11

Fair Social Contracts and the Foundations of Large-Scale Collaboration

Eric D. Beinhocker

Abstract

Large-scale collaborations with non-kin are a unique feature of human societies and foundational to human civilization. Individual relationships with collectives can be thought of as "social contracts." This chapter argues that perceptions of social contract fairness are essential for effective large-scale collaboration and that factors likely to create perceptions of fairness are subject to empirical analysis. Drawing on empirical behavioral and social science literature, the chapter proposes nine dimensions of social contract fairness. Each dimension is distinct, imperfectly substitutable, and universal, although with individual and cultural variations in interpretations and preference weightings. Here, these nine dimensions are applied to the breakdown in political collaboration in the United States. It is argued that for large segments of the U.S. population, all nine dimensions of social contract fairness were broken during the mid-1970s–2010s. The chapter concludes with thoughts on social contract repair and further research into perceptions of social contract fairness.

Introduction

Modern society is built on large-scale, complex, collaborations among "strangers," people who are neither kin nor with whom one has thick personal bonds (Seabright 2010). Firms, global supply chains, governmental bodies, scientific collaborations, religious communities, cultural organizations, and many other institutions provide examples of thousands to millions of people collaborating toward some shared end. Most people in these networks will have never met, let alone be related or known to each other personally. Such large-scale collaboration among strangers appears to be a uniquely human capability that developed during the Neolithic period (Bowles and Gintis 2011; Gintis 2011; Henrich et al. 2010a). In traditional, pre-agricultural societies, group sizes

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

typically ranged from a dozen to 150 individuals, although larger collections of groups gathered into societies that could number in the thousands (Bird et al. 2019; Diamond 1997; Dunbar 1992, 1993; Hamilton et al. 2007; Graeber and Wengrow 2021). These groups were comprised mostly of related individuals and individuals with thick, personal bonds. In contrast, in most present-day societies, individuals not only have kinship and personal relationships but also abstract relationships with large collectives of strangers, such as their employer, government, or nation.

We can refer to these abstract relationships between an individual and a collective as a "social contract." In this chapter I will argue that our ability to collaborate with such large groups of strangers depends on social contracts being perceived to be "fair" by the individuals in the groups. If individuals perceive a social contract to be fair, then they are more likely to engage in high-functioning collaborative behaviors; in contrast, if individuals perceive the arrangements to be unfair, then they are more likely to withdraw their collaboration or engage in destructive behaviors. In this sense, the perceived fairness or unfairness of social contracts is foundational to establishing and maintaining large-scale collaborations.

Whether a social contract is likely to be perceived as fair or unfair is a question subject to empirical analysis. Drawing on a growing literature in moral psychology, social psychology, neuroscience, anthropology, organizational studies, and behavioral economics, this chapter advances an empirically informed hypothesis that judgments about the fairness of social contracts are based on nine dimensions arising from underlying moral instincts and cultural norms for relational fairness, process fairness, and distributive fairness, that evolved in humans to support cooperation with non-kin. I will further claim that while there may be individual preferences and cultural variations in interpretations and weightings across the nine dimensions, they are highly universal, distinct, and only imperfectly substitutable (i.e., they do not collapse to a unidimensional notion of utility). In addition, I argue that when a social contract satisfies these nine dimensions, it enables participants to trust the collective they are interacting with and make prosocial collaborative choices, such as making costly investments in collective action with uncertain future payoffs and engage in altruistic behaviors. When, however, the nine dimensions of fairness are violated, not only may this result in a loss of trust and withdrawal of collaboration, but it may also trigger antisocial behaviors that undermine capacities for large-scale collaboration or stoke conflict within and between groups.

After describing this empirically informed hypothesis, I will apply it to the specific problem of political discord in the United States. I will argue that a critical reason the United States has experienced a widespread loss of institutional trust, breakdown in political collaboration, and rise of political populism is that all nine dimensions of social contract fairness were degraded for large segments of the population during the mid-1970s–2010s. This implies that restoring social contract fairness is an essential step to restoring trust and

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

functional politics. The nine dimensions provide a useful guideline for such a program of social contract renewal.

The Problem of Complex Collaboration at Scale

To see why fair social contracts are a necessary condition for collaboration at scale, it is helpful to clarify some of the specific challenges associated with initiating and sustaining such collaborations. The term "collaboration" is used in varying ways in behavioral and social science (and in this volume), so I will briefly define how I am using the term. Collaboration involves agents aligning their behaviors to achieve some mutual end. Yet, one can think of a spectrum of such behavioral alignment, from simple and mechanical, to complex and cognitively demanding. I will use the term "collaboration" to refer to more complex, cognitively demanding forms of behavioral alignment and distinguish it from "coordination" and "cooperation" as follows:

- *Coordination* occurs when agents align their behaviors to achieve some collective end. Processes of coordination may be quite mechanical and not require complex cognitive capacities. For example, honeybees regulate their hive temperature by generating heat from their muscles when the temperature is too low and beating their wings when it is too hot. Individual bees are genetically programmed to engage in thermoregulation at varying temperature points such that when the temperature deviates from the target by a small amount, a small number of bees thermoregulate; when the temperature deviation is greater, more bees join in. This feedback mechanism coordinates the bees to smoothly respond to temperature fluctuations, maintaining the hive at roughly 35ºC.
- *Cooperation* occurs when agents align behaviors in mutually beneficial ways, *anticipating or understanding the behavior of other agents*. Imagine a dog and a human playing a game of fetch. It is a cooperative game: if both behave in certain ways, both get pleasure (although perhaps with asymmetric payoffs for the canine). This setting is more complex and cognitively demanding than the simple coordination example (Moll and Tomasello 2007; Tomasello and Carpenter 2007). For *shared intentionality*, the players need to understand the rules of the game and payoffs and then voluntarily *choose* to enter the game (e.g., the dog initiates the game by dropping a ball at the human's feet and wagging its tail). Furthermore, such cooperative games require that each player has a *theory of mind* about the other (e.g., the dog has a theory or expectation about how the human will behave when the dog drops the ball at the human's feet). In addition, each agent must *understand their own causal role and that of other agents*, enabling them to see, among the large set of possible actions, the sequence of actions

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

that will yield the cooperative payoff (e.g., the human understands if they pick up the ball and toss it, the dog will fetch it). Other examples of cooperative behaviors in both humans and nonhumans include group hunting, mutual grooming, mutual defense, shared tool use, and shared care for the young.

- *Collaboration* can then be thought of as a subset of cooperation that occurs when agents align behaviors in mutually beneficial ways, but where the structure of the game is not given and static; instead, *the players themselves are inventing, co-creating, and evolving the structure of the game over time*. Imagine, for example, a group of people who come together to create a new business. While such an activity certainly requires both coordination and cooperation, many of the rules and future payoffs of this game are unknown and perhaps unknowable (Knightian uncertainty). In fact, the rules of the game will be, at least partially, co-created by the players themselves as the game progresses and are likely to evolve over time. An activity such as building a business is not a single game; it involves *multiple, interlinked, repeated games and subgames*, both within the entrepreneurial group itself as well as through dynamic interactions with other games and players in the environment. Furthermore, the *causal links between player actions and future payoffs may be significantly separated by time and space*, *be noisy and complex, and it may be difficult or impossible to disentangle individual contributions to collective results*.

What I call collaboration is thus clearly more cognitively demanding than either coordination or playing simple cooperative games with fixed rules and clear payoffs. There is ample evidence of what I have called coordinating and cooperative behavior in many species (see Chapter 2, this volume). But what I call collaborative behavior appears at least unique to primates and possibly unique to humans (some might argue that nonhuman primate behavior is more correctly viewed as proto-collaborative rather than fully collaborative in the sense I have described). Clearly, what is uniquely human is our ability to engage in collaborative behaviors at scale with strangers. Chimpanzees, for example, cooperate (or possibly collaborate) in small groups of kin, near-kin, and known individuals in troupe sizes of 20 to 30. But humans can collaborate to build an Airbus A380 aircraft, assembled from four million parts and manufactured by tens of thousands of people in 1,500 companies from around 30 different countries.

Social Contracts as the Foundation of Large-Scale Collaboration

How then can agents align their actions in complex, dynamic settings with large groups of strangers and evolving, co-created rules with imperfect information

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

on the goals, contributions, and abilities of those strangers? Collaboration at scale requires agents to make a mental leap and see themselves as having collaborative relationships not just with other agents *individually* but with a *collective* of agents as an *entity*. As those entities may contain many agents not known to them, and the composition of agents may change over time, agents must be able to abstract the entities from the individuals who form them, such that the agent sees themself in a relationship with a collective entity (e.g., a tribe, firm, government, school, or a sports team).

We can define that relationship between the individual and the collective entity as a *social contract*. The idea of a social contract goes back to the Ancient Greeks, but modern discussions have their roots in concepts introduced by figures such as Thomas Hobbes, John Locke, and Jean-Jacques Rousseau, and then further developed in the twentieth century by John Rawls. Describing the relationship between an individual and a collective as a "contract" implies a mutuality of commitments: the individual voluntarily aligns their behaviors with the interests of the collective, agrees to contribute effort and resources toward collective goals, and submits to being governed by collectively enforced social arrangements in exchange for some set of future benefits. A core claim of social contract theory is that if an individual voluntarily submits to being governed by such collectively enforced social arrangements, then they must ipso facto view those arrangements as fair and legitimate, or at least "fair enough" (D'Agostino et al. 2021). In contrast, if people do not view the arrangements as fair, then they will either withdraw their cooperation or only submit to collective governance involuntarily.

To make our use of the term "social contract" more precise, we can think of it as follows: There is a "game" where if a group of players collaborates, they will potentially generate some nonzero-sum gains. An individual offers their collaboration to the collective group of players, conditioned on the following terms:

- I consent to play the game,
- I agree to play by the rules of the game, and
- I promise to play the game to the best of my abilities,
- *If* the game is fair.

The *social contract* thus defines the set of arrangements for that conditional offer of collaboration between the individual and the collective. The individual then makes judgments on the fairness of that contract based on their moral intuitions and cultural norms. Those feelings of fairness or unfairness, in turn, influence the agent's collaborative or noncollaborative behaviors.

As discussed above, collaborations involve a significant degree of uncertainty and imperfect information. In economic terms, this would imply that we cannot write a complete contract between the individual and the collective for the collaboration. Thus, an agent's doxastic representation will be incomplete. As such, the agent cannot simply make a self-interested rational choice as it would, say, in a cooperative game setting where the rules and payoffs are

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

known in advance, probabilities can be assessed, a complete contract written, and a rational choice can be made (Binmore 1994). I am thus hypothesizing that the evaluative criteria for the individual's commitment to a collaboration is *fairness*, which may incorporate aspects of self-interested rationality (e.g., I might not think the game is fair if the costs of my contributions outweigh the benefits) but involves a broader set of evaluative criteria (e.g., even if I receive net-positive payoffs, I might not think the game is fair if I am treated less well than others).

Equality, Process Fairness, and Deservedness

If fair social contracts are a necessary condition for effective, complex, large-scale collaborations, then the next question is: What do we mean by "fair"? This question has been widely explored philosophically, for example, by asking what kinds of social contracts would lead to a morally just distribution of resources, power, and rights in society (e.g., Rawls 1971), or what types of social contracts individuals would rationally choose to enter (e.g., Binmore 1994; Gauthier 1986). Here we will take a different approach (more Hume than Plato) and start from an empirical question: What are the characteristics of social contracts that most people are likely to perceive as fair?

This empirical perspective does *not* imply that people's individual moral intuitions about fairness will necessarily lead to social contracts that are just, from a societal or philosophical perspective. Nor does it imply that empirically observed moral intuitions about social contract fairness will be logically consistent, noncontradictory, or economically rational. Instead, I am making a simpler claim: *If individuals view their social contract arrangements to be fair, then they are more likely to engage in effective collaborative behaviors.* Therefore, in designing policies and institutions to maximize collaboration, it is useful to know what kind of arrangements are likely to be viewed as fair.

Human instincts about fairness appear to have deep evolutionary roots, and they likely evolved to facilitate cooperation and collaboration (Bowles and Gintis 2011; Gintis 2003, 2004, 2011; Gintis et al. 2008). Fairness instincts are found in nonhuman primates (Brosnan 2011, 2013) and appear early in child development (Gredebäck et al. 2015; Shaw et al. 2012). Feelings of fairness or unfairness also evoke distinctive neurophysiological responses, including the production of hormones associated with trust, pleasure, stress, or anger as well as heightened activity in the amygdala brain region (Chang et al. 2015; Crockett 2009; Haruno and Frith 2010; Tanaka et al. 2017). Additionally, there appears to be some genetic heritability in cooperative norms (Cesarini et al. 2008). Certain fairness norms appear to be highly universal, although their specifics may be more culturally variable. For example, an experimental study of resource sharing by children, conducted in seven diverse societies, demonstrated a universality of preferences for equal outcomes and rejection of

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

unequal outcomes that disadvantaged individuals (Blake et al. 2015). However, rejection of unequal outcomes that advantaged individuals was more culturally variable. Similarly, a large cross-cultural study of reciprocity norms showed high universality in the structure of those norms but with cultural variability in parameters for what specifically was considered fair, reciprocal behavior (Henrich et al. 2004).

One specific finding that is important for our purposes is that judgments about fairness are, to a significant extent, judgments about *process* fairness, and assessments of distributive outcomes are used as *signals* of whether a process is fair or unfair, based on a priori expectations of process outcomes (Starmans et al. 2017). Two examples illustrate this point: Imagine a group playing a coin-flipping game. Our reasoned expectation would be that if the game were played fairly, the outcome would be a roughly equal distribution of heads versus tails among participants. If, however, the outcome was significantly unequal, with, say, a large number of people flipping statistically unlikely, long streaks of heads, we would then suspect the game was not being played fairly—that they somehow cheated. So based on participants' understanding of the process, the expectation is an equal outcome, and unequal outcomes are a signal of potential process unfairness. Now imagine a second game, a 100-meter running race between a random group of people and Usain Bolt, the world record holder. Our a priori expectation would be that a fair race would yield an unequal outcome, with Bolt winning by a lot. If, on the other hand, the race yielded an equal outcome, with everyone crossing the line at the same time, we would suspect that something about the race was unfair—it was rigged. So, an equal or unequal outcome is not *inherently* fair or unfair but may instead be a signal as to whether a given *process* is fair or not.

In games involving distributions of resources, however, people express strong preferences for equal outcomes as a kind of default setting (Blake and McAuliffe 2011; McCrink et al. 2010; Shaw et al. 2012). Exceptions are made to the equality default rule based on perceptions of *deservedness* or *merit*, and unequal distributions may then be regarded as fair (and equal outcomes as unfair). For example, imagine a group of friends sitting around a table and one places a large cookie in the middle. How do they divide the cookie? The default answer would be equally; if one person grabbed more of the cookie, they would be viewed as greedy and their actions as unfair. We have strong instincts for relational fairness or more specifically, moral equality—the idea that we are each of equal worth and moral standing (Cook and Hegtvedt 1983; Killen et al. 2001; Konstantareas and Desbois 2001). Intuitions about moral equality arguably appear in primates (Brosnan 2011, 2013), develop in young children (LoBue et al. 2011), appear across cultures (Kim and Leung 2007), and are a central idea in many religions (i.e., all are equal before God). The intuition for moral equality tells us that, in the absence of any other information, we each deserve an equal share of the cookie. Now imagine it turns out that one person just returned from a long run; the others might view her as

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

deserving more of the cookie. Or one of the group members had lost his job and to cheer him up might need more of the cookie. Or someone starts acting in an objectionable way and the group expresses its displeasure by saying, "you don't deserve any of the cookie!" While equal distribution may be the default rule, we also have instincts for *deservedness*, for merit-based exceptions to the equality default. A fair process also takes account of information on differences in circumstance, merit, luck, and the nature of the game being played, to adjust the outcome based on such factors, ending in a result where everyone "gets what they deserve." Denying these merit-based exceptions to the equality default rule would itself likely be viewed as unfair. In a small group of people personally known to each other (as in our cookie example), such a fair process may be quite informal, although still with potential for contested views as to what counts as "deserving" and what that implies for allocation. In a large group where people are not personally known to each other, information is imperfect and behavioral monitoring is limited; challenges become significant, and confidence in process fairness becomes critical.

Nine Dimensions of Fair Social Contracts

We can use these findings to construct a simple framework for assessing social contract fairness. We have preferences for *relational fairness*, which includes the principle of moral equality as a precondition (e.g., it is hard to have a fair process with unfair power relations); for *procedural fairness*, which includes the principle of deservedness or merit; and for *distributional fairness*, which relates perceived outcomes to expected outcomes to make assessments about the fairness of the game. Building on these general underlying preferences, we can ask: What are the specific attributes of social contracts that are likely to be viewed as fair?

Table 11.1 summarizes the nine attributes that I propose contribute to perceptions of social contract fairness. A brief description is given for each dimension and their supporting evidence. I have phrased these as "I" statements as they are from the perspective of the individual agent facing the collective.

Relational Fairness

1. *Agency: I can choose to play the game and have choices within the game.* If I am forced to play the game (e.g., a slave), I am unlikely to view the social contract as fair. Likewise, if I enter the game but all choices are made for me (particularly if I cannot predict the outcomes from such involuntary choices), I am unlikely to view the contract as fair. One can think of agency as an aspect of relational fairness, as it answers the question of who has the power to make decisions that affect an individual. The literature shows that agency is critical to healthy

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

Table 11.1 Summary of the proposed nine dimensions of fair social contracts.

Underlying Moral Preferences	Dimensions of Fair Social Contracts	Description
Relational fairness	1. Agency	I can choose to play the game and have choices within the game.
	2. Inclusion	I have an opportunity to play the game. I am not excluded.
	3. Dignity	If I play by the rules and contribute to the best of my abilities, I will be valued, respected, and have status.
Procedural fairness	4. Rules-based	I know the rules of the game and they are applied equally to everyone.
	5. Meritocratic	I, and everyone else, will receive rewards and punishments in the game based on merit.
	6. Security	If I play by the rules and contribute to the game, but suffer misfortune through no fault of my own, I will be protected.
Distributional fairness	7. Capabilities	I have the capabilities to play the game or the opportunity to acquire them.
	8. Reciprocity	If I play by the rules and contribute, others will reciprocate, and I will share in the game's rewards.
	9. Progress	If I play by the rules and contribute to the best of my abilities, my life and the lives of those I care about will improve.

human functioning and sense of identity, motivation, and engendering cooperative behaviors (Akbas et al. 2019; Bandura 1997; Bandura 2006; Ryan and Deci 2000). In economic experiments, subjects valued agency to be a key element in determining the fairness of the game (Akbas et al. 2019; Konow 2000). One may be able to create a kind of large-scale forced coordination with an army of slaves, but it is not possible to create true collaboration capable of solving complex problems (e.g., an army of slaves could not develop a novel vaccine). There are, of course, degrees of agency. People may have choices of employment but still must work to make a living or people may have democratic political choices but still must obey the law regardless of who wins an election. Nonetheless, an ability to make choices, within a set of options limited by agreed rules, remains a critical component of both fairness and effective collaboration.

2. *Inclusion: I have an opportunity to play the game, I am not excluded*. If one chooses to play the game but is excluded for unjust (i.e., not based on merit) reasons, one is likely to view the game as unfair.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

An obvious example is the long history of economic, political, and social exclusion for reasons of race, gender, religion, ethnicity, class, or sexual preference or identity. Inclusion is an aspect of relational fairness in that unjust exclusion violates the principle of moral equality. There is significant social psychology evidence on the detrimental effects of exclusion on subjective well-being (Bellani and D'Ambrosio 2011; Gross-Manos 2017). Furthermore, unjust exclusion of others appears to trigger people's sense of fairness and prompt them to action (MacDonald and Leary 2005; Moor et al. 2012; Tuscherer et al. 2016; Williams 2007). There is a link to procedural fairness, as non-merit-based exclusion triggers feelings of unfairness. When, for example, Jackie Robinson was excluded from Major League Baseball simply because of the color of his skin, that was widely viewed as deeply unfair. If, however, a middle-aged professor were to be excluded from Major League Baseball because of his terrible performance, that would be fair, particularly if (see Pt. 7 below) he was previously given access to acquiring capabilities (e.g., opportunities to play Little League), and the process for judging players is meritocratic (see Pt. 5 below).

3. *Dignity: If I play by the rules and contribute to the best of my abilities, I will be valued, respected, and have status*. Humans are status-conscious and status-seeking creatures. Status and dignity evoke strong emotions tied to feelings of fairness (Folger and Cropanzano 2001; Stets 2004). Feeling like a valued contributor to the collective is a powerful motivating force in collaborative behavior and a critical element in forging a common identity with the collective (Axelrod and Hamilton 1981; Fox and Guyer 1978). Violations of dignity (i.e., feeling underappreciated or disrespected) can evoke strong negative emotions and feelings of injustice and lead to behaviors detrimental to collaboration (Greenberg 1988; Milinski et al. 2002). Affording people dignity recognizes their worth and standing and is thus an aspect of relational fairness.

Procedural Fairness

4. *Rules-based: I know the rules of the game and they are equally applied to everyone*. This attribute bridges preferences for relational fairness and process fairness. If everyone is of equal moral worth, then the rules of the game must apply equally to everyone, and a fair process is one in which the rules are known, followed, and equally enforced. There is evidence from cognitive science and social psychology that people have strong preferences for such procedural fairness (Engel 2005; Folger 1986; Folger and Cropanzano 2001; Greenberg 1987, 1990; Henrich et al. 2010a; Marwell and Ames 1981). People's degree of association between respecting the rules and fairness varies by

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

culture (e.g., cross-cultural studies of tax compliance; Cummings et al. 2005), but the idea that the same set of rules should apply to everyone (even if not always complied with) does appear to be widely held. Situations where rules are unevenly applied, manipulated, or ignored by privileged groups, or opaque and subject to arbitrary interpretation or enforcement, are widely viewed as unfair.

5. *Meritocratic: I, and everyone else, will receive rewards and punishments in the game based on merit.* People appear to have intuitive notions of merit and deservedness. Rewards should go to those, for instance, who contribute to the collective effort, engage in reciprocal behaviors, have relevant capabilities, play by the rules, and are of good character (Adams 1965; Baumard et al. 2012; Cohn et al. 2011; Kulik and Ambrose 1992). While Sandel (2020) argues that meritocracy can lead to excessive individualism, reinforce inequities, and harm collective endeavors, people tend intuitively to see meritocratic processes as providing a basis for distributive justice. For example, most people would see a university admission process based on some notion of merit (e.g., student academic achievement, potential to contribute to the student body) as fairer than one based on non-meritorious criteria (e.g., a parent's donations to the university). What constitutes "merit" is, however, highly contestable and context-dependent (e.g., some might argue that conventional measures of academic merit favor students born to wealthy parents who can afford private schools). Nevertheless, most people view a meritorious process as more likely to lead to distributional fairness.
6. *Security: If I play by the rules and contribute to the game, but suffer misfortune through no fault of my own, I will be protected.* There appears to be widely shared instincts for luck egalitarianism, the recognition that bad luck can strike any of us for reasons not of our own making (Anderson 1999; Dworkin 1981; Nagel 1979; Tinghög et al. 2017). One might get cancer, be laid off in a recession, or face hunger in a drought. While we cannot protect against all unlucky situations, humans have strong empathetic instincts in such situations and are often willing to act charitably and altruistically (Boyd and Richerson 2005; Dovidio 1984; Fehr et al. 2008; Masten et al. 2010; Pavey et al. 2011; Zak 2011). Furthermore, there are strong instincts for mutual protection of fellow members in one's group, particularly if the unlucky individual is seen as a contributor to the group's welfare. However, there are sensitivities to the potential for free riding and abuse of empathetic feelings. Thus, government social safety net programs tend to have higher political support when they insure against bad luck that could strike anyone, require reciprocity, and monitor against abuse (Batson et al. 2007; Fehr et al. 2002; Fehr and Gachter 2000; Fong et al. 2006; Sasaki and Uchida 2013).

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

Distributional Fairness

7. *Capabilities: I have the capabilities, or opportunities to acquire the capabilities, to play the game*. A fair game requires the capabilities to play. As Amartya Sen has argued, positive freedom requires capabilities to provide the functionings necessary for a fulfilling life (Sen 1985, 2008). This is particularly important for games we play out of necessity, notably the "earn a living game." Yet, there is a birth lottery in the distribution of capabilities (e.g., you might be born to a poor family or where a good education is not available). So distributional fairness requires that people have opportunities to acquire capabilities and fulfill their potential. Likewise, it is unfair to expect people to play a game for which they do not have the capabilities or do not have the opportunities to acquire them. For example, systematic underinvestment in female education violates distributional fairness (Nussbaum 2002, 2003; Robeyns 2006; Sen 2008). While I am not aware of literature that provides direct evidence of people's perception of capabilities as an attribute of fairness, there has been work in psychology connecting capabilities to feelings of well-being (Jayawickreme and Pawelski 2013) and on capabilities as a basis for agency and empowerment (Shinn 2015). One can hypothesize that a social contract that requires certain actions or behaviors but does not provide the capabilities to fulfill those expectations would be generally regarded as unfair.
8. *Reciprocity: If I play by the rules and contribute, others will reciprocate, and I will share in the game's rewards*. Reciprocity can be categorized as a form of distributive fairness, as an observable and expected outcome in a fair process. If the process is fair, I will observe reciprocal behaviors in the contributions and the sharing of rewards between players (e.g., players only have information on their own contributions and observations of distributive outcomes; Guth and Tietz 1990). Intuitions and norms of reciprocity develop in early childhood (House et al. 2013; van den Bos et al. 2010; Warneken and Tomasello 2013), appear across cultures (Chen et al. 2009; Kuwabara et al. 2007), and are foundational to establishing cooperation and collaboration (Adams 1965; Axelrod and Hamilton 1981; Bowles and Gintis 2011; Greenberg 1990; Trivers 1971). Furthermore, evidence shows that when reciprocity norms are violated, agents not only withdraw from cooperation, but they also punish the individuals or institutions that have violated their expectations of fairness (Adams 1965; Axelrod and Hamilton 1981; Bowles and Gintis 2011; Greenberg 1990; Trivers 1971). The literature further shows that such punishment may even be to the punisher's detriment, altruistically bearing a cost to enforce norms of reciprocity.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

9. *Progress: If I play by the rules and contribute to the game, my life, and the lives of those I care about, will improve*. Progress can be thought of as a form of distributional fairness over time. In economic reality, and in its perception as either a moral good or right, progress is a phenomenon and concept that appears to have developed in certain societies during the seventeenth to nineteenth centuries (Maddison 2007; Wootton 2018). It is not clear that similar notions exist in traditional societies, and it has historically been viewed differently in many non-Western cultures. Nonetheless, today the "right to progress" has become a widely held idea across the globe (Alesina et al. 2004; Day and Fiske 2017; Rodon and Sanjaume-Calvet 2020; Wegener 1991). Furthermore, expectations of progress in one's life, and emotions related to hope for the future, are strongly associated with subjective well-being (Pleeging et al. 2021).

Discussion

Another way to consider the impact of these nine dimensions on perceptions of fairness is to think of a social contract with the opposite characteristics. Imagine being offered a social contract to play a game where the following is true:

Relational unfairness

1. You do not have agency to make choices
2. You are excluded from critical aspects of the game
3. You will not be respected for your role and contributions

Procedural unfairness

4. You do not know the rules and/or they are unequally applied
5. You and others will not receive rewards and punishments based on merit
6. You will not be protected from misfortune

Distributional unfairness

7. You do not have the capabilities necessary to play successfully nor opportunity to acquire them
8. You are not reciprocally rewarded for your contributions
9. And, finally, even if you play and contribute to the best of your abilities, your life and those you care about will not improve

Would this be a fair game? Would you accept a social contract to play it? Probably not. Would anyone *voluntarily* agree to play such a game? It is highly unlikely.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

The next question then is: If even *one* of the above negative statements is true, would you regard the contract to play the game as fair? My hypothesis is that if even one of these negative statements is true, then that would be sufficient to make the game unfair for most people. This in effect means that the nine dimensions are distinct and non-substitutable. I do not mean to imply that in the real world people cannot or do not make trade-offs across the attributes—they can and do. Instead, what I am proposing is that *all nine are necessary, to at least some degree*, for a contract to be viewed as fair. A zero value for any of the nine will trigger moral intuitions of unfairness. For example, even if a social contract is very high on meritocracy, that is no substitute for not having capabilities. Nor will investing in more capabilities make up for being excluded.

How universal are these nine dimensions? Moral psychology researchers have observed a high degree of universality in moral intuitions, social-emotional responses, and neural-cognitive patterns as well as significant individual and cultural variability in how people weigh, trade-off, and interpret moral preferences (Crockett et al. 2014; Gintis et al. 2008; Greene et al. 2004; Molnar-Szakacs 2011; Shenhav and Greene 2010; Singer 2005; Zak 2011). Jonathan Haidt (2012) likens findings on the universality of dimensions of moral preferences to "taste buds." Every human has the same five taste receptors (sweet, sour, salty, bitter, and umami), but individuals and cultures vary in their preferences as to how these universal tastes are combined in specific foods. Likewise, in the case of the above nine dimensions, I would propose that they are highly universal and applicable across cultures, but individuals and cultures will vary in their preferences for how they are interpreted, weighted, and traded off in specific social contracts.

It is important to note that, despite the universality of dimensions, different interpretations and weightings can nonetheless result in highly contested views as to what specifically constitutes a fair contract. Examples include people with differing political views interpreting the provision of access to capabilities in different ways or debating how much security is "enough" in the welfare state. Or one branch of a religion that interprets sacred texts as justifying the exclusion of women from education or certain occupations (i.e., adherents may view these texts as providing a merit-based justification for exclusion, the "merit" being "God says so") might be in conflict with another branch of the same religion that interprets the texts as promoting moral equality of both women and men and therefore inclusion. Again, my claim is *not* that there is universality to the *specific* social contracts that people perceive as a fair or unfair, but rather that there is universality to the *evaluative framework* people use when making such judgments.

Social Contract Violation and Political Populism

This universality of an evaluative framework offers insights into how and why collaboration breaks down, and why the fairness of specific social contracts

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

may be contested. Here, I will apply the nine dimensions to briefly analyze the breakdown of political collaboration in the United States. I argue that a major deterioration in the fairness of social contracts in the United States from the 1970s to 2010s led to widespread perceptions of contract violation. This, in turn, laid the emotional foundations for a drop in political collaboration and rise in political populism.

Over the past decades, the United States and various other countries have witnessed a breakdown in political collaboration, increased polarization, a loss of faith in democracy, a loss of trust in key institutions, and a rise of populist and authoritarian political figures (Hawkins et al. 2019b; Edelman Trust Barometer 2022; Pew Research Center 2016). Using a variety of metrics, Putnam and Garrett (2020) identify the late 1960s to early 1970s as a peak in U.S. social, cultural, and political cohesion. By 2015, this cohesion had deteriorated to levels not seen since the Civil War. A variety of explanations have been put forward to explain this broad trend, including increases in economic inequality, economically "left behind" regions, cultural and demographic factors, and changes in the media landscape. Surveys and studies of recent election results find, however, that instead of material explanations (e.g., economic, education, demographics), the most explanatory variables are attitudinal and emotional (Cox et al. 2017; Green and McElwee 2018; Hawkins et al. 2019b; Inglehart and Norris 2016; Mutz 2018; Ward et al. 2020). Notably, voters who support populist candidates report feelings of a loss of agency over their lives and communities (e.g., the Brexit slogan, "take back control"), alienation and exclusion from the broader culture (e.g., perceptions that their racial, ethnic, cultural, or religious group is becoming a "persecuted minority"), a loss of reciprocity (e.g., the sentiment, "we pay our taxes while others get benefits"), a view that powerful elites are playing by different rules (e.g., "the game is rigged"), significant fears of status loss, and a loss of feelings of security and hope for the future. Such attitudes align very closely with a negation of the attributes of fair social contracts and indicate significant feelings of contract violation (Table 11.2).

Such feelings of social contract violation have become widespread but have been most heavily concentrated over the past decades in two broad groupings. The first group consists of white, working-class, largely Christian, largely male, noncollege-educated ex-urban voters. For many of these these voters, the "others" who have violated the contract are people of different political beliefs, racial groups, religions, immigrants, and gender identities, as well as foreign countries (Silver et al. 2021). The overall feeling of these voters is that their own group has worked hard, contributed to society, and played by the rules but has lost opportunities and status because of unfair play by the "others." Furthermore, the rule-setters and referees who are supposed to ensure a fair game—the "cultural elite" of political, business, media, and academic leaders—have not only allowed the contract to be broken, but are perceived to have been complicit in breaking the contract to serve their own interests.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

Table 11.2 Attitudes of supporters of populist political candidates and causes align closely with feelings of social contract violation.

Unfair Game/Broken Contract	Sample Attitudes
Loss of agency	Others are controlling our lives. We need to take back control.
Exclusion	My group is being discriminated against and excluded from opportunities.
Loss of dignity, status	People like me used to be valued members of society. Now we are not.
Rules violations	The game is rigged. Powerful people and favored groups play by different rules.
Less meritocratic	I work hard but cannot seem to get ahead while less deserving people do.
Decreasing security	I worry about my finances, health, retirement, crime, and our nation's security.
Insufficient capabilities	I have worked hard all my life but my skills are no longer valued.
Loss of reciprocity	I work hard and deserve what I get but others do not and get a free ride.
Loss of progress, hope	Things are getting worse not better. I fear for my children's future.

Thus, perceptions (and misperceptions) of contract violation have contributed to increases in racism, anti-immigrant sentiment, misogyny, and an anti-elite backlash. Right-wing political figures, parties, and media were the first to notice these growing sentiments in the 2000s, when they began to exploit them (e.g., Pat Buchanan's 2000 presidential campaign), creating a major political realignment that shifted many white, working-class voters from left-leaning to right-leaning political parties in the United States and Europe. This led to a dramatic rise in right-wing populism, exemplified by Brexit and the 2016 election of Donald Trump.

Whereas white, working-class voters drove the rise in right-wing populism, they were not the only ones who felt a sense of social contract violation. The second broad group with such feelings (Table 11.2) includes struggling lower-income families, citizens in deprived urban communities, people from historically excluded racial, religious, and gender groups, and young people who fear for their future. For this group, the "others" violating the social contract include billionaires who do not pay their taxes, large corporations who exploit workers and profit at the expense of others, "privileged" groups (i.e., white males) who benefit from historical injustices, as well as a political class that rigs the game. These voters have aligned with left-wing populists such as Bernie Sanders in the United States and Jeremy Corbyn in Britain.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

Preceding the 2016 populist wave, it is notable that the 2008 financial crisis resulted in angry, grassroots populist movements on both the political right (e.g., the Tea Party movement) and the political left (e.g., the Occupy Wall Street movement), both of which articulated broken contract narratives but differed as to who was doing the violating. While the politics and policies of the right-wing and left-wing populists differ starkly—and the racism and sexism of certain right-wing populist figures and some segments of their supporters must be condemned—the *emotional structure* of popular support has been similar on both sides, founded on feelings of moral outrage over a broken social contract.

These feelings of a broken contract among large segments of voters are in many cases justified. Although material explanations may not be directly causal in explaining the populist rise, underneath each of these attitudinal dimensions are changes and trends in the structure of the economy and society that have arguably provoked these feelings. Putnam and Garrett (2020) identify the late-1960s to early 1970s as the turning point in U.S. social cohesion. Beginning in the mid-1970s, the structure of the U.S. economy underwent a profound shift: productivity growth and worker income growth decoupled, incomes for about 90% of households stagnated in real terms while almost all the gains of growth flowed to the top 1% of earners, the middle-class shrank as a percentage of the population, social mobility declined, and various measures of economic insecurity increased.[1] While technological change and globalization contributed to these trends, particularly from the 1990s onward, cross-country studies suggest that much of this change resulted from shifts in economic ideology and policy that began earlier, in the 1970s and 1980s (Nolan 2018). A shift toward more so-called "neoliberal" economic policies, both in right-wing (e.g., Reagan, Thatcher) and left-wing (e.g., Clinton, Blair) governments, resulted in

- shifting the tax burden away from the wealthiest individuals and corporations to middle- and lower-income workers,
- relative reductions in public investment (e.g., education, infrastructure),
- weakening of the social safety net,
- changes to labor market regulations that reduced union and worker power,
- central bank policies that prioritized low inflation over employment and wage growth, and
- trade policies that favored corporate over worker interests.

At the same time, changes in corporate practices (e.g., moving from balanced stakeholder to shareholder value-maximizing governance, outsourcing, offshoring, reductions in pension and health benefits, less secure employment)

[1] Research on these various trends and the impacts of neoliberal policies is too voluminous to cite individually. For various studies and data sources see, e.g., the University California Berkeley Center for Equitable Growth (http://ceg.berkeley.edu/index.html), Washington Center for Equitable Growth (https://equitablegrowth.org/), the Economic Policy Institute (https://www.epi.org/), and World Inequality Database (https://wid.world/).

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

shifted gains in productivity away from workers and toward shareholders, while reducing worker power and security (Lazonick and O'Sullivan 2000). These economic changes coincided with a growing influence of money in U.S. politics and various failed attempts to regulate it in the 1980s–2000s (culminating in the Supreme Court's 2010 Citizen's United decision), as well as increasingly effective gerrymandering of congressional and state legislative districts, and demographic shifts that make the U.S. Senate less representative of the population (Mansbridge 1999; Smith 1995; Teachout 2016). Together, this has made the U.S. democratic system less responsive to citizen concerns and more responsive to those of well-funded interests (Lindsey and Teles 2017).

At the same time, major social, cultural, and demographic shifts were underway:

- The civil rights and feminist movements of the 1950s–1960s empowered historically excluded and underrepresented groups.
- Demographic shifts saw the nonwhite population in the United States grow from 20% in 1980 to 40% by 2019 (Frey 2020).
- The immigrant population of the United States tripled during this period from 4.8% in 1970 to 13.7% in 2020 (Budiman 2020).

Furthermore, there were significant population movements from rural to urban and suburban, and from rustbelt regions to the sunbelt. While studies show that voters with racist, anti-immigrant, and sexist attitudes were a significant factor in Donald Trump's election (Bock et al. 2017; Cassese and Barnes 2019; DeSante and Smith 2020; Hooghe and Dassonneville 2018), a significant numbers of voters were expressing feelings of economic injustice and insecurity, status loss, cultural disorientation, resentment of elites, and perceptions of a corrupt and unresponsive political system. Again, such feelings were not limited to voters traditionally on the political right.

Table 11.3 shows how these dimensions of economic, political, and social change map onto the attributes of fair social contracts. For a very large numbers of citizens, there was thus a factual basis to perceptions of a deteriorating social contract.

This leads to a key conclusion: the United States and other similarly affected countries cannot heal political divisions, renew faith in democracy, and reinvigorate collaboration at a national scale, *unless the social contract is restored*. A key aspect of the psychology of broken contracts is that feelings of contract violation must first be acknowledged and empathized with before people are willing to listen and engage in contract reconstruction. Populist candidates have succeeded electorally by giving voice to resultant feelings of moral outrage (Brady et al. 2018; Crockett 2017), promising to fix the violation ("only I can fix it"), and contrasting themselves with "out of touch elites" who "don't get it." It is essential that political and other leaders, who *genuinely* want to restore the social contract, acknowledge first that the social contract has been broken. They must demonstrate that they hear and empathize with the resulting

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

Table 11.3 Each dimension of the U.S. social contract was broken or weakened for a large proportion of the population during the 1970s–2010s.

Unfair Game/ Broken Contract	Sample Trends
Loss of agency	Loss of worker power, de-unionization; loss of local government autonomy, more centralized, less responsive political power
Exclusion	Racial and gender barriers, demographic change, cultural alienation, identity politics
Loss of dignity, status	Perceived relative status loss (especially for white, working-class, males)
Rules violations	Different rules for rich and powerful (e.g., corporate behavior, political capture)
Less meritocratic	Lower social mobility, "opportunity hoarding" by top 5%
Decreasing security	Weakening social safety net (public and private), greater downward mobility
Insufficient capabilities	Declining public investments, education quality, worker training
Loss of reciprocity	Decoupling of wages and productivity growth; declining tax fairness
Loss of progress, hope	Wage stagnation, declining optimism for future generations

emotions and then provide specific solutions to restore the contract that map onto the nine attributes of contract fairness. If this hypothesis is correct—that the attributes of fair social contracts have high universality—then those nine attributes can provide a template for reducing social divisions and increasing collaboration by pointing to areas of broad agreement on goals while allowing debate on specific policies. For example, restoring perceptions of reciprocity could be aided by both increased tax fairness (a traditional cause of the left) and welfare system reform (a traditional cause of the right). Increasing agency could be helped both by increasing worker power (a traditional cause of the left) and devolving central political power (a traditional cause of the right). Greater agreement on ends (a fairer social contract) and more constructive debates on means (specific policies) could help facilitate the return of a more functional politics.

Conclusions and Directions for Future Research

Social contract fairness is foundational to large-scale collaboration, and social contract unfairness is a key factor in collaboration breakdown. Understanding what leads to perceptions of social contract fairness is amenable to empirical study. This chapter has presented an empirically informed hypothesis that there are nine universal dimensions to social contract fairness. This hypothesis

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144

is testable. Tools and methods from various disciplines could be brought to bear to prove, disprove, or modify this hypothesis. For example, behavioral experiments could test the willingness of players to collaborate in games that varied in design along the nine dimensions. Other experiments could test the substitutability of the dimensions, as well as individual preference weightings. Sociological surveys could be used to test perceptions of fairness or unfairness against the dimensions and their universality or variation across individuals and cultures. Organizational studies or anthropological observations could seek to document and analyze social contract designs "in the wild," assessing participant perceptions of fairness against the dimensions. Finally, it would even be possible to imagine field experiments, where social contract terms are varied for different groups to assess the impacts on perceptions, collaborative behaviors, and outcomes.

Findings from such work could yield prescriptive insights for identifying risks for social contract breakdown. The example of U.S. political polarization illustrates the stakes involved when social contract fairness in societal-scale collaborations is allowed to degenerate. Practical insights and strategies are needed for social contract repair in many contexts—only through collaboration can we solve our greatest challenges.

Acknowledgments

The author acknowledges the Institute for New Economic Thinking and the Open Society Foundations for funding support and would like to thank participants at The New Institute, Foundations of Value and Values Symposium, Hamburg, Germany, April 2022, as well as the Ernst Strüngmann Forum, How Collaboration Arises and Why it Fails, Frankfurt, Germany, May 2022, for their valuable feedback. Finally, the author thanks Fiammetta Brugo and Jan Meyerhoff for research assistance.

From "The Nature and Dynamics of Collaboration,"
edited by Paul F. M. J. Verschure et al. Strüngmann Forum Reports, vol. 33,
Julia R. Lupp, series editor. Cambridge, MA: MIT Press. ISBN 9780262548144