

How Collaboration Breaks Down

Simon DeDeo, Jenna Bednar, Eric D. Beinhocker,
Esther Chevrot-Bianco, L. Zachary DuBois,
Dorthe Døjbak Håkonsson, Jônatas Manzolli,
and Ferdinand von Siemens

Abstract

Humans may be “super-cooperators,” but no collaboration lasts forever. This chapter summarizes the outcome of an interdisciplinary collaboration between political, social, economic, and cognitive scientists into the question of collaboration collapse. It locates the breakdown of collaboration downstream from the failure to align on either values or actions. A fourfold taxonomy is presented of the consequence of these failures: catastrophic collapse, generative reboot, contested persistence, and sputter on launch. Each failure mode is illustrated by case studies (e.g., the breakup of the Beatles, the collapse of the Hawai‘ian *kapu* system, the failure of the Kyoto Protocol, kinship taxation, resistance to antibiotics) to demonstrate how general principles of our taxonomy unfold over a range of historical, political, and economic contexts. Understanding the mechanisms that underpin successful collaborations and the taxonomies of dysfunction might inform efforts in pursuit of stable collaboration and enable interventions that do not disrupt or enfeeble alignment mechanisms.

Introduction

Turning and turning in the widening gyre
The falcon cannot hear the falconer;
Things fall apart; the centre cannot hold;
Mere anarchy is loosed upon the world.

—W. B. Yeats (1920)

The heartbreak of a failed collaboration can be hard to forget. The emotional engagement of individuals in a collaboration and the feelings involved when things break down are often intense. Because we rely on collaborations for everything, from the most rarefied artistic and scientific endeavors to finding

food in an emergency, a breakdown may do more than emotional damage: it may endanger our flourishing or even our survival. Yeats's poem, written in the aftermath of the collapse of European society prior to World War II, captures both the emotional and material devastation of the "mere anarchy" that follows a breakdown.

Every collaboration is a unique combination of cooperation and coordination. For our purposes here—and drawing on a long tradition in organizational design that dates back to Nadler and Tushman (1980)—we define "cooperation" as an *alignment of values* that can range from agreement on immediate goals to the sympathetic sharing of enduring moral and ethical principles. We define "coordination" as an *alignment of actions*; that is, the more or less efficient combination of resources and capabilities that can involve everything from the simultaneous entrance of two voices in a choir to the complex synchronization of a vaccine supply chain.

With these definitions in hand, this chapter unfolds a basic insight of organization design: that all breakdowns in collaboration are causally downstream of *failures to align* on these two distinct axes of values or action (Figure 18.1). A collaboration may involve a multitude of imperfections, self-interest, and sins. It may even be literally dedicated to sin, but unless, and until, these antisocial forces combine, perhaps with external pressures, to produce misalignment in either values or actions, the collaboration can persist and even thrive. Perfect alignment may be neither possible nor desirable, and misalignments may be papered over or lie latent for decades, but significant levels of misalignment, along one or both of these axes, provide the causal mechanism of breakdown.

Cooperation failures and coordination failures can arise independently. A group may be well aligned in what it values but unable to coordinate successfully: a group of well-meaning restaurant owners may want to salvage waste food for a local homeless shelter but be unable to coordinate their kitchens at the end of a chaotic day. Other groups may have highly efficient coordination mechanisms but be unable to agree on values: a factory may have the capacity to coordinate a fantastically complicated series of tasks, but if workers and management have different notions of fair compensation, it may lead to a strike.

To some, Figure 18.1 may seem too complex, with a seemingly indeterminate number of causes that can lead to misalignment. As we show in our series of case studies, however, it appears that, while the causal structure varies from one case to the next, the structure is often relatively sparse and relatively fixed, in any particular case. Much of the complexity is explicable in terms of the emergence of feedback loops.

To others, Figure 18.1 may seem too simple: surely breakdown can have more than two possible proximate causes? The breakdown of collaboration, however, is best separated (at least conceptually) into a "before," when the precursors to misalignment grow, and an "after," when misalignment has become so severe that people no longer want or are able to collaborate under the same

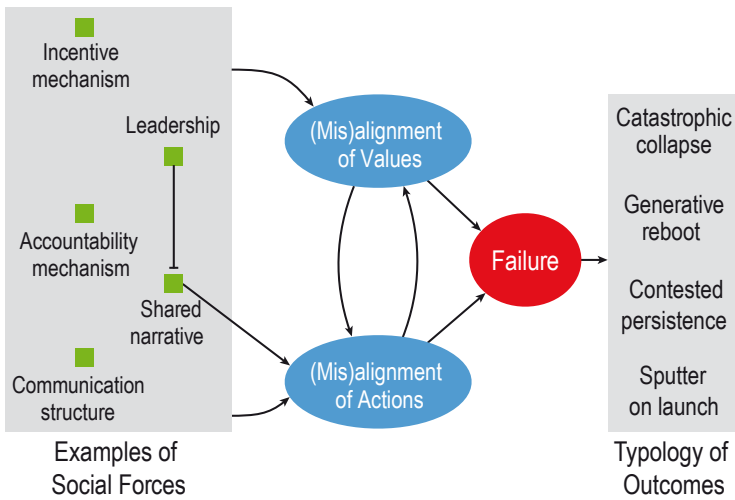


Figure 18.1 The causal topology of collaboration breakdown. The complexity of human cognition and culture means that an enormous variety of causes interact to both sustain and interfere with collaboration (the alignment of values) and coordination (the alignment of actions). Misalignment of values and actions may arise separately: in this case, we have indicated the particular causal pathway for the catastrophic collapse of the Hawai‘ian *kapu* system, where elites acted to undermine a shared narrative (i.e., the *kapu* system was backed by divine punishment) by publicly holding a mixed-gender feast. The destruction of this shared narrative led to novel misalignment of actions within the aristocracy and between the aristocracy and the people. This led to the catastrophic collapse of ancient Hawai‘ian society and its replacement by a new, centralized monarchy. In other cases, misalignment of values and actions may have a hidden common cause. The emergence of misalignment in one leads to feedback effects on the other, either directly (e.g., failure to coordinate may lead to the divergence of values among the collaborators) or via social forces (misalignment of values may cause a legitimacy crisis that makes it impossible for a leader to correctly coordinate action).

terms. What happens afterward is of great interest, and we identify three major classes of breakdown (catastrophic collapse, generative reboot, contested persistence) along with an edge case, sputter on launch. These are introduced through a series of brief case studies, as outlined in Table 18.1.

We begin with a discussion of the key aspects of any collaboration analysis, where our focus is on the aspects of collaboration that can lead to failure. Next, we present the four forms of collaborative failure through a series of examples from different scholarly fields that span a range of scales, historical periods, and sociotechnical contexts. In our discussion of the Franklin Gap, we discuss how a source of robustness can also serve as a source of failure. This provides a cautionary note against thinking too simplistically about the underlying sources of collaboration failure. We conclude with a brief discussion and a novel policy recommendation: alignment review.

Table 18.1 A taxonomy of breakdown by failure type, with samples of each and the type of mismatch created.

Failure Type	Example	Mismatch Type
Catastrophic collapse	Breakup of the Beatles	Values
	Collapse of the Hawai'ian <i>kapu</i>	Actions
Generative reboot	Copenhagen Climate Conference	Values
	Scaling Wikipedia	Actions
Contested persistence	Kinship tax	Values
	Fencing among the Maasai	Actions
Sputter on launch	Antibiotic resistance	Actions

The Fragile Architecture of Collaboration

Humans are not the only species to collaborate. Honeybees, for example, collaborate in hives; as eusocial insects, their goals are aligned through genetic similarity and kin selection. Trout appear to collaborate in hunting (Vail et al. 2014), and crows collaborate to mob predators (Jelbert et al. 2015). Dominance hierarchies that enable complex coordination in the presence of otherwise debilitating conflicts (DeDeo and Hobson 2021) are found throughout the animal kingdom, from ants to elephants (Hobson et al. 2021).

Many of these cases can be understood in the axiomatic language of game theory. For *Homo economicus*, one-shot cooperative games with aligned incentives often require only the recognition of mutual interest and the ability to spot strictly dominated strategies. Once the payoff structure aligns the value, robust collaborations emerge as a matter of course.

What draws our attention as students of human nature, however, are the often far more sophisticated collaborations enabled by capacities well beyond what we might ascribe to trout, crows, or *H. economicus*: collaborations that span multiple levels of organization and generations of participants, and can be adaptive and resilient in the presence of diversity and changing contexts. In many cases, human abilities not only make new forms of collaboration possible, they also make older forms newly impossible. For example Gintis et al. (2019) hypothesize that the combination of social interdependence and the invention of lethal weapons is the causal origin of zoon politikon, the complex political life of humans. It is zoon politikon, and its failures, that concern us here.

The Gadgets of Collaboration

At the cognitive level, the “gadgets” that underlie collaboration include counterfactual thought and mental simulation, mental time travel, theory of mind, language, and the mental abilities necessary to support joint intentionality and cultural norms (Dor et al. 2019; Heyes 2018; Tomasello et al. 2005). At the

social-institutional level, these include communication systems, repositories of cultural memory, and epistemic institutions like public argument-making, debt, and markets (Bellah and Joas 2012; Boy and Torpey 2013; Graeber 2012). Four key epistemic tasks appear to be in play when we seek to initiate and maintain a collaboration:

1. We have to imagine a future we want and to share that imagination with other agents. (We have to be able to set goals and to share those representations with others.)
2. We have to foresee how collaboration can make real that desired future and again share that perception. (We have to understand the structure of potential collaboration and its benefits.)
3. We have to predict the behaviors of others regarding their willingness and abilities to collaborate, how they might collaborate, and whether they can be trusted. (We have to understand the values and capacities of others.)
4. We have to manage the risks associated with uncertainty due to the temporal and spatial extent of a collaboration. When rewards to oneself depend on the actions of others, for example, collaboration may require that participants contribute without specific knowledge of reward but instead with a general expectation of reciprocity. When rewards to others are uncertain, we have to rely on proxies and signals (e.g., market prices) to see how we might contribute. (We have to see any particular aspect of a collaboration as part of a larger context of information- and reward-sharing.)

The combination of a theory of mind, norms of behavior, abilities to communicate, and shared memory enables agents to overcome these challenges, at least for a time. Theory of mind allows us to understand and predict others and, when combined with mental simulation, allows us to foresee how collaboration might make their desired imagined future real. Cultural norms create regularities and predictabilities in behavior, engendering trust and risk-taking. Communication allows imagined futures, strategies for collaboration, and cultural norms to be shared. Shared memory enables agents collectively to track progress toward goals, to trace contributions and rewards, to create alignment of goals and values, and to foster meaning.

Embedded within all of these capacities, however, are weaknesses as well. While the complexity and range of our collaborations can far exceed those of the other animals, we have yet to construct one that has a hope of lasting as long as the genetically induced collaboration of the honeybee, in existence for thirty million years.

The honeybee has advantages. A primary one is time, that enabled it to align the values of its individuals through the remorseless logic of evolutionary kin selection against a reasonably slow-changing environment. Human value systems, both younger and more dynamic, are often maintained in alignment, by

contrast, through the use of meaning-making narratives and stories, which are subject to powerful and unpredictable pressures from cultural evolution (Milton 2022). Coordination technologies are similarly fragile: where bees can rely on their line of sight and the use of their bodies and pheromone emissions as symbol systems, human coordination often relies on far more ambiguous and hard to interpret signals such as prices in a market (Hayek 1945), or the poetry of a national creed (Nussbaum 2013).

Taken together, our species' reliance on higher-level cognitive and cultural architectures suggests that our collaborations will be not only more complex than the other animals, but that such complexity will also bring with it more points for potential failure. The emotional pain and suffering that accompanies the breakdown of a collaboration is indicative of how important collaboration is to human evolution and the strong motivations we have to make collaborations succeed.

Levels of Collaboration

To ask questions about how collaborations fail is, at least implicitly, to invoke three distinct levels of analysis: (a) the level of the collaboration and its emergent laws of behavior, (b) the level of the individuals that compose the collaboration, and (c) the wider social world within which the collaboration is embedded. While the proximate causes of misalignment crises come from the individual level—it is, after all, the individuals that need to coordinate and cooperate—more complex forms of failure can emerge from the interaction between collaborations.

Figure 18.2 captures some of the basic features of how different levels can relate to each other. Individuals belong to multiple collaborations, which may themselves interact, and potentially compete, within a wider context.

Membership in multiple collaborations often comes with benefits for both the individual and the group, as happens in the case of “board interlock” in corporate and nonprofit settings (Fennema and Schijf 1978), and is a key component of Tocquevillian civil society (Ma and DeDeo 2018). However, those who lie at the intersection of multiple collaborations may also have competing demands that interfere with each other, particularly when they find themselves in a relatively subordinate position in both. This intersectionality can be a source of collaboration failure, as discussed further below.

Out of necessity, Figure 18.2 simplifies the complex relationships that can emerge between levels, where interactions and frustrations can cross multiple levels. Consider, for example, the case of political parties (i.e., collaborative organizations where party elites work together to represent their constituents and maximize the likelihood of reelection). Depending on the form that they take, political parties can be effective at one level but dysfunctional failures at another. In particular, with two forms of political parties—cartel and personalized—what is good for the party can be bad for society at large.

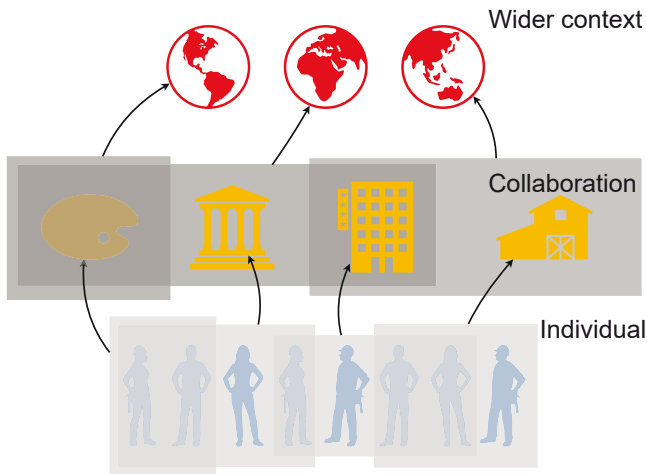


Figure 18.2 Three levels of analysis: Collaborations are composed of individuals, whose psychologies and capacities define and limit how their values and actions can be aligned. Individuals may be members of multiple collaborations at the same time, leading to both synergies and challenges: one might be simultaneously a member of a tribe, a business endeavor, and a political party. This can lead to competing demands and new potentials for misalignment between value systems. Collaborations themselves exist in a common wider context, interfering or coordinating with each other both through ties of common membership.

The conventional political party is the mass party; to acquire political power, it attempts to maximize voter support by appealing to a broad array of interests. Intraparty diversity may make consensus hard to achieve, but when it is able to act, policy tends to serve society broadly, as a reflection of the mass party's representational aims.

A cartel party or set of parties, in contrast, co-opts the state to limit interparty competition. These systems suffer from more limited policy input as parties no longer seek to represent a broad set of voters, and the party's innovation declines (Katz and Mair 1995). In personalized parties, the party elites align around one person (Pedersen and Rahat 2021; Rahat 2022). The dynamics of internal coherence means that internal dissent is less tolerated; the party becomes "tighter." In extreme versions, the party becomes synonymous with a single person.

Both cartel parties and personalized parties have electoral advantages for the collaborating units. They are successful at their own level of fostering collaboration within the party because they either have solved a value alignment problem by eliminating competition (cartel parties) or narrowed the vision to a single person (personalized parties). This enables them to solve the action alignment problem by presenting a coherent and often populist political message.

Such collaborations are, however, harmful at the social level. By rejecting new and diverse information, a party's inventiveness and adaptability in

solving social problems is reduced. In both cartel and personalized parties, they often lose interest in policy achievements and set their goal at elite reelection rather than social problem-solving. Personalized parties have a secondary and significant negative effect at the social level: in personalizing and narrowing the political space, voters are put in a position of supporting or opposing a person. This leads to political polarization, the fracturing of society, and declining trust (Iyengar et al. 2019; Kingzette et al. 2021).

Alignment of Collaboration

Alignment of beliefs and motivations does not require that all participants be the same. Indeed, group diversity might be important for successful collaboration because more diverse individuals might generate knowledge complementarities which improve problem-solving. Alignment of values and actions means that the individuals involved share compatible goals and common knowledge on the ways these goals might be achieved. This does not preclude that the individuals differ in their knowledge, personal experiences, cultural values, or preferences and traits.

In general, both business consultants and management scholars typically conclude that diversity can benefit collaboration (Desvaux et al. 2007; Page 2007a). The particular mechanisms used by high-diversity groups to achieve and maintain alignment may, however, lead to new weaknesses (see discussion below on the Franklin Gap).

Taxonomy of Breakdown

Our basic contention is that collaboration breakdown follows one of a small number of basic patterns: catastrophic collapse, generative reboot, contested persistence, or sputter on launch. We introduce each pattern in turn using case studies to show how the generic forces of misalignment play out in a range of historical and contemporary domains.

Catastrophic Collapse

The most obvious kind of failure is one that leads to complete and rapid dissolution, which we refer to as “catastrophic” failure. When collaborations are small, catastrophic failure may mean that participants in the collaboration simply go their own separate ways, dispersing out into the wider world, perhaps both wiser and sadder, to find new collaborations. When collaborations are large, the participants may continue to encounter each other—and may even form new collaborations—but under radically changed circumstances.

The Breakup of the Beatles

In the early 1970s, the breakup of the Beatles was catastrophic not just in a technical sense (i.e., a complete, unrecoverable breakdown in musical collaboration) but on an emotional level for many fans at the time. It provides our first example of how we can trace a breakdown from its catastrophic collapse backward to misalignment (in this case, of values) and the causal drivers of this misalignment.

As shown in the documentary *The Beatles: Get Back* by Peter Jackson, the Beatles faced strong frictions in the collaboration for a new live album in 1969, largely due to conflicting goals among the members. Although making music together still seemed fulfilling, the group members could not agree on whether to have a concert or where such a concert might take place. Underlying these disagreements were strong misalignments in the artistic values. Paul McCartney disliked John Lennon's experimental work with Yoko Ono, whereas Lennon's more sophisticated tastes had little room for the simple melodies that McCartney preferred (Emerick and Massey 2006). Both Lennon and McCartney lacked respect for the musicianship of George Harrison and at one point considered replacing him.

Failures in value alignment spread to fundamental forms of personal disrespect, and Harrison was the first to announce in 1969, to the other members of the group, that he intended to quit the band (Miles 2009). By the beginning of the 1970s, the Beatles, as a collaboration, had collapsed. Notably, the value misalignments that led to group collapse were not at first accompanied by coordination failures: the group was still able to produce iconic music of enduring popularity, including the material for the album *Let it Be* which appeared in 1970.

A key factor in this breakup was lack of a generally accepted leader or moderator after the sudden death of Brian Epstein in 1967. Epstein had been with the group since 1962 and was sometimes referred to as a "fifth Beatle"; he had a particular talent for balancing the competing goals of the different group members and helping them find and align around values. Though he was somewhat disastrous as a business agent, the band members viewed Epstein as a "spiritual, psychic force" that even Lennon could respect (Geller 2014). His combination of charisma and oddity enabled him to play the role of value-aligner, preventing prior differences in values between the members from percolating up to the level of the collaboration. When Epstein died, value misalignment cascaded into catastrophic collapse.

Collapse of the Hawai'ian Kapu System

Our second example of catastrophic collapse comes from the nineteenth-century history of the Hawai'ian peoples. The collaboration in question was the Hawai'ian *ali'i* (chieftain) aristocracy, which was sustained by the *kapu*

system—an extensive, interconnected system of behavioral taboos that regulated behavior ranging from what foods men and women could eat to the permissible forms of contact between commoners and chiefs. *Kapu* prohibitions underwrote the power of the local aristocracy through their psychological symbolism (e.g., sumptuary regulations on dress and food, stratified by rank and sex), their authorization of violence (even minor violations of *kapu* could be punished by death), and their centrality to the system’s political economy in providing a “grammar” (Crawford and Ostrom 1995) for regulating common pool resources like fisheries (Handy and Pukui 2012).

The collapse of the old aristocracy was triggered by a breakdown in the *kapu* system. It signaled the final stage in the unification of the Hawai‘ian islands and resulted in the termination of the ancient Hawai‘ian aristocracy in favor of a centralized and hereditary monarchy. This process is sometimes referred to as the “Hawai‘ian Cultural Revolution” (Davenport 1969).

The breakdown was intentionally triggered by a series of ritual violations that culminated in November 1819 when the reigning monarch at the time, King Kamehameha II, his regent and foster mother Ka‘ahumanu, and his mother, Queen Keōpūolani, violated *kapu* by staging a meal together, thus violating the prohibitions on men and women eating together. This ritual act led, in turn, to a period of bacchanalian ‘*Ai Noa* (“free eating”), which included not only widespread violation of the eating taboos but also the permanent suspension of religious rituals, the disenfranchisement of the priesthood that underwrote the power of the aristocracy of high chiefs, and the subsequent disintegration of their political economy.

Like the fall of the Bastille in revolutionary France, the feast of November 1819 was a historical “event” in the classic Sewellian sense (Bishop 1976; Sewell 1996). It led to a complete remaking and reorientation of Hawai‘ian society. This move was staged by elites with clear strategic goals. On the surface, the *kapu* system was a system of values, yet in practice it coordinated and organized political, legal, and economic action in Hawai‘ian society. Thus, in their quest to modernize and open the economy to benefit from pan-Pacific trade, the elites attacked it (Levin 1968). The catastrophic collapse of the prior *ali‘i* aristocracy in Hawai‘ian history was, in short, an example of a collapse engineered by a misalignment of actions.

Generative Reboot

The art of losing isn’t hard to master;
so many things seem filled with the intent
to be lost that their loss is no disaster.

—E. Bishop (1976)

While catastrophic breakdowns involve a kind of memory wipe, making what has come before illegible, not all failures take this form. Participants in a

failed collaboration may retain a clear memory of their roles and identities and use their experiences to create a new collaboration after the collapse of the previous one.

With a little bit of luck, the new collaboration will develop new and distinct alignment structures. If the previous collaboration collapsed because of a failure to align on values, the reboot may well experiment with a different set of values that (one hopes) are more robust or easier to hold. If there was a failure to align on actions, collaborators may rely on those prior experiences to create new institutions that they think will answer particular problems that previously emerged.

The Copenhagen Climate Conference

Our first example of generative failure is the collapse of the Kyoto Protocol in Copenhagen in 2009, and its subsequent reboot through the Paris Agreement of 2015. In our framework, the collaboration instantiated by the Kyoto Protocol collapsed because of value misalignments that could not, in the end, be finessed. What makes this interesting is what happened next.

In 1992, the UN established a process to negotiate global climate agreements. This treaty, the United Nations Framework Convention on Climate Change (UNFCCC), serves as a framework for international cooperation to combat climate change. The UNFCCC holds periodic Conferences of the Parties (COP) to assess progress, modify and operationalize existing agreements, or negotiate new agreements.

The Kyoto Protocol, adopted at COP 3 in 1997, represents UNFCCC's first major agreement. The architecture of the Kyoto Protocol was negotiated top down at the international level; emissions limits placed on individual countries were intended to be legally binding under international law. The Protocol included a moral framework of "common but differentiated responsibilities and respective capabilities" that recognized the historic responsibility of the developed countries for historic cumulative emissions and their greater economic and technological capabilities for reducing future emissions. The Protocol set binding targets for 37 industrialized and some middle-income countries, while developing countries did not have targets but were promised financial and technical assistance to build their mitigation and adaptation capacities.

The Kyoto Protocol was rife with value misalignments. The top-down allocation of emissions targets created a zero-sum negotiating mentality. Making the targets binding in international law clashed with sovereignty concerns in some countries (e.g., the U.S. Senate refused to ratify the Protocol). Developed countries viewed major emitting developing countries (e.g., China) as free riders, while developing countries viewed developed countries as abdicating their moral responsibilities to lead in emissions reductions and support the rights of developing countries to develop.

This clear value misalignment quickly resulted in dysfunction: The Kyoto Protocol failed to meet its ultimate collective goal and global emissions continued to rise dramatically. It also missed most of its subgoals as well in that most countries failed to meet their emissions targets and very little financial and technical assistance was actually delivered to developing countries.

COP 15, held in Copenhagen in 2009, was viewed as the last chance to reach agreement on what would come after Kyoto's expiration. World leaders were invited to a summit to create political pressure on the COP to deliver a result.

There were two broad camps at COP 15. The first camp thought that the Kyoto architecture must be preserved at all costs, and that the job of the COP was to renew and strengthen it. This camp was dominant in the proceedings and was backed by a lot of political, institutional, and ideological momentum. A second, smaller, less influential camp saw Kyoto as fundamentally flawed and that it needed to be replaced with a different architecture.

The proposed alternative was a "pledge and review" system; there would be a collective overall goal, but countries would pledge "nationally determined contributions" of emissions reductions toward meeting the collective goal. These pledges would be encoded in national rather than international law, thus avoiding sovereignty concerns. The UN's job would then be to review the bottom-up pledges versus the collective goal, name and shame those not carrying their weight, and attempt to ratchet up ambition over time. The idea was that such a system would create a positive political dynamic, a "race to the top" of virtue competition (and shame those not pulling their weight toward the agreed-upon collective goal), rather than the negative responsibility allocation, zero-sum dynamic of Kyoto.

While there were informal, high-level political discussions of a potential pledge and review type agreement among major emitting countries, the concept was strongly resisted in the formal UNFCCC process, which instead focused on trying to save the failed Kyoto architecture. Because of this, negotiations at COP 15 collapsed in failure, and with it the Kyoto Protocol. No significant agreement could be reached. The event itself was widely viewed as a major blow for global climate cooperation and raised questions about the future existence of the UNFCCC.

Crucially, however, the failure of COP 15 did not mean that nations abandoned the goal of climate change action. Its dramatic failure shifted the gestalt of the process and created the space for the "pledge and review" architecture to develop. Players advocating it gained new positions of authority in the UNFCCC and key national governments, and eventually, the "pledge and review" architecture became the basis of the 2015 Paris Agreement (COP 21). It seems clear that those who survived the debacle of "Brokenhagen" learned from their mistakes. New entrants were brought in to replace those whose ideas and approaches failed. Together, this led to a new logic for climate change action (Falkner 2016) epitomized by the Paris Agreement's new approach to value alignment.

While the Paris Agreement maintained many of the moral principles that animated the Kyoto Protocol, including common but differentiated responsibilities, it responded to the demonstrated failures of the top-down value structure by enacting a bottom-up system that could take advantage of increasing demands for climate action that occur at the subnational and regional level. The diplomatic success of the meeting was driven in part by a more inclusive approach. In response to debacles blamed on the Danish presidency's leadership, for instance, the French government brought in a wider range of stakeholders earlier in the process.

The entrenchment of the prior power structure—institutionally, in terms of who set the agenda and structure of the negotiations; intellectually, in terms of what could conceivably be on the table—meant that Paris could not have happened without the failure at Copenhagen. It is far too early to view Paris a success, given the enormous gap between national pledges versus what is required (Climate Action Tracker 2023), and delivery on developing country assistance. Nonetheless, the architecture has demonstrated more positive political dynamics than Kyoto (as evidenced by the increasing ambition of pledges between the Paris and Glasgow COPs). A generative reboot in response to value misalignment, Paris can thus be viewed as a potential platform for increasingly effective international collaboration on climate, but if and only if major players have the political will and alignment to make it succeed.

Scaling Wikipedia

While some generative reboots end political careers in a spectacular fashion, others spawn thousands of online angry exchanges in the process of finding new solutions. In 2000, Wikipedia became “the encyclopedia that anyone could edit.” By 2005, exponential growth saw the number of editors increase by orders of magnitude. This coincided with a period of highly egalitarian coordination practices, sometimes described as an “adhocracy” (Konieczny 2010), where formal roles and responsibilities were strongly discouraged. During this early period, values were consonant with the unusual coordination demands: the near impossible task of creating a total archive of all human knowledge meant that there was simply no time to refer even major questions and ambiguities to a central authority.

By the early 2010s, the collaboration had undergone a complete realignment. Wikipedian practice—both as described and legislated by users, and as seen in actual patterns of behavior—had shifted to a completely novel oligarchic mode, reminiscent of similar transitions described in terms of Michel's *Iron Law of Oligarchy* (Shaw and Hill 2014). A small group of administrators, endowed with special editing powers by the system software, became gatekeepers of new content. The organizing norms of the Wikipedian collaboration itself, as measured by user invocations of named “Meta” pages, shifted away from an emphasis on ad hoc creation toward a form of bureaucratic

rationalization, an example of the “iron law” of oligarchy supported by Weber’s “iron cage” of rationality (Heaberlin and DeDeo 2016). Both innovations—the concentration of power in an oligopoly centered around early administrators, and the bureaucratization and routinization of task coordination—amount to a dramatic revision in the mechanisms of aligning actions in response to the rapid rise of the Wikipedian population.

Through this dramatic change in action coordination, users remained aligned around the basic values of the encyclopedia. The norms that were central in 2005 remained central a decade or more later (Heaberlin and DeDeo 2016). Although many early users left during this period of transition, the system never lacked for participants, and there was no collapse. The transition from coordination through ad hoc rule invention and rule breaking (an early rule, codified in the Wikipedian Meta pages, was IAR, “ignore all rules”) to a new coordination around more bureaucratic and formal systems happened gradually. At the microscale, of course, the evolution was anything but gradual: there were spectacular “rage quits,” resignations, and public controversies, including a split between founder Jimmy Wales and cofounder Larry Sanger. One might describe this as a transition from Copenhagen-to-Paris through a thousand intermediate steps—a microevolutionary version of the Copenhagen-to-Paris macroevolutionary leap.

Contested Persistence

The best lack all conviction, while the worst
Are full of passionate intensity.

—W. B. Yeats (1920)

Failures take time to unfold, and in some cases, the failure itself may stall. Some fraction of the collaboration may retain high levels of alignment, whereas others may experience significant failures of alignment between the collaboration’s values and their own, or may encounter great difficulty in coordinating correctly with their supposed colleagues. One response to this misalignment is exit (e.g., people may leave the collaboration for greener pastures, as happened, in part, during Wikipedia’s scaling woes). It is not always, however, possible to walk away from a collaboration. This leads to a form of failure we term “contested persistence”: when collaboration, for some of the participants, becomes coercion.

Contested persistence is distinct from both catastrophic collapse and generative reboot. In some cases, some participants may be unaware that the “collaboration” either involves a high degree of coercion or fails to satisfy the aims of many of its participants. One symptom of contested persistence in corporations is the phenomenon of *innere Kündigung* (inner withdrawal or internal resignation; Richter 1999), where in response to organizational conditions,

employees do the minimum necessary to avoid being fired, consciously withdrawing committed effort and renouncing initiative.

Collaborations in contested persistence may be more or less permanent features of the world, lasting many decades or even centuries, such as those beset by systemic injustices such as racism or sexism. They may also be understood as a temporary stage in a longer-term spreading failure. When the failure culminates in a more complete failure of coordination or cooperation, some participants will describe it as the collapse of a collaboration, while others will ask “what collaboration?”

Just as with catastrophic and generative failures, contested persistence may emerge from misalignment in either the value or action domain, as demonstrated by the following cases.

The Kinship Tax

Our first example of contested persistence comes from development studies. In many low-income countries, informal collaboration within a kinship group plays an important role, and the pressure to share resources and income is high. Sharing of resources and income is one mechanism that supports successful group collaboration, as it provides informal insurance and redistribution in the absence of state support (Jakiela and Ozier 2015). Such a “kinship tax” may be enforced within the kinship group, as assiduously as any Western nation-state polices governing tax evasion, albeit through different tools such as familial guilt or psychological shame.

A member of a kinship collaboration may simultaneously be embedded within a Western-style capitalistic market, which puts a premium on personal accumulation of capital and its use and growth by the individual. Assuming the member adopts the values of the capitalist system, they may soon find themselves misaligned with the values of the kinship system but may be prevented from leaving by the familial network of other commitments and values. As cases like this accumulate, collaborations that levy a kinship tax, in the midst of a larger society that is moving toward market capitalism, become contested failures.

The kinship tax is an example of contested persistence created when two systems intersect at the level of the individual and cause value misalignment on a person-by-person basis. In Figure 18.2, the individuals at risk (i.e., those who make the collaboration contested) are those with collaborative “intersectionality” (i.e., memberships in two or more different collaborations).

We expect contestation-from-intersectionality to be a widespread phenomenon that generalizes beyond the specific case of kinship tax, because most people belong to multiple groups (e.g., family, friends, companies, associations). While we expect the “collaborative ecosystem” of stable, successful societies to square the different values and demands of the groups, periods of

social change, where the values of some of these groups shift relative to the others, provide ripe opportunities for contested failure.

Traditionally, men are expected to participate in the public sphere whereas women are perceived to belong to the private, “domestic” sphere (Arendt 2013). They play a crucial role in enabling the functioning of the household: as a key member of the “domestic collaboration,” their value and action alignments are critical for preventing failure. In cases where this traditional social norm does not prevent them from participating in more modern organizations (e.g., associations or the labor market), emergent value misalignments create new conflicts.

The consequence of such conflicts, as in the case of the kinship tax, is often contested persistence. A woman in this situation may shift her values toward those of the new organizations, leading to new forms of value misalignment in the domestic sphere. She may also divert time, attention, and resources to these new organizations, upsetting traditional modes of coordinating household duties. In other cases, she may attempt to participate in the new group without adjusting her values or family coordination, thus causing a misalignment with the new group that may be sufficient to lead to catastrophic collapse.

An example of this gender-based contested persistence, with structural parallels to the kinship tax, can be seen in self-help groups (SHG) for women in rural South India (Coley et al. 2021; see also Chapter 7, this volume). In these cases, women are participants in two collaborations: the private household with traditional gender expectations and the SHG with more modern goals for women’s empowerment. Understanding how these two systems intersect—and how the two collaborations adjust to minimize value misalignment—is crucial to making sense of the particular success, or failure, of an SHG.

Fencing among Maasai Pastoralists

The Maasai refer to themselves as *iltung’ana loo ngishu* (people of cattle). Since livestock brings great prestige, the Maasai tend to own very many animals. In the Greater Mara area of Kenya, due to increasing land tenure uncertainties and a disruption of an earlier nomadic lifestyle (Løvschal et al. 2019) the area is experiencing a severe strain on grazing areas. This makes it increasingly important for the inhabitant pastoralists to coordinate livestock grazing in grazing areas that are still managed as a common resource. The need to coordinate is made increasingly apparent by fencing processes across the Greater Mara, which are appropriating land at an unprecedented and accelerating speed and scale, threatening the very foundation of the ecosystem (Løvschal et al. 2017).

Fencing, in theory, represents an extremely effective coordination technology. However, because fenced land is often used for cultivation and pasture, rather than the informal grazing practices of the Maasai, fencing makes the majority of Mara pastoralists even more dependent on continued access to and availability of spaces for adequate grazing land, which are diminishing.

Therefore, fencing makes coordination around the remaining open access areas even more important.

For pastoralists who depend on large-scale grazing, fencing seems counterintuitive. It has been argued, however, that fencing represents a defense against land tenure uncertainties (Løvschal and Gravesen 2021). In that sense, fencing forms a contested persistence where people continuously misalign their actions in ways that directly impact the future survival of the collaboration itself.

In contrast to the value misalignments of the kinship tax, the Maasai fencing problem appears to originate in the misalignment of actions. One way to envision this was provided by (Håkonsson et al. 2021). In a field experimental study, they had teams of Maasai livestock owners in the Greater Mara area play a computer-supported board game, where the task was to buy and sell livestock, then allocate the livestock across three different grass areas in such a way that the commons was sustained.

When the Maasai participants were provided with an information infrastructure, which enabled them to learn about how grazing behaviors influenced the ecosystem, they were able to collaborate better. Yet, it was not until they were also provided information about each other's actions—how many livestock each team had allocated to a particular area in the previous round—that they were able to coordinate their actions well enough to avoid a collapse of the commons.

Sputter on Launch: Antibiotic Resistance

We are the hollow men
 We are the stuffed men
 Leaning together
 Headpiece filled with straw. Alas!
 —T. S. Eliot (1925)

A final form of collaboration breakdown is when players agree on a collective goal but cannot effectively organize a collaboration to achieve the goal. We call this “sputter on launch.” (More rarely, sputter on launch can emerge in the presence of coordination but lack a common goal. For example, when a group of friends with high levels of trust, technical skill, resources, and spare time cannot figure out what kind of business to start.)

A clear—and dangerous—sputter on launch is the failure to create a global collaboration to act on antibiotic resistance. For decades, the scientific community has issued increasingly dire warnings about the growth of antibiotic resistance (French 2010; Levy 2001), which happens when antibiotic overuse in humans and animals breeds drug-resistant strains. In 2019, a major study attributed five million deaths worldwide to antibiotic resistance (Murray et al. 2022). Some fear that our current usage of antibiotics could render the treatment largely ineffective in coming decades and that humankind would return to the miseries of the pre-antibiotic era.

Despite widespread agreement on the threat and clear need for global coordination to fight antibiotic resistance, no effective collaborative architecture has yet emerged to address the problem at the scale or timeline required. To help coordinate research toward this end, the World Health Organization (WHO) has created the Global Antibiotic Research and Development Partnership (GARDP) and various plans, studies, and initiatives have been announced and carried out by WHO, national health bodies, private companies, and NGOs. Still, the scientific community continues to warn that the scale and speed of actions and levels of coordination remain dangerously inadequate.

The Franklin Gap: Ambiguity, Robustness, and Failure

Civilization is hooped together, brought
Under a rule, under the semblance of peace
By manifold illusion

—W.B. Yeats (1934)

Healthy collaborations are often characterized by a tolerance of noise and ambiguity. If it is possible to finesse misalignments of value through strategic use of language, for example, collaborations may survive changing circumstances. An example, taken from the autobiography of Benjamin Franklin (2012), describes Quaker leaders from Philadelphia in the eighteenth century. Pacifist to the core, when faced with the danger of enemy attack, they allowed the purchase of gunpowder only after it was included under a budget item valid for “wheat or other grain.” Moving from political performance to music, we find that tolerance of noise, error, and ambiguity in coordination is a core part of polyrhythmic performance traditions: musicians maintain microscale coordination through a “pulse” mechanism, rather than a more top-down bar-and-measure structure that demands strict coordination with a single leader (see also Chapter 9, this volume). Indeed, the psychological act of listening to music at all may partially depend on continuously adjusting an abductive process, where listeners revise their prior understanding in response to the flow of novel sonic events (Oliveira et al. 2010).

Ambiguity like this may be a source of robustness, but it can also lead to a particular form of contested failure, co-optation, introduced above in the context of party politics. Z. DuBois (pers. comm.) described an example of co-optation involving labor unions and management. As a city bus driver and member of a local amalgamated transit union in Detroit, Michigan, DuBois experienced union leadership that were paid off by management. This led to high levels of frustration among union members and prevented the union from creating positive change for its members, either via negotiation or advocacy for better working conditions. It also severely limited the likelihood that union leadership would draw on strategies, such as work stoppages, to effect change

for its members. The “Franklin Gap” between values and actions worked, in this case, to stall rather than enable collaboration.

Another example of a more ambiguous failure is the co-optation of city LGBTQIA+ Pride parades, and their local organizing committees, by corporate power. Historically, Pride events commemorated acts of resilience and resistance and were community organized. In that way, they also mirrored their communities where they took place (e.g., San Francisco looked very different from Boston which looked different from Berlin). More recently, however, with increased visibility and acceptance, Pride events have become celebrations whose floats and events are dominated by investor companies eager to advertise their inclusivity and support. The result may reflect many of the aims of the original event, but the very open-endedness of the original collaboration, a source of its robustness in earlier times, allowed for a drift in meaning that vitiated much of its original power.

Ambiguity, in other words, cuts both ways. In the generative reboot of Wikipedia discussed above, the free play of “ignore all rules” was replaced with a more rationalized structure. This did not mean, however, that ambiguity was eliminated. Heaberlin and DeDeo’s (2016) claim is that the modern incarnation of Wikipedia plays out a version of Meyer and Rowan’s (1977) institutional myth, where a useful ambiguity can reemerge in, for example, the logic of “good faith.” Wikipedia’s persistence may come in part from new forms of the Franklin Gap.

Collaborations, of course, can fail when ambiguity enables co-optation. The most free-flowing and open styles stand likely to both benefit and suffer in this respect. A classic essay on anarchism cites an old adage: “Anarchism has a broad back, like paper it endures anything” (Chomsky 1970:1). That endurance comes with costs. Among other things, it includes, as the essay notes, many people whose acts, undertaken in the name of anarchism, are such that “a mortal enemy could not have done better” (Chomsky 1970:1).

Discussion

If we can identify the generic features of how collaborations break down, we can better identify which collaborations might be in danger, plan when failure is near, and use this information to build more robust collaborations in the future. To that end, we offer a simple message: collaboration failure is best understood through the causal drivers responsible for misalignment—either of values or actions—and that the causal sequelae of these misalignments trace out a small number of basic patterns.

One consequence of our analysis is the idea that policy makers conduct what we call an “alignment review”: an assessment of how a policy change will affect the alignment of values and actions of the various collaborations that exist within the society in question. Such a review would be a natural

extension of already existing practices. It is increasingly common to consider the effect policy changes might have on the culture as it currently exists, and policy makers are increasingly aware of the phenomenon of institutional crowding-out (Bowles 2008; Ostrom 2000; Wrzesniewski et al. 2014)—the perverse effects that novel external incentives can have on behavior previously guided by longer-standing traditions. A classic example comes from a study of day-care centers in Haifa by Gneezy and Rustichini (2000), who showed that parents were more likely to violate the social norm of being late to collect their children at the end of day-care center hours when an additional external fine was imposed.

Our work suggests a larger framework for conducting this kind of analysis. By becoming familiar with the values and coordination mechanisms that underlie a society's successful collaborations—as well as the taxonomies of dysfunction described in this chapter—policy makers have a chance to investigate how necessary alignments might be stabilized. This will also enable them to make sure that policy interventions do not disrupt or enfeeble these alignment mechanisms, or, if required, that truly necessary interventions provide new mechanisms to replace old ones.