturally, in the former 3 cases we will say that the player was  $L$ -revealed; in the latter 3 cases,  $R$ -revealed. Note that upon joining the organization, an employee's type is unknown and thus unrevealed, but the employee's actions can 'reveal' her type (partially or fully). We will say that there is an  $L$ -majority (and analogously for an  $R$ -majority) if the number of  $L$ -revealed employees is strictly higher than the number of  $R$ -revealed employees.

We can now define norms. An  $L$ -norm – and analogously for an  $R$ -norm – means that all regular employees who are  $L$ -revealed or unrevealed choose  $a_n = L$  and vote to eliminate those and only those who have been  $R$ -revealed (which will include anyone who just chose  $a_n = R$ ). (Note that strong  $R$ -type employees and already- $R$ -revealed employees will violate the norm. The first because they always choose  $R$ , the second – which occurs only off equilibrium – because they have already been revealed to be of the 'wrong' type anyways.)

Let us now describe the 'standard norm equilibrium'. (We will discuss the interpretation at the end of this section, after the proposition.) The following description applies both on and off the equilibrium path, except for a few unusual cases (e.g., if a majority-type player is wrongly perceived to be of the minority type) that we deal with in the online appendix.

- Consider any decision point prior to the third elimination round. If there is a  $L$ -majority, then the employees follow the  $L$ -norm, and analogous for an  $R$ -majority and the  $R$ -norm. If there is no majority of either type, then all employees choose a course of action according to their real type and vote to eliminate anyone who was revealed to be of the opposite type. (So an  $l$ -type will choose  $L$  and vote to eliminate anyone who was  $R$ -revealed.)
- In the (potential) third elimination round and for the final private action choices, all employees vote and choose according to their true type. (So an  $l$ -type will choose  $L$  and vote to eliminate anyone who was  $R$ -revealed.) In this case, norms do not arise because there is no threat of future elimination to enforce norms.

Beliefs (about employees' types) are determined by Bayes' rule whenever possible. When Bayes' rule doesn't have any bite – following off-equilibrium outcomes – the equilibrium specifies 'skeptical' beliefs: a player who has ever violated – or failed to enforce – the norm in any round with a majority is perceived to be the minority type. (We specify the details in the online appendix.) Such skeptical beliefs are intuitive, and essentially prevent employees who have violated a norm from 'rehabilitating' themselves by later adhering to the norm.

The outcome of our norm equilibrium is then as follows. *First round:* Once employees have been randomly selected, they all choose a course of action according to their actual type, which thus reveals their type and which always results in either an (absolute)  $L$ -majority or an (absolute)  $R$ -majority. The majority-type employees then vote to eliminate all minority-type employees in the first elimination round. *Subsequent rounds, until third elimination round:* In subsequent action rounds, the majority norm is observed. Note that employees of the regular minority type, by following the norm, remain unrevealed and will not be eliminated. But employees of the *strong* minority type will violate the norm, thus revealing their type. When such a strong minority type violates the norm, the regular minority types will vote with everyone else to eliminate that employee – that is, they enforce a norm that goes against their own preferences – in order to remain unrevealed. *Final round:* If the game proceeds to the third elimination round, then everyone votes their type. In the final (private) action round, everyone takes their preferred action.

The following proposition not only shows that the ‘standard norm equilibrium’ indeed exists but also gives some relevant comparative statics for when the equilibrium exists and when it doesn’t. As mentioned before, we focus on a vanishingly small probability of strong types,  $\gamma \downarrow 0$ . In terms of notation, let  $\Omega$  be the allowed parameter space  $\{(V, v, W, w, \epsilon) : V > 0, v > 0, W > 0, w > 0, \epsilon > 0\}$  (excluding  $\gamma$ ) and let  $\omega$  be some generic point in  $\Omega$ .

**Proposition 1** *A ‘standard norm equilibrium’ exists for sufficiently large  $U$ , small  $V$ , large  $v$ , small  $W$ , and intermediate  $w$ , in the following sense:*

*There exists a subset  $S$  of  $\Omega$  with non-empty interior such that:*

1. *The ‘standard norm equilibrium’ exists for any element in the interior of  $S$*
2. *If  $\omega = (U, V, v, W, w, \epsilon) \in S$  then a)  $w \geq \epsilon$  and b) if  $\omega' = (U', V', v', W', w', \epsilon) \in \Omega$  is such that  $U' > U, V' < V, v' > v, W' < W, \epsilon < w' < w$  then  $\omega'$  is also in  $S$*
3. *The ‘standard norm equilibrium’ does not exist for any  $\omega \in \Omega$  outside of  $S$*

So what does this proposition say? Let us start with the comparative statics on when the ‘standard norm equilibrium’ exists. The norm equilibrium exists for high  $U$ , i.e. for attractive organizations, because that makes it worthwhile to follow the norm to avoid being eliminated; for low  $V$ , i.e., low cost of the personal public actions, because that makes it cheap to signal a type; and for large  $v$ , i.e., large benefits from personal private actions, because that again makes the organization attractive. Consider, finally, the comparative statics on  $W$  and  $w$ , i.e., how much an employee is impacted by others’ actions, which captures the degree of interdependence across employees. These are driven by two forces. On the one hand, you only have reason to enforce a norm or eliminate anyone if their actions affect you. So norms are only possible when  $w$  is large enough ( $w > \epsilon$ ).<sup>23</sup> On the other hand, being part of an organization where everyone follows norms that go against your personal beliefs and preferences is frustrating, especially when their actions really affect you. So very high  $W$  and  $w$  make an organization unattractive to an employee who disagrees with the norms, so that that employee won’t bother to follow the norm. (In such case, all minority types will get eliminated, so the organization ultimately has no norms but perfectly shared beliefs/preferences.) We had originally overlooked this second intuition but the formal analysis brought it out. Overall, norms are thus more likely with intermediate interdependence. And as interdependence increases, an organization moves from having no shared beliefs or preferences and no norms, to having partially shared beliefs and preferences with norms that make everyone behave ‘as if’, to ultimately having completely shared beliefs and preferences but again no norms.

There are also a few interesting observations about the equilibrium itself. First, people often get eliminated not so much because of their public action ( $a_n$ ) that triggered the elimination, but because of what that action reveals about their type (and thus about their future private action  $b_n$ ).<sup>24</sup> Second, the flipside is, of course, that people follow a norm to avoid revealing that they have the ‘wrong’ beliefs or preferences. And with elimination votes public, they have to also enforce the norm as if they held the ‘right’ beliefs and preferences. That makes the norm thus serve as a lever on the shared beliefs of the majority type, so that these beliefs appear to a naive observer more widespread than they really are. Third, the off-equilibrium results that an  $L$ - or  $R$ -majority results in a norm but that a stalemate doesn’t, start to capture the idea that you need a critical mass for a norm to emerge. (Models with, e.g., more than 2 types would make this point more clearly.)

<sup>23</sup>The fact that  $W$  is not subject to that condition has to do with the specific setup. In most settings, there would be a similar condition on  $W$ .

<sup>24</sup>In the third round, for example, norm violators get eliminated even though their public action  $a_n$  is frozen.

## 7 Conclusion

This paper showed that culture as ‘shared beliefs and preferences’ can give rise to norms, including norm enforcement even by people who disagree with it. It then explored the broad implications of this connection for the nature and functioning of culture and for its relationship with strategy.

The paper raises a number of interesting further research questions, such as:

- Can we develop and test theory on the dimensions along which culture and norms are more or less likely to evolve – assuming a broad view of culture as ‘perceived shared beliefs and preferences’ and the norms that follow from it?
- Can, in some settings, formal organization be used to counter the effect of culture in cases where strategy and culture conflict?
- Would experiments that change the task environment be able to encourage the emergence or extinction of social norms?
- Which aspects of culture are (theoretically and/or empirically) most likely to interfere with strategy?

On a higher level, this paper hopes to have made two contributions: 1) encourage the view that microfoundations can really improve our understanding of complex phenomena, 2) contribute new questions and insights to the discussion on how to most effectively define culture and how to think about the interaction between culture and strategy.

## References

- ABERNETHY, M., H. DEKKER, AND A. SCHULZ (2015): “Are Employee Selection and Incentive Contracts Complements or Substitutes?,” *Journal of Accounting Research*, 53(4), 633 – 668.
- BARNEY, J. B. (1986): “Organizational Culture: Can it be a Source of Competitive Advantage?,” *Academy of Management Review*, 11(3), 656– 665.
- CAMPBELL, D. (2012): “Employee Selection as a Control System,” *Journal of Accounting Research*, 50(4), 931–966.
- CARRILLO, J. D., AND D. GROMB (1999): “On the Strength of Corporate Cultures,” *European Economic Review*, 43, 1021– 1037.
- CHATMAN, J. A. (1991): “Matching People and Organizations: Selection and Socialization in Public Accounting Firms,” *Administrative Science Quarterly*, 36, 459– 484.
- CHATMAN, J. A., AND C. A. O’REILLY (2016): “Paradigm lost: Reinvigorating the study of organizational culture,” *Research in Organizational Behavior*, 36, 199–224.
- COLEMAN, J. S. (1990): *Foundations of Social Theory*. Belnap Press.
- CRÉMER, J. (1993): “Corporate Culture and Shared Knowledge,” *Industrial and Corporate Change*, 3(3), 351– 386.
- DEUTSCH, M., AND H. B. GERARD (1955): “A study of normative and informational social influences upon individual judgment,” *The Journal of Abnormal and Social Psychology*, 51(3), 629–636.
- DEWDNEY, A. K. (1989): “Computer Recreations,” *Scientific American*, 260(1), 106–109.



- EPSTEIN, G. S., AND R. LINDNER-POMERANTZ (2017): "The Survival of Unique Corporate Cultures," *Managerial and Decision Economics*, 38(4), 622–629.
- FAYOL, H. (1916): "Administration Industrielle et Generale," *Bulletin de la Societe de l'Industrie Minerale*, 10, 5–164.
- GHEMAWAT, P. (2001): "Distance Still Matters: The Hard Reality of Global Expansion," *Harvard Business Review*.
- GIBBONS, R., AND R. HENDERSON (2012): "Relational Contracts and Organizational Capabilities," *Organization Science*, 23(5).
- GOLDBERG, A., S. B. SRIVASTAVA, V. G. MANIAN, W. MONROE, AND C. POTTS (2016): "Fitting In or Standing Out? The Tradeoffs of Structural and Cultural Embeddedness," *American Sociological Review*, 81(6), 1190–1222.
- GRANT, R. M. (2016): *Contemporary Strategic Analysis*. John Wiley & Sons.
- GROYSBERG, B., J. LEE, J. PRICE, AND J. Y.-J. CHENG (2018): "The Leaders Guide to Corporate Culture," *Harvard Business Review*.
- GUIO, L., P. SAPIENZA, AND L. ZINGALES (2006): "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives*, 20(2), 23–48.
- (2015): "The value of corporate culture," *Journal of Financial Economics*, 117(1), 60–76.
- HAMEL, G., AND C. PRAHALAD (1996): *Competing for the Future*. HBR Press.
- HOGG, M. A., AND J. C. TURNER (1987): "Social Identity and Conformity: A Theory of Referent Informational Influence," in *Current Issues in European Social Psychology*, ed. by W. Doise, and S. Moscovici. Cambridge: Cambridge University Press.
- HUSTON, T. L., AND G. LEVINGER (1978): "Interpersonal attraction and relationships," *Annual Review of Psychology*, 29, 115–56.
- KREPS, D. M. (1990): "Corporate Culture and Economic Theory," in *Perspectives on Positive Political Economy*, ed. by J. E. Alt, and K. A. Shepsle, pp. 90–143. Cambridge University Press, Cambridge.
- LAZEAR, E. P. (1995): "Corporate Culture and Diffusion of Values," in *Trends in Business Organization: Do Participation and Cooperation Increase Competitiveness?*, ed. by H. Siebert, pp. 89–133. Mohr Publishing Company, Tubingen.
- LEVINE, S. R. (2019): "A Superb Strategy Without The Culture To Support It Leads To Failure," *Forbes*.
- LI, H. (2016): "Developing Shared Knowledge in Growing Firms," *Journal of Law, Economics, and Organization*, 33(2), 332–376.
- MCPHERSON, M., L. SMITH-LOVIN, AND J. M. COOK (2001): "Birds of a Feather: Homophily in Social Networks," *Annual Review of Sociology*, 27, 415–444.
- O'REILLY, C. A., AND J. A. CHATMAN (1996): "Culture as Social Control: Corporations, Cults, and Commitments," *Research in Organizational Behavior*, 18, 157–200.
- PARKS, L., AND R. P. GUAY (2009): "Personality, values, and motivation," *Personality and Individual Differences*, 47(7), 675–684.
- PARSONS, T. (1951): *The Social System*. New York: Routledge.

- PRASAD, S., AND M. TOMAINO (2019): “Resources and Culture in Organizations,” Working Paper.
- SALONER, G., A. SHEPARD, AND J. PODOLNY (2001): *Strategic Management*. John Wiley & Sons, New York, book draft : New York : John Wiley & Sons.
- SCHEIN, E. H. (1985): *Organizational Culture and Leadership*. Jossey-Bass Publishers, San Francisco, CA.
- SIEGEL, J. I., A. N. LICHT, AND S. H. SCHWARTZ (2013): “Egalitarianism, Cultural Distance, and Foreign Direct Investment: A New Approach,” *Organization Science*, 24(4), 1174–1194, Published online in 9/2102.
- SRI (2009): *Strategy Reader*. Strategy Research Initiative.
- SRIVASTAVA, S. B., A. GOLDBERG, V. G. MANIAN, AND C. POTTS (2018): “Enculturation Trajectories: Language, Cultural Adaptation, and Individual Outcomes in Organizations,” *Management Science*, 64(3), 983–1476.
- VAN DEN STEEN, E. (2010a): “Culture Clash: The Costs and Benefits of Homogeneity,” *Management Science*, 56(10), 1718–1738.
- (2010b): “On the Origin of Shared Beliefs (and Corporate Culture),” *Rand Journal of Economics*, 41(4), 617–648.
- VAN DEN STEEN, E. (2017): “A Formal Theory of Strategy,” *Management Science*, 63(8), 2616 – 2636, Published online in Articles in Advance in July 2016.
- (2018): “The Strategy in Competitive Interactions,” *Strategy Science*, 3(4), 574–591.
- WEBER, R. A., AND C. F. CAMERER (2003): “Cultural Conflict and Merger Failure: An Experimental Approach,” *Management Science*, 49(4), 400– 415, USC Center for Law, Economics, and Organization Research Paper C01-10.