CHAPTER

4

# Examining how to enhance collective action

# Examining how to enhance collective action

Examining how to enhance collective action to manage interdependence can be explored through different assumptions about human behaviour's interactions with institutions. Different explanations for behaviour can inform ways of advancing collective action to provide global public goods.

Insights from recognizing how behaviour and institutions are contingent on the changing social context over time can help address shared challenges. A broader perspective on choice informed by these insights also shows how risks associated with domestic patterns of political polarization may harm collective action across countries.

"...the making of a treaty is the treaty. It doesn't matter what the terms are, just that there are terms. It's the goodwill that matters. When that runs out, the treaty is broken, whatever the terms say."

—Hilary Mantel[1]

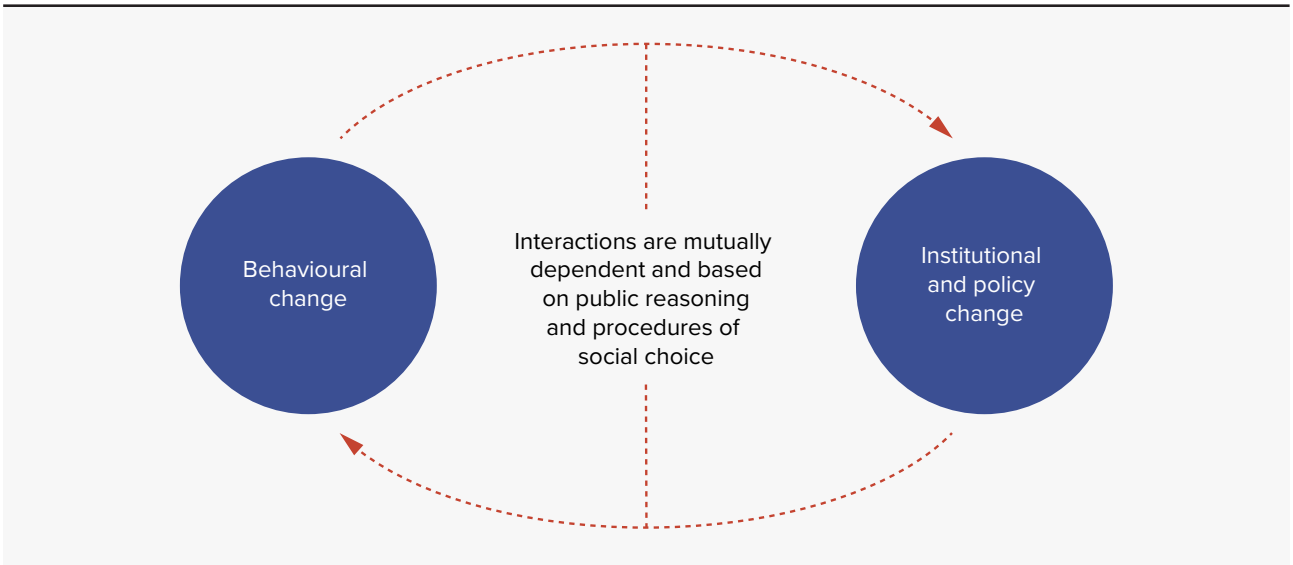
Examining how to enhance collective action to manage interdependence can be explored through different assumptions about human behaviour and its interactions with institutions.[2] This chapter considers how different explanations for behaviour can inform ways of advancing collective action[3] for the provision of global public goods.[4] It explores three perspectives on behaviour and the interventions to enhance collective action that emanate from these perspectives.[5]

- *Selfish choice.* Under a standard selfish choice model of behaviour, enhancing collective action depends on interventions that reshape incentives by providing information or resources to align narrow self-interest with improved collective outcomes. International treaties mobilize interventions that reshape incentives. For climate change, incentives can be altered by pricing carbon; applying information from scientific syntheses, such as those produced by the Intergovernmental Panel on Climate Change; and using resources from funds that support countries in mitigating climate change, such as the Green Climate Fund.
- *Behavioural choice.* Human behaviour often deviates from the assumptions of the standard selfish choice model, deviations that are sometimes described as behavioural biases. For instance, providing new information alone does not always lead people to update their beliefs.[6] And providing financial rewards to change incentives can undermine cooperation that is motivated by a social norm.[7] Even though large swathes of debate in the social sciences and humanities take issue with the emphasis of behavioural science, providing explanations for behaviour and institutions that explore culture, context and power,[8] insights from behavioural science yield a richer description of behaviour than the selfish choice model and thus suggest other ways of intervening that supplement incentives by also changing what people focus on and how they feel and think.[9] For example, changing social norms can enhance collective action by activating a social tipping point, as when reaching a threshold of enough solar panels flips the community norm to making solar panels the social standard.
- *Encultured choice.* Explicitly bringing in culture can explain how people's beliefs result from experience and exposure to different social contexts, shaping their perceptions, self-image, aspirations and meanings.[10] This perspective accounts for why some behavioural biases, thought to be universal and hard wired under the behavioural choice perspective, are culturally contingent.[11] It also explains how behaviour is sometimes constrained by people's inability to imagine more prosperous and fulfilling lives, curtailing their aspirations and their agency.[12] This perspective has implications for cooperation, too, as when people's affiliation with a group is tied to a salient aspect of their identity—such as opposing vaccination as a marker of belonging to a group that is sceptical of government intervention, resulting in the less cooperative behaviour of not being vaccinated.[13] Understanding how these dynamics take hold and change points to recognizing the social context, including patterns of political polarization and mistrust within countries that may stand in the way of enabling collective action at higher scales.

## "Changes in behaviour and in institutions can foster collective action that enhances the provision of global public goods

Fostering collective action for the provision of national public goods is one of the primary roles of governments, in part through centralized enforcement.[14] But since countries are sovereign,[15] they have to voluntarily agree to collective action without centralized enforcement.[16] So, the lens has to move towards an exploration of the processes of social choice that can enhance the provision of global public goods.[17] That implies changes in behaviour (countries shifting from not contributing to contributing to a global public good) and in institutions (establishing a treaty or a creating a multilateral organization that enhances the provision of a global public good), along with the interaction between the two (figure 4.1).[18] Behaviour and institutions are interdependent, as argued

**Figure 4.1 Behavioural change and institutional reform influence each other—jointly shaping and being shaped by social choice procedures**




**Source:** UNDP 2022a.

conceptually[19] (and demonstrated through models[20]) and experimentally.[21]

Where to start? The three sets of assumptions about behaviour discussed in this chapter point to three different answers.[22] The simplified set of assumptions of the standard selfish choice model begins by thinking about the design of institutions to enhance collective action. By contrast, a behavioural model of choice opens the possibility of directly changing behaviour to enhance individual and collective outcomes. While insights from both perspectives are useful, the recognition of how behaviour and institutions interact in different social and cultural settings supplements the first two sets of assumptions by emphasizing the contingent nature of both behaviour and institutions.

## Start with a standard selfish choice model of behaviour

In a standard selfish choice model of behaviour, a decisionmaker seeks to do as well as possible to fulfil a fixed and stable set of preferences and assumes that everyone behaves the same way (box 4.1).[23] This behavioural model is the foundation for much economic and political science analysis associated with collective action. And it is implicit in the discussion in chapter 3 of the prospects for providing global public goods under different aggregations.[24] So, when can collective action without enforcement from above happen under these assumptions?

It is crucial to distinguish two different situations. One in which everyone desires the same thing, but some common standard needs to be set (such as deciding which side of the road to drive on or which language to communicate in).[25] The other in which there are different interests on what is desired and where the pursuit of those individual interests does not yield what is most desirable collectively, posing social dilemmas.

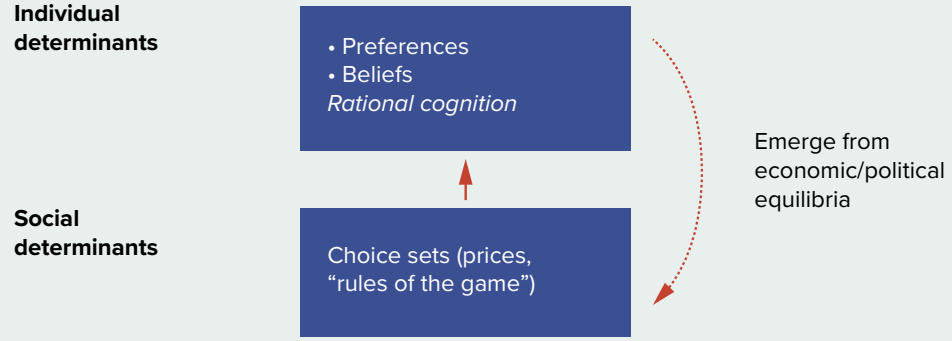
What matters in the first situation is that everyone adopts the same standard. While multiple standards may exist (driving on the left or on the right), all that matters is that everyone chooses the same side of the road.[26] Once a standard emerges, there is a strong incentive to comply with it—for instance, to comply with the standard of driving on the right rather than defecting and driving on the left. The difficult bit is setting the standard to begin with, a challenge of getting everyone in sync. Collective action in this situation needs to overcome a coordination problem.

The key obstacle to overcoming the coordination problem is not diverging interests—interests are aligned. Even though everyone wants the same thing, uncertainty about how others will act can lead to coordination failures that impede collective action.[27] Measures to enhance collective action associated

**Box 4.1 A standard selfish choice model of behaviour**

Preferences are exogenous and drive each decisionmaker (or agent) to pursue individual self-interest (box figure 1). The agent's beliefs, separate and independent from preferences, are based on information collected to help the agent make a specific decision. For instance, given a preference not to get drenched, an agent needs to form a belief about whether it is going to rain before choosing whether to take an umbrella when going out. The belief is based on the collection of information, such as by consulting a weather forecast in the evening. And the preference to not get drenched has no bearing on how the belief is formed (so things such as wishful thinking, where the preference not to get drenched shapes the belief that it is not going to rain, are not allowed in this model of behaviour). Beliefs are updated if the information changes—if the weather forecast consulted in the morning is different from the one consulted the previous evening. Rational cognition is defined by a set of axioms implying, among other things, that preferences can always be ordered in a consistent way.

**Box figure 1 In a standard selfish choice model, behaviour is determined by the exclusive pursuit of self-interest**




**Scope of interventions to change choices**

- Correct "market faliures" (externalities, information asymmetries and the like) through regulation/decentralized approaches
- Governance

- Preferences are stable
- Beliefs are informed by optimal (not too little, not too much) collection and processing of information, for any given preferences
- Rational cognition: doing as well as (she rationally believes) she can, assuming that everyone else is doing to the same

**Source:** Human Development Report Office elaboration based on Elster (2015a, 2020) and Hoff and Stiglitz (2016).

with coordination challenges include those directed towards ameliorating this uncertainty, through interventions that seek to get everyone on the same page (or side of the road).[28]

## Transforming cooperation challenges into coordination problems can enhance global collective action

Coordination challenges related to transportation and communication have been successfully overcome at the global level, resulting in collective action for the adoption of regulatory practices and regimes that enable air travel, maritime shipping and digital communication.[29] The resulting standards yield very high benefits and very few constraints (a country is constrained in not opting out of the standard but would derive no benefits from doing so). And though these standards are sometimes derided as instances of shallow international cooperation,[30] their existence may suggest that global collective action that results from addressing coordination problems is something that sovereign countries can readily do.[31]

But a very different scenario emerges in the second situation, when the pursuit of varied selfish interests is not aligned with what would be more

desirable collectively. When what most advances each decisionmaker's self-interest is not the best possible collective outcome, this situation creates a social dilemma between self-interest and collective action. Chapter 3 showed that this situation plagues many global public goods. Collective action in this case requires cooperation, so that decisionmakers choose an action that is suboptimal for their selfish interests but superior for the collective. Recognizing that countries find it relatively easy to coordinate their actions—even while struggling to voluntarily cooperate and sometimes enforcing agreements to do so[32]—opens the possibility of designing institutions (such as multilateral organizations or international treaties) that shape incentives so that a cooperation challenge becomes a coordination problem.[33]

"Summation global public goods typically require cooperation, but institutions can be designed in a way that reshapes incentives to turn a social dilemma into a coordination problem, as with the Montreal Protocol

How can challenges of international cooperation be reshaped as problems of international coordination? It is possible to learn from multiple successful examples of such reshaping (spotlight 4.1). For weakest-link global public goods, such as eradicating a global (and eradicable) communicable disease, we are already close to a coordination problem. This gives insights into the kind of reshaping that could be pursued. In disease eradication all countries share the same objective, this objective can be defined with certainty and precision and each country has an incentive to contribute if it can be sure that other countries will do their part.[34] The key challenge is for countries to coordinate their actions in a way that sustains incentives to shore up the weakest links until the disease is eliminated.

Summation global public goods typically require cooperation, but institutions (for instance, international treaties) can be designed in a way that reshapes incentives to turn a social dilemma into a coordination problem, as with the Montreal Protocol.[35] Asking countries to contribute to the summation global public good of avoiding depletion of the ozone layer by limiting emissions of ozone-thinning chlorofluorocarbons (CFCs) is a call for international cooperation. But the treaty did not simply do that. It also banned trade in CFCs and products containing CFCs between countries that were parties to the treaty and countries that were not, effectively providing incentives for high-income countries to sign the treaty. This is because trade interactions between countries are bilateral and reciprocal, with compliance easier to monitor and enforce (spotlight 4.1).[36]

Stipulating a minimum number of ratifying countries for the Montreal Protocol to enter into force meant that a tipping point was eventually reached, aligning incentives to make (high-income) nonsignatory parties better off by signing the treaty—thus resolving a coordination problem.[37] To provide incentives for low- and middle-income countries to sign the treaty, a later amendment established a financial mechanism (the Montreal Fund) to compensate countries for the incremental cost of participation. These side payments induced virtually universal participation. Finally, technological alternatives to CFCs were widely shared and advertised, including by firms that stood to gain from adopting these alternatives.[38]

### Uncertainty can harm international collective action

Setting thresholds can motivate collective action, as seen in disease eradication (where the threshold for full provision is eliminating the disease in the weakest-link country)[39] or in the Montreal Protocol (with the establishment of a minimum number of countries for the treaty to come into effect, plus the trade provisions). The underprovision of some global public goods, such as climate change mitigation or biodiversity preservation, is often framed as the need to stay within boundaries or limits[40] to avoid reaching tipping points in planetary systems that could result in catastrophic outcomes.[41] Presenting thresholds that, once crossed, can result in catastrophic societal collapse could galvanize collective action.[42] But there are two critical conditions.[43] First, the thresholds must be known with little uncertainty. Second, each country must share the burden of not passing the threshold.[44]

When there is uncertainty about where the thresholds lie, collective action becomes more difficult. For disease eradication, to achieve zero cases globally,

each country has to achieve zero cases nationally. But when there is no unambiguous way of determining how much each country should contribute to ensure that the world stays under a threshold for catastrophe, that calls for some agreement about how to allocate effort across countries.[45] These two factors imply a very different set of incentives for countries.[46]

"The standard selfish choice model can provide insights about how to enhance the provision of global public goods. Its behavioural assumptions emphasize the use of incentives, resources and information to make provision more likely

While much effort centres on estimates of damages from crossing climate thresholds or whether damages are overstated or understated,[47] uncertainty about these damages does not seem to matter as much for cooperation as uncertainty over the thresholds.[48] Because even though the decision on whether to cooperate has no influence on the scale of the damages (which depends only on whether the threshold is crossed), whether countries cooperate does bear on whether the threshold is surpassed.[49]

In the case of existential risks, these insights could inform ways to structure incentives through institutions to enable cooperation associated with providing global public goods that reduce those risks.[50] These insights also suggest that it is far more important to reduce threshold uncertainty than damages uncertainty, a challenging task given the underlying ambiguity in many of the thresholds of interest.[51]

The standard selfish choice model can thus provide insights about how to enhance the provision of global public goods. Its behavioural assumptions emphasize the use of incentives (trade provisions in the Montreal Protocol), resources (the Montreal Fund) and information (about the damage caused by CFCs and the availability of alternatives) to make provision more likely. But these assumptions also have limits, not only in not accounting for actual behaviour but also in lacking power to account for some of the obstacles to collective action that call for a broader understanding of the drivers of human behaviour.[52] Behavioural science provides an initial steppingstone towards a broader explanation of behaviour.

## Apply insights from behavioural science, but handle with care

Insights from behavioural science reveal how decisions depart from the behaviour predicted by the standard selfish model of choice (box 4.2).[53] These departures give added insights in understanding when and why collective action takes place and provide opportunities to design interventions that make the provision of global public goods more likely.[54] At the same time, as the discussion will make clear, despite the enormous interest in designing interventions based on behavioural science, there are severe limitations as well, that range from the lack of replicability of some findings, questions about their validity beyond specific experiments and the assumption that the policymaker "knows better" than individuals—among others. That is why it is important to consider insights from behavioural science but to also handle them with care.

### Beyond selfishness—recognizing social preferences

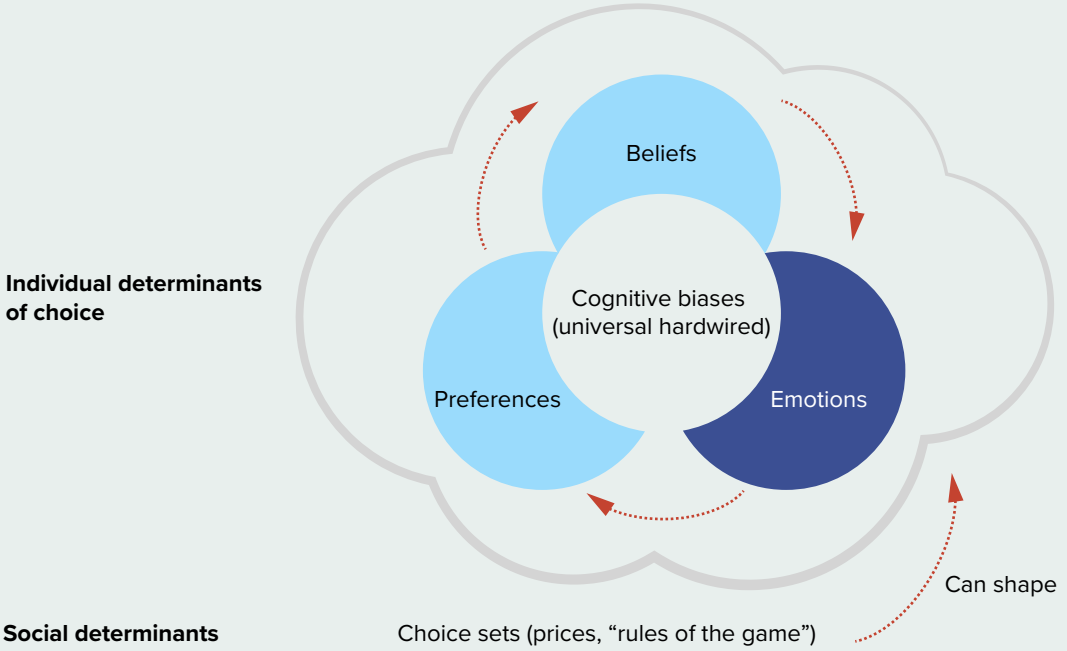
With social preferences decisionmakers consider the welfare of others, not just their self-interest, and are prosocial when that evaluation is positive.[55] Pure self-interest can motivate cooperation through reciprocity in repeated interactions (giving something today while expecting to get something in return tomorrow).[56] But people often reciprocate more generously when others behave in a friendly way and punish more harshly those who do not (even if at great cost), indicating that social preferences are likely at play.[57]

People vary in the extent to which they have either selfish or prosocial preferences and in how they express social preferences.[58] A recent comprehensive review of social preferences provides some key insights.[59] When, at the turn of the 20th century, researchers concluded based on systematic evidence from experiments that some people had social preferences, respondents who expressed these preferences were described as being "crazy."[60] Yet evidence from nationally representative samples suggests that in many countries people holding purely selfish preferences are in the minority (representing 5–20 percent of the population in countries with data).[61]

## Box 4.2 A behavioural choice model of decisionmaking

Preferences, beliefs and cognition interact to shape how people make decisions (box figure 1). Preferences can be social, meaning that the individual takes account of other people's welfare. Preferences can be fickle and influenced by (sometimes transitory) emotions. Fear triggered by the belief that there is a threat tends to make people more risk averse, while anger tends to make them more risk seeking. Beliefs are driven not only by the processing of information but also by one's identity (perceptions of belonging to a particular group that holds a particular view) or preferences (for instance, motivated reasoning, as in wishful thinking, which makes people believe that a goal they are pursuing is more likely to be achieved). When beliefs are deeply held or linked to a salient aspect of a person's identity, they can trigger strong emotional responses (as when challenging deeply held religious or political beliefs triggers anger, disgust or hate). There are multiple deviations from rational cognition—for example, how a decision is framed affects choices, and how people discount the future is not consistent along different time horizons.

**Box figure 1 Social context shapes what people think and do at the moment of choice**




- Preferences can be fickle (emotions "interfere")
- Beliefs can be motivated (by preferences)
- Cognitive biases and limitations (loss aversion, hyperbolic discounting, representative heuristics, ...) are hardwired and universal

**Note:** The text in bold to the right of "Social determinants of choice" and "Scope of interventions to shape choices" denotes the new elements that are added to the selfish choice model of behaviour (which remains relevant).
**Source:** Human Development Report Office elaboration based on Elster (2015a, 2020) and Hoff and Stiglitz (2016).

Some evidence suggests that holding prosocial preferences is a direct determinant of wellbeing, with a positive effect similar in size to the effects of parenthood, income and education.[62] In addition, there are very strong positive links between prosocial preferences and cooperation.[63] When social preferences take the form of aversion to inequality, more heterogeneous collectives (in either resources or benefits

derived from public goods) are less likely to achieve and sustain cooperation, which provides a rationale for reducing inequalities to foster cooperation.[64]

Do these relationships between individual social preferences and cooperation scale from the individual to more aggregated levels? Particularly relevant for the provision of global public goods is whether the relationships scale up to countries. Some evidence suggests that they can (box 4.3). With the assumption of prosocial preferences, the prognosis for summation global public goods can change. For instance, in the standard selfish choice model a country's unilateral increase to a summation global public good (say, abatement of greenhouse gases to mitigate climate change) will not only not incentivize other countries to contribute but will also likely provoke a reduction in their efforts. But that expected outcome changes if the other countries behave as if having prosocial preferences: in that case contributions to summation global public goods are no longer pure substitutes but become complements.[65]

## Harnessing social norms

Social norms set shared expectations of behavior, providing structure to people's beliefs and bearing on the decision to cooperate (box 4.4).[66] In many circumstances, they can be stronger determinants of behaviour than the individual pursuit of material wellbeing.[67] For instance, social norms can establish what may be required to earn a reputation as a cooperator.[68] Concerns with social image can also drive prosocial behaviour: because most people care about what they believe others will think of them, they tend to make more prosocial choices in public.[69] The effectiveness of social norms in shaping behaviour depends in part on norm enforceability (or beliefs about

### Box 4.3 Social preferences can scale up

A complicating factor in addressing whether social preferences scale up is the great variation across people in every country. The outcome of cooperation at the group level depends on the number and intensity of individuals with prosocial preferences.[1] Another challenge is that while cooperation may be strong within groups in society (as for people sharing the same political beliefs), it may be difficult to get cooperation between groups. In fact, strong negative reciprocity (punishing, or threating to punish, another group harshly) may trigger retaliatory action (or even preemptive aggression)[2] by those who are punished or threatened with punishment.[3] And that can result in intergroup conflict.[4] Chapter 6 explores in more detail the implications of this type of intergroup dynamics, which are particularly challenging in polarized societies.

Still, recent advances in measuring differences in preferences at the global level are starting to provide some answers about what scales up to countries. In an experimentally validated survey[5] on the social preferences of 80,000 people in 76 countries, cross-country variation in charitable giving is correlated with prosocial preferences, after factors that could also explain charitable giving are controlled for.[6] In addition, after the same factors are controlled for, countries with a higher degree of negative reciprocity have suffered more violent conflicts.[7] A study of 40 countries found that people everywhere were more likely to return a wallet with money than what a standard selfish choice model would predict,[8] with prosocial preferences (in this case, measured by the extent to which concerns for welfare extend beyond one's ingroup) playing a role.[9] And in another study of 31 countries, prosocial preferences were associated with better environmental performance—a proxy for cooperation to manage environmental externalities—and material interests mattered less than appeals to everyday cooperative behaviour.[10]

**Notes**

**1.** In an experiment Fehr and Fischbacher (2003) show that a minority of altruists can force a majority of selfish individuals to cooperate but that a few egoists can induce a large number of altruists to defect—and that the context matters in both cases. Hauser and others (2014) show that mechanisms can be designed to ensure that those with prosocial preferences can restrain defectors in an intergenerational public goods game. And Gächter, Kölle and Quercia (2017) show that it matters whether the challenge is to provide or maintain a public good. **2.** Böhm, Rusch and Gürerk 2016. **3.** Nikiforakis 2008. **4.** For a broader review of the psychological foundations of intergroup conflict, see Böhm, Rusch and Baron (2020) and De Dreu and others (2022). There is growing evidence of differences between dispositions to defend or to attack. For instance, consistent with loss aversion, experiments suggest that people invest more resources to protect against losses than to achieve victory (Chowdhury and Topolyan 2016; De Dreu and Gross 2019). **5.** See Falk and others (2023) for details. **6.** Falk and others 2018. **7.** This variation in preferences appears to be deeply rooted in history and to be path dependent (Becker, Enke and Falk 2020). **8.** Whether the interaction was in person or mediated through computers made a difference. Interaction through computers increased cheating threefold compared with in-person interactions (Cohn, Gesche and Maréchal 2022). **9.** Cohn and others 2019. **10.** Van Doesum and others 2021. The findings were contested (Komatsu, Rappleye and Silova 2022) but appear to hold after scrutiny (Van Doesum and others 2022).

## Box 4.4 "It's not a lie if you believe it"—Beliefs, social norms and collective action

The formation of beliefs and their interaction with preferences and emotions has implications for collective action.[1] For beliefs about how others will behave during social interactions, the standard selfish choice approach assumes that everyone behaves the same way.[2] The behavioural approach allows for more nuance and variability in how we expect others to behave, which can be influenced by factors ranging from the ability of agents to communicate[3] to perceptions of trust among agents[4] and assumptions about the preferences of others (whether they are conditional cooperators).[5]

Mechanisms of controlling and selecting those with whom to cooperate are key to sustaining cooperation,[6] with reputation a key driver of beliefs about whether counterparties are likely to cooperate (or reciprocate in the future).[7] These mechanisms can also support cooperation across groups (including countries) in what has been termed "universal cooperation."[8]

People differ in how much they are influenced by the decisions and behaviours of others.[9] There is substantial evidence that social comparison is a powerful driver of changes in individual behaviour, including changes aimed at addressing climate change.[10] Shifts in social norms can also drive changes towards more cooperative behaviour[11] in the face of threats.[12] In addition, social contagion appears to be a strong mechanism leading to proactive cooperative behaviour not only when responding to threats but also as evidenced when behaviour by neighbours is replicated, as in the adoption of solar panels.[13] People are more driven to change their behaviour when they observe others acting than when they simply receive information on the benefits of the cooperative action.[14] Policy interventions have the potential to tip social norms towards more desirable outcomes (including more cooperation).[15] And this potential has been studied across a wide range of challenges, from handling misinformation to advancing public health and fostering collective action for sustainability.[16]

**Notes**

**1.** Isler and others 2021. The quote in the title of the box is a line by the character George Costanza in the television series "Seinfeld," as used in Bicchieri, Dimant and Sonderegger (2019). **2.** This also includes subjective beliefs, for instance, about whether countries are optimistic or pessimistic, as explored in Im, İriş and Ko (2022). Fehr and Charness (forthcoming) discuss belief-dependent social preferences (where beliefs about the intentions of other players matter) using models of reciprocity and guilt aversion (related to theories that include emotions as part of social preferences). **3.** Barbieri 2023; Crawford 2019; Ellingsen, Östling and Wengström 2018. **4.** Bose and Camerer 2021; Schilke, Reimann and Cook 2021. Emotions affect how trust beliefs are formulated, with angry people typically being seen as less trustworthy (Kausel and Connolly 2014). Some evidence suggests that people who have a preference both for being honest and for being seen as honest are more likely to be truthful (Abeler, Nosenzo and Raymond 2019). **5.** Engelmann and others (2019) show the neural signals when beliefs about conditional cooperation are violated. **6.** Reviewing the experimental literature on infinitely repeated games, Dal Bó and Fréchette (2018) find that while cooperation can be supported in equilibria, it does not imply that most subjects will cooperate to begin with—cooperation will emerge only when the structure of the game is robust to strategic uncertainty. **7.** Balliet and Van Lange 2013; Gross and De Dreu 2019; Jordan and Kteily 2023; Rand and Nowak 2013; Roberts and others 2021. **8.** Gross and others 2023. **9.** Kendal and others 2018; Mesoudi and others 2016. In part because that much social information is "wasted," in that it is not used in individual decisionmaking (Morin and others 2021), leading to heterogeneity across the population. **10.** In a second-order meta-analysis of 10 meta-analyses of 430 primary studies, Bergquist and others (2023) found that social comparison was one of the most important mechanisms in driving changes in behaviour, such as towards sustainable transportation or circular consumption to mitigate climate change. **11.** For a review of evidence, mechanisms and potential to inspire interventions to harness social norms, see Frank (2021). See also UNDP (2020b). For the social dimensions of fertility choices and consumption patterns, see Barrett and others (2020). **12.** Szekely and others 2021. **13.** Allcott 2011; Barnes, Krishen and Chan 2022; Bollinger and Gillingham 2012. **14.** Kraft-Todd and others 2018. **15.** Andreoni, Nikiforakis and Siegenthaler 2021. **16.** Nyborg and others 2016. See the review in Efferson, Vogt and von Flüe (forthcoming).

its enforceability),[70] on how strict the norms are[71] and on the balance between rewards for compliance with norms and punishments for deviations from them.[72]

Emotions play a central role in compliance with social norms, with people complying to avoid shame or guilt and people motivated to enforce norms out of anger or disgust for norm violators.[73] A social norm of conditional cooperation—full cooperation as long as others also fully cooperate and reduced cooperation as others' contributions go down—can account for a large set of regularly documented cooperation-related behaviours.[74] In repeated interactions the observed behaviour of others can inform the decision on whether to cooperate and by how much. But in one-shot interactions or when the behaviour of others is not observed, beliefs about how others will behave are determinant. This insight is crucial to the discussion in chapter 6 on the potential of misperceptions about what others believe to hinder collective action.[75]

Social norms can be harnessed to change collective action at scale[76] (see box 4.4) and have distinctive characteristics that aggregate to countries, which

accounts for differences in cooperative behaviour.[77] One such characteristic is the tightness of social norms (as measured by the harshness of punishment of behaviour that deviates from the norm), which appears to vary systematically across countries[78] and change over time.[79] When facing a collective threat,[80] countries with tighter norms may cooperate better internally because of the cohesive glue of strong social norms.[81] But extreme tightness can make cooperation across groups or countries more challenging (or can even trigger conflict).[82] Tighter social norms can also make adaptations to a changing context more difficult, potentially resulting in a mismatch between internal and international cooperative arrangements in the face of new threats and challenges, with implications for the provision of new global public goods.[83]

> "Leadership can propel and sustain social norms that are supportive of international cooperation: that can shift norms and trigger reciprocal actions from other countries that further entrench the norm of contributing to the global public good

Moreover, global norms often influence countries' decisions.[84] For instance, norms against gender inequality spread globally.[85] But as with any social norm, progress cannot be taken for granted. And it can be subject to contestation, particularly when polarization fuels backlash against more inclusive norms, as discussed in chapter 6.[86] But leadership can propel and sustain social norms that are supportive of international cooperation, for instance, when a country takes the lead on providing a summation global public good such as mitigating climate change: that can shift norms and trigger reciprocal actions from other countries that further entrench the norm of contributing to the global public good.[87]

### Drawing on cognitive biases

Almost 200 cognitive biases have been identified to explain several puzzles in the social sciences,[88] opening windows for new policy interventions and motivating a wide range of organizations[89] and initiatives around the world that seek to enrich public policy with these insights (figure 4.2 and spotlight 4.2).[90] For instance, loss aversion (people caring more about a loss than an equivalent gain) has been empirically documented in a wide range of studies.[91] This insight has explained behaviours where the framing as a loss or gain influences decisions ranging from how much to work[92] to political choices[93] to why people tend to hold on to their beliefs[94] and to the design of strategies to foster learning.[95] More generally, behavioural insights have informed policy[96] through new policy tools (such as nudges), enabling better predictions about the impact of policies and generating new implications for how to enhance welfare[97] and its distribution.[98]

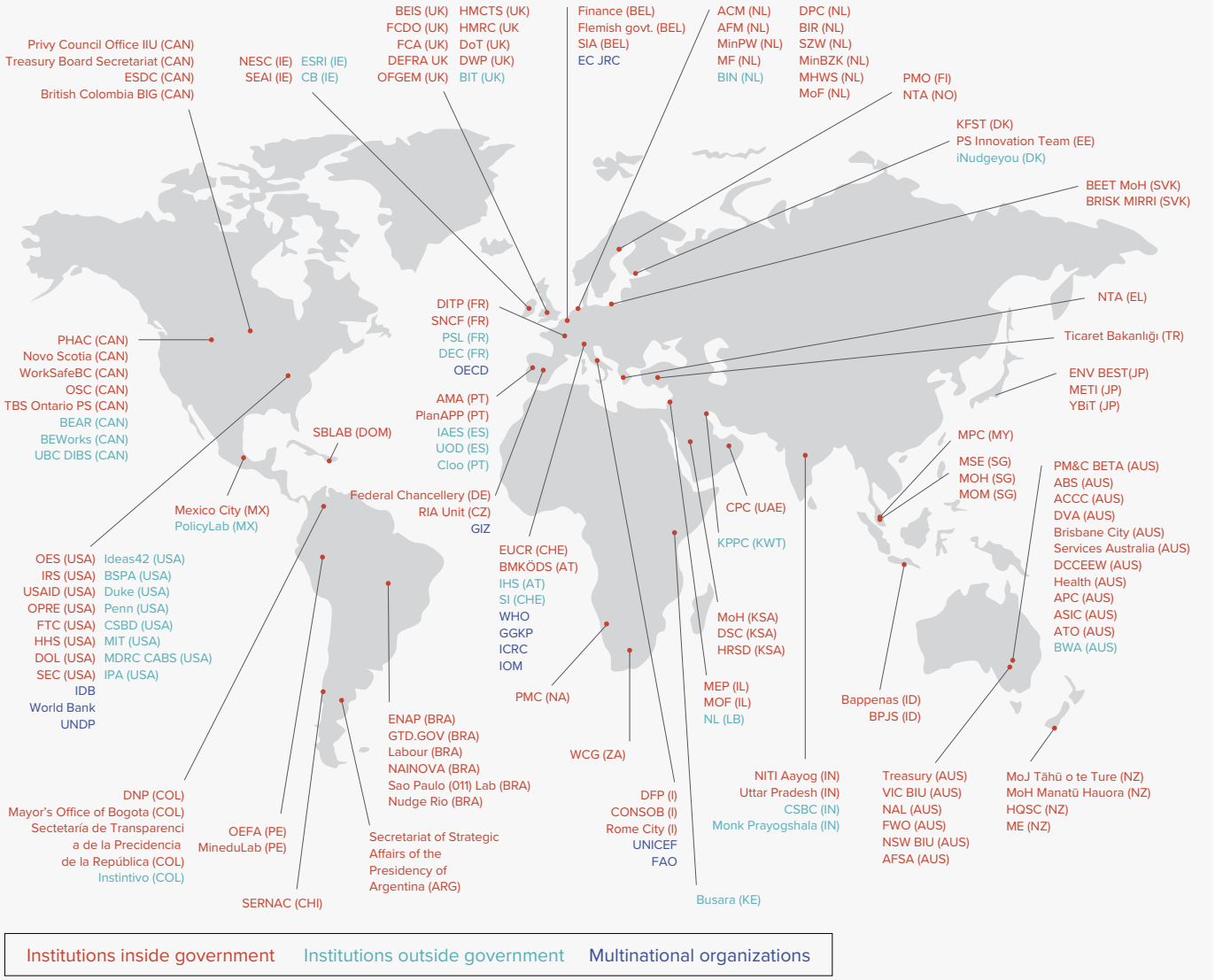
Some interventions informed by recognizing cognitive biases seek to directly change individual behaviour, many of them through nudges, which have had encouraging results in some domains.[99] Despite advocacy, including to support the response to the Covid-19 pandemic,[100] the uptake of insights derived from identifying cognitive biases has been mixed (box 4.5).

In fact, insights from behavioural science not only have to confront challenges associated with the replicability of several studies, but they also face a problem in the lack of an overarching theory that can account for the multiplicity of biases that are being documented (spotlight 4.2).[101] The relevance of behavioural science findings may lie less in providing descriptions of behaviour that is empirically more realistic, or a catalogue of ills to be addressed by nudges, and more as a guide to help decisionmakers achieve desired collective outcomes.[102] But that requires a framework to interpret how these biases interact with institutions and broader determinants of human behaviour (discussed thus far as if they were universal and hardwired, an assumption that will be relaxed in the next section).[103]

## Recognize how culture shapes behaviour and institutions

Behaviour during the Covid-19 pandemic illustrates the importance of having a broader understanding of behaviour that goes beyond selfish choice and behavioural insights and extends to an explicit consideration of the role of culture and its change over time (box 4.6).[104] There are many approaches to doing this, from sociologists interested in structuration to anthropologists interested in cultural economy and

**Figure 4.2 Widespread efforts draw on behavioural insights to inform public policy**




**Source:** Organisation for Economic Co-operation and Development, Observatory of Public Sector Innovation (https://oecd-opsi.org/blog/mapping-behavioural-insights/).

politics to approaches drawing attention to structural political economy.[105] Cultural evolution (spotlight 4.3) is one such approach among many that consider historical, social and relational perspectives that have been pursued across many disciplines. It is one way of accounting for how behaviour and culture interact in different societies and create packages adapted to address cooperative challenges at scale, with distinct cultural and behavioural traits.[106]

Insights from recognizing how behaviour and institutions are contingent on the social context and its change over time can be mobilized to address shared challenges, including the provision of such global public goods as climate change mitigation. These insights suggest that people can be expected to react differently to different interventions, as opposed to assuming that all people behave according to the standard selfish choice model or that they are all constrained by universal and hard-wired cognitive and other biases.[107] Another insight is that it is important to understand the interplay between social norm psychology and social identity to understand drivers of cooperation.[108] To see why and how, consider first the perils of interventions that start from either end of the behaviour–institution interaction (see figure 4.1).

## Box 4.5 The promise and peril of nudges in changing behaviour

Nudges attempt to change the choice context to increase the likelihood of people making decisions that contribute to meeting a policy goal without precluding any other choices or relying on economic incentives (for instance, changing default options on organ donations or retirement savings or framing incentives as losses instead of gains, given loss aversion).[1] Nudges not only seek to improve individual welfare but also tackle collective challenges, including green nudges to change behaviour towards climate and environmentally friendly choices.[2] They have the potential to increase the effectiveness of price-based interventions to mitigate climate change, such as carbon taxes, including by enhancing the public acceptance of taxes.[3] Thus, insights from behavioural science linked to cognitive biases are now regularly considered in the design and implementation of environmental policy[4] and in the provision of global public goods such as climate stability[5] and biodiversity conservation.[6] The potential to derive insights from behavioural sciences has been explored for enhancing the provision of global public goods within international law[7] and international relations.[8]

Once again, the question is the extent to which individual behaviour aggregates into biased aggregate outcomes. For instance, people may self-select or be sorted into groups with similar degrees of cognitive bias. If this is the case, some groups might deviate less, in the aggregate, from the selfish choice behaviour than others. Even with this type of sorting, whether biases matter depends on the decision being considered for accomplishing a certain task. For some tasks a group that gathers individuals who behave more according to the selfish choice model does not produce biased aggregate outcomes. But for other tasks biases can be amplified in the aggregate even when selfish choice and biased people are sorted into different groups.[9]

Understanding how and why this sorting matters for some tasks and not others is an important area for research. It is particularly relevant in the international context, where decisions on behalf of countries negotiating treaties are made by individuals empowered to represent those countries. In negotiations for climate change, negotiating peers perceive the credibility of country commitments to mitigate climate change to be determined by the quality of institutions in that country—with economic factors such as economic benefits and costs of those commitments bearing less on credibility.[10]

Whether decisionmakers are subject to biases is thus particularly important. It has been argued that decisionmakers among the elite may be less prone to biases and act more in line with the selfish choice model.[11] But this does not mean that they are not influenced at all by biases,[12] particularly when their decisions touch on issues salient in people's lives (such as climate change or management of a pandemic). Public opinion[13]—or, at a minimum, elites' perceptions of public opinion[14]—matters and is often conditioned by cognitive biases.[15]

There is an ongoing debate on the extent to which nudges and other behavioural interventions are effective.[16] In a study of 73 randomized controlled trials in 67 US cities implemented in collaboration with a national nudge unit, fewer than a third of the nudges were adopted in policy.[17] There are several barriers in translating insights from behavioural science into policy,[18] but recent debates on the size of the effects of interventions reported in the literature have further moderated policymakers' enthusiasm.[19] Information gathered from more than 200 studies reporting 440 effect sizes remains inconclusive.[20] There is also a difference between effects reported in small samples and effects realized when interventions are taken to scale. In 126 randomized controlled trials covering 23 million people, the average impact of interventions (that, is, at scale given the number of people covered) was 1.4 percentage points, compared with 8.7 percentage points in literature that typically relies on small samples.[21]

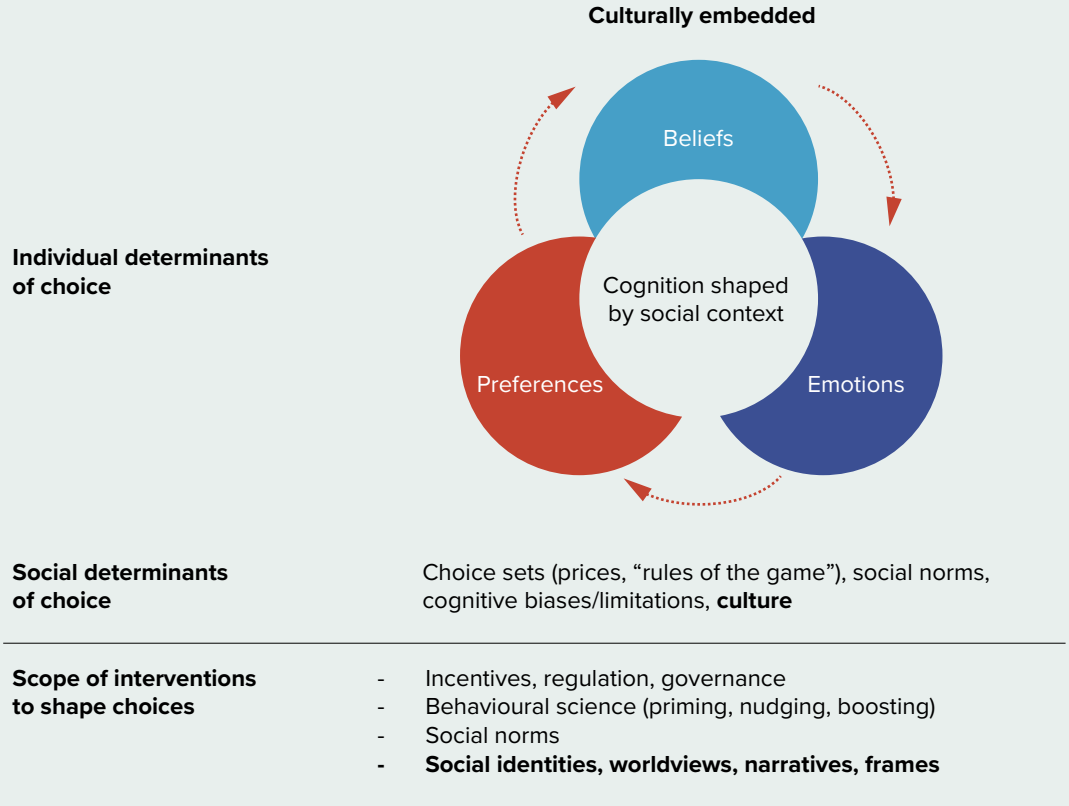
**Notes**

**1.** Thaler 2018; Thaler and Sunstein 2003. **2.** Carlsson and others 2021. Some green nudges are reportedly very effective. For instance, in China green nudges increased individuals' share of food orders with no cutlery (thus reducing plastic waste) more than sixfold (He and others 2023). **3.** Gravert and Shreedhar 2022. **4.** For a general review, see Carlsson and Johansson-Stenman (2012). **5.** See Brekke and Johansson-Stenman (2008) for a framework and early review. **6.** Travers and others 2021. **7.** van Aaken 2018. Although attention has also been drawn to some potential limitations. See the introduction to a symposium on this theme in van der Zee, Fikfak and Peat (2021) as well as Yildiz and Yüksel (2022). **8.** Davis 2023; Davis and McDermott 2021. **9.** Enke, Graeber and Oprea 2023. **10.** Victor, Lumkowsky and Dannenberg 2022. **11.** Hafner-Burton, Hughes and Victor 2013. There is also evidence that some elites appear to have weaker social preferences (Fisman and others 2015). **12.** As found in Mildenberger and Tingley (2019). **13.** Anderson, Böhmelt and Ward 2017; Oehl, Schaffer and Bernauer 2017. **14.** Hertel-Fernandez, Mildenberger and Stokes 2019. **15.** Webster and Albertson 2022. **16.** See, for instance, Dimant, van Kleef and Shalvi (2020), Guttman-Kenney and others (2023) and Bicchieri and Dimant (2022). **17.** DellaVigna, Kim and Linos 2022. **18.** As reviewed in Linos (2023). For a set of proposals on how to address some of the challenges, see Mažar and Soman (2022). **19.** And behavioural interventions have been unable to address some major challenges, such as how to reduce economic inequality (Ruggeri and others 2022). **20.** The original study by Mertens and others (2022a) reports a Cohen's *d* (a standardized measure of the difference between the mean of the untreated group and the treated group) of 0.43, but after reanalysing the data and correcting for publication bias (that only statistically significant results are published; more surprising results are more likely to be published) and heterogeneity (whether the findings extend beyond the sample used for the study), Szaszi and others (2022) find an effect of 0 (in a response, Mertens and others (2022b) agree with the importance of addressing issues associated with publication bias and heterogeneity). **21.** With respect to two nudge units in the United States: DellaVigna and Linos (2022) and Webster and Albertson (2022).

## Box 4.6 Social context shapes what people do and how they see themselves

Taking account of the interplay between behaviour and culture suggests that human choices need to be understood within the social context that shapes not only individual preferences and the architecture of choice but also how people see themselves and how they see others (box figure 1).[1] That is, who they are.[2] The entanglement of behaviour and culture opens the possibility of understanding the processes of social choice and the potential scope for interventions—in ways that recognize when and how they can be mutually reinforcing, as opposed to pinning all hopes on either institutions or behaviour to enhance collective action. This approach also implies recognizing that some processes of social choice change endogenously, so the interventions may have unintended consequences. Even if these are not possible to predict with precision, being aware of this possibility and understanding the mechanisms for them to emerge can enhance policy design and implementation.[3] This implies that criteria for the design of interventions (either behavioural or institutional) should consider efficiency and equity as well as efficiently evolving institutions to account for both a changing world and the endogenous dynamics of change between behaviours and institutions.[4]

**Box figure 1 Social context shapes who people are**




**Note:** The text in bold to the right of "Social determinants of choice" and "Scope of interventions to shape choices" denotes the new elements that are added to the selfish choice model and to the behavioural choice model (which remains relevant).
**Source:** Human Development Report Office elaboration based on Hoff and Stiglitz (2016).

**Notes**
**1.** Lamont 2023. **2.** This is where the discussion arrived, but it is the point of departure for sociology (Lamont 2019). This makes insights from sociology also relevant, particularly recent developments on understanding culture as a toolkit from which people draw cultural resources to navigate their life (Swidler 1986). These have been inspired in part by insights from the cognitive and behavioural sciences (DiMaggio 1997; Dimaggio and Markus 2010; Lamont and others 2017). These insights have already been incorporated into models and accounts of institutional change by economists (Acemoglu and Robinson 2022, 2023). **3.** Hébert-Dufresne and others (2022) present a model of this endogenous process of social choice. **4.** Schimmelpfennig and Muthukrishna 2023.

### Ignoring the interaction between behaviour and institutions is perilous

The perils of starting with institutions is perhaps more widely understood. Transplanting institutions[109] from the context in which they emerged to settings with different behavioural, social and economic contexts has been widely documented. Institutions, including legal institutions and mechanisms to enforce formal law, typically work in shaping behaviour if there is already an equilibrium resulting from an underlying set of beliefs that sustains cooperation.[110] This may very well be the case in an international context, except perhaps outside interactions where direct reciprocity fosters cooperation; some evidence suggests that the effectiveness of international agreements sometimes does not depend on whether there is an enforcement mechanism.[111]

Formal institutions set very important structural features of contemporary societies, so understanding flaws in those structural features (which can exacerbate inequalities in human development, perpetuate exclusion or impede collective action) and how to change them is crucial.[112] The contribution of the discussion in this section towards this goal is not to assume politics away or to minimize the importance of formal institutions but rather to probe how assumptions about behaviour also shape how those flaws are identified and what to do to correct them (box 4.7).[113]

But there also are perils in attempting to start from the other end, towards changing behaviour to foster collective outcomes, without taking into account the institutional and broad cultural context in which the changes are pursued. As noted, changes in behaviour can be pursued directly (creating nudges, for instance) or indirectly (making people change choices voluntarily based on their observation of others, particularly when social norms reach tipping points that make individual and social beneficial behaviours ubiquitous).[114] These processes can be mobilized to support the provision of global public goods.[115] Imitative adoption played a crucial role in the spread of solar panels in Germany around the 2000s, advancing from an initial slow adoption to a rapid spread that led the country to generate more solar power per capita than any other country by 2009.[116]

### Enhancing collective action requires understanding differences in preferences and beliefs shaped by social contexts

Interventions can trigger rapid shifts in social norms,[117] but identifying when and how tipping occurs requires understanding how preferences and beliefs are distributed across the population. Both preferences and beliefs can be shaped by cultural and social contexts, and ignoring differences can result in ineffective or, worse, misguided interventions (box 4.8).[118] Often, experimental studies draw on university students or segments of the population that may be more prosperous than average. There is also variation in the strength of behavioural effects across the population according to education and income (figure 4.3).

Variation also occurs across countries when effects based on one intervention were not observed when the intervention was implemented in an alternative way[119] or when interventions were explored across countries. These different outcomes point to the importance of recognizing how behaviours and institutions interact with culture.[120] It has long been recognized that signature findings of behavioural insights from experiments in high-income countries in Europe and North America are not generalizable, as shown by a failure to replicate the results in different contexts.[121] Moreover, over the course of human history and even today, most people have not lived in such settings,[122] implying the need for caution in generalizing claims from results based on samples from these settings.[123]

> “Both preferences and beliefs can be shaped by cultural and social contexts, and ignoring differences can result in ineffective or, worse, misguided interventions

Recent work uncovered substantial cultural differences in preferences and beliefs associated with economic inequality, supporting the notion that cultural processes are at play in shaping this diversity across and within countries.[124] For preferences on how much inequality people accept or are averse to, much depends on the kinds of inequality that people consider to be unfair.[125] Representative surveys in 60 countries documented variation across countries in the extent to which people subscribe to one of three views

## Box 4.7 Where are the politics?

One simplified way of identifying where the politics lie is to assume that there are two types of interactions relevant for collective action within countries.[1] One pertains to setting up the rules of the game—the conditions under which society is governed. This can be considered the realm of politics, which determines who holds power to do what and how.[2] These rules are codified in documents, from constitutions to civil and penal codes to jurisprudence (in short, the law). The creation, execution and enforcement of the law are ensured by formal institutions. Interactions of the other type then unfold within the law—the social and economic decisions undertaken by people and other formal institutions (those with legal status, such as firms or civil society organizations). Each of these realms is the subject of entire disciplines, including political science for the first, and much of economics for the second.

The two sets of interactions are mutually constitutive. For instance, rules can enable the accumulation of wealth and resources by certain agents that, in turn, can mobilize those resources to further advance their economic advantage in the domain of political interactions, through direct capture of political office, lobbying or the use of the media.

Still, as important as the law and rules are, there is a growing appreciation that contracts are notoriously incomplete (and externalities are pervasive), with the irreducible incompleteness of the law and formal institutions particularly relevant in contexts of uncertainty.[3] So, economic and social behaviour is also regulated in part by social norms in which the formation of beliefs and preferences and how they change over time and across people and countries have crucial importance.[4]

But behavioural assumptions, and the role of beliefs, matter even without assuming the irreducible incompleteness of the law. Why do people comply with the law, and how can social order be maintained in diverse societies? The selfish choice model suggests that people are motivated to seek individual gains and avoid losses, so these assumptions would suggest the use of strategies that deter law violators.[5] While these strategies matter, so do beliefs about the legitimacy of formal institutions: "Legitimacy is a concept meant to capture the beliefs that bolster willing obedience."[6]

Under this perspective people obey the law due in part to a common commitment to obey formal institutions, sustained by the belief that there is an obligation to obey (value-based legitimacy) that is then reflected in actual compliance (behavioural legitimacy). Within this framework antecedents to value-based legitimacy include components of how the formal institutions are perceived (motivations of leaders, administrative competence and the performance of formal institutions in delivering on their public purposes, including the provision of public goods) and views about procedural justice (whether the exercise of authority is perceived as fair). Within views about procedural justice, the perception that government procedures are unfair often motivates disobedience, evasion and resistance to legal demands, with deterrence motives overwhelmed and ineffective in these cases.[7]

The role of beliefs also comes to the fore when formal institutions undergo change. Fundamental institutional change often takes place during critical junctures when there is uncertainty about the shape that future institutions will take. A recent strand of literature shows that the dispersion of beliefs about future institutions can help identify these critical junctures. How these beliefs diffuse and get consolidated around particular views shapes in part the rules of the game that societies end up with.[8] Some of the evidence comes from contexts where people can choose to rely on formal state institutions or on nonstate entities (for instance, in dispute resolution) or where there are competing claims to the formal governance institutions, which shows that beliefs (about which arrangement is more effective or more enduring) not the formal institutions themselves causally determine behaviour.[9]

In sum, there is growing recognition of the importance of beliefs in shaping the two set of interactions and a recognition that they are shaped by the dynamic interaction between behaviour and institutions. Political scientist Margaret Levi titled a recent account of her intellectual journey "The Power of Beliefs."[10] And economist Kaushik Basu titled a deep reflection on the relationship between law and economics "The Republic of Beliefs" because "The might of the law, even though it may be backed by handcuffs, jails, and guns, is, in its elemental form, rooted in beliefs carried in the heads of people in society [...], creating enormous edifices of force and power, at times so strong that they seem to transcend all individuals, and create the illusion of some mysterious diktat enforced from above. In truth, the most important ingredients of a republic, including its power and might, reside in nothing more than the beliefs and expectations of ordinary people."[11]

**Notes**

**1.** Inspired by Hurwicz (1996), as described in Powers, van Schaik and Lehmann (2016), who distinguish the political game from the economic game. Above the political game Ostrom (2009b) posited a constitutional game. To simplify the discussion, the constitutional game is subsumed under the political game. **2.** See Powers, Perret and Currie (2023) for a discussion of how playing the political game in societies of increasing size leads to the emergence of political inequality. **3.** We are grateful to Charles Efferson for emphasizing these points. **4.** For discussions of how differences within countries on cooperative versus conformist preferences relate to differences in political ideology and how these differences may have emerged, see Claessens and others (2020) and Claessens and others (2023). For an account of the diversity across 99 countries in the (lack of) correlation between cultural and economic conservatism, see Lelkes, Malka and Soto (2019). **5.** The framing and discussion in this paragraph draw from Tyler (2023). Deterrence is typically understood to mean punishing violators as a means to enhance compliance, but rewarding a commitment to cooperate could also be effective (Han 2022). **6.** Levi, Sacks and Tyler 2009, p. 354. **7.** Levi, Sacks and Tyler 2009, p. 360), with numerous examples, including several related to tax avoidance and evasion. For further elaborations related to the need to raise fiscal revenue to provide for public goods, see Levi (1988, 1999). For a debate on the relevance of procedural justice, see, for instance, Hagan and Hans (2017). **8.** Reviewed in Callen, Weigel and Yuchtman (2023). **9.** Acemoglu and others 2020. **10.** Levi 2022. **11.** Basu 2018, p. 40.

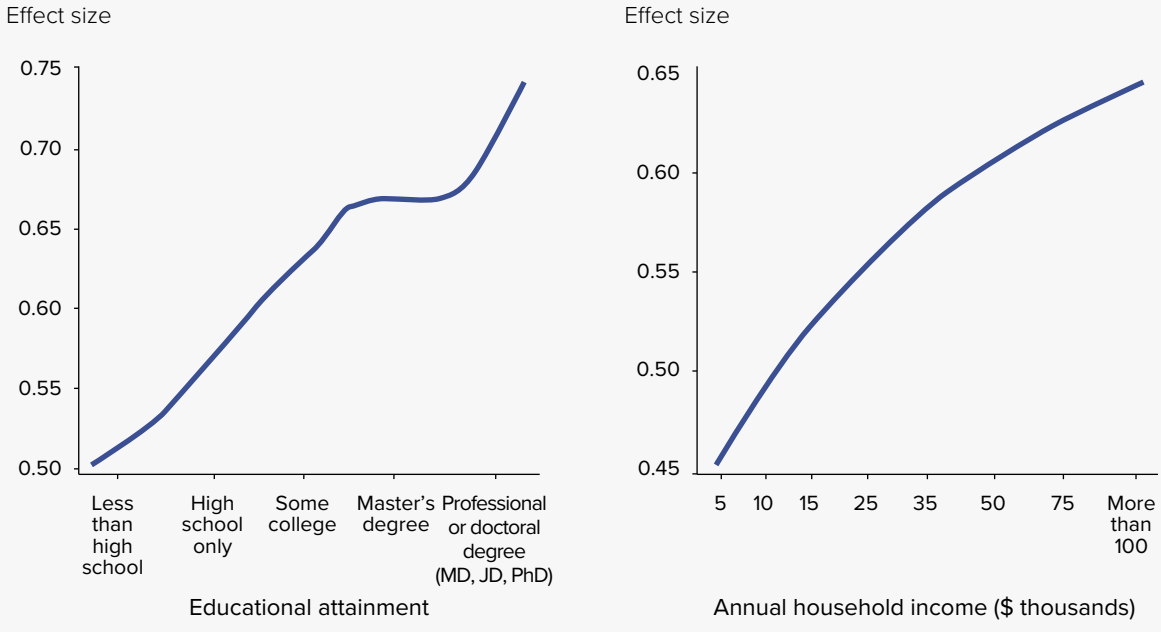
## Box 4.8 It seemed such a good idea at the time: The dangers of ignoring heterogeneity when pursuing social tipping

An intervention heralded as very successful in harnessing the potential of social norms to change behaviour was the firm Opower's provision of information about how each customer's energy use compares with that of its neighbour, along with messaging that signalled that conserving energy was a desirable goal.[1] An initial evaluation of 600,000 households that compared the behaviour of households that received the information with that of households that did not found that this nonprice intervention had a substantial effect in encouraging energy conservation.[2] However, when the intervention was scaled to more than 8 million people, the average effect—and its practical importance—turned out to be much lower than in the initial evaluation.[3]

This was not a replication failure, since both evaluations were rigorous and stood up to independent analysis.[4] But the initial evaluation was based on the communities that were the first to adopt the measure. They were already inclined to value energy conservation, had large homes and were relatively prosperous, thus they had many opportunities to conserve energy. The effect of the intervention declined substantially when it was expanded to include people with a broader set of beliefs and much wider range of incomes. Even when studies are carefully conducted, the choice of convenience samples seems to be particularly problematic in behavioural interventions.[5]

**Notes**
**1.** Featured, for instance, in Chetty (2015). Thus, the intervention relied on both descriptive and injunctive social norms. See Constantino and others (2022) for a discussion and Bhanot (2021) for the role of injunctive social norms in promoting conservation. **2.** Allcott 2011. An initial smaller expansion beyond the 600,000 also suggested that the effects persisted (Allcott and Rogers 2014). **3.** Allcott 2015. **4.** As reported in Bryan, Tipton and Yeager (2021), which inspires the analysis in this paragraph. **5.** Sometimes simply because there is not enough contextual information, as Vivalt (2020) showed in an analysis of 635 studies of impact evaluations of development interventions, posing challenges to the generalizability of results.

## Figure 4.3 Effects of several behavioural phenomena are stronger in more educated and wealthier segments of the population



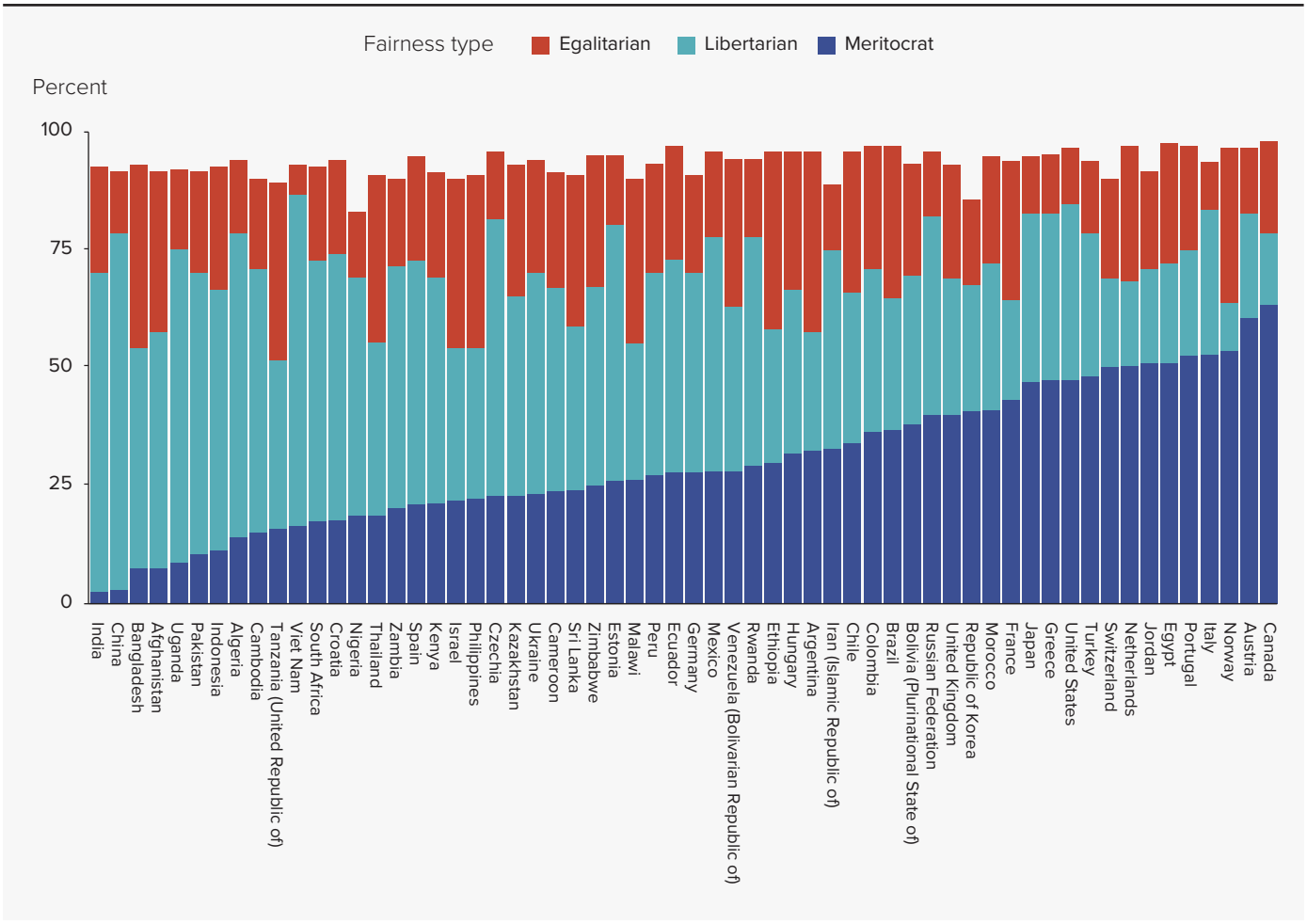

**Note:** The figures show the effects from seven classical studies in behaviourial science (conformity with a descriptive social norm, impact of argument quality on persuasion, base rate underutilization, conjunction fallacy, underappreciation of the law of large numbers, false consensus and ease of retrieval) according to educational attainment and annual household income in a representative sample of the US population. The vertical axis reports Cohen's *d* effect sizes (the standardized difference between the means of the treatment and control groups; Cohen 1988).
**Source:** Yeager and others 2019.

on fairness: egalitarians (find all inequalities unfair), meritocratic individuals (accept inequalities due to differences in performance as fair but those due to luck unfair) and libertarians (accept all inequalities as fair). A large share of the population in several high-income countries adhere to the meritocratic view—which is also the foundation for some normative theories of distribution—holding that people should not be considered responsible for outcomes beyond their control (figure 4.4).[126]

But this view is not well represented in many other countries. And even among countries with similar shares of people holding a meritocratic view, there are large differences in the other two categories. For instance, although Norway and the United States have similar shares of meritocratic individuals, the United States has a much larger share of libertarians, and Norway has a much larger share of egalitarians.[127] Moreover, there are differences within societies. In Norway the share of egalitarians is much higher among 15-year-olds from low socioeconomic households than among those from high socioeconomic households,[128] and while most grade 5 children are egalitarians, the meritocratic share increases in higher grades and is largest by grade 13.[129]

So, experience and social context shape people's views of fairness, again pointing to cultural processes in shaping preferences over the lifecycle. Separately from preferences, what people believe about the sources of inequality also matters. A meritocratic individual who believes that inequality is driven by luck, not effort, would find inequality unfair. As with preferences, there is a wide disparity across and within countries on beliefs about the drivers of inequality.[130]

**Figure 4.4 There are widespread differences in fairness preferences around the world**




**Note:** Egalitarians find all inequalities unfair, meritocratic individuals accept inequalities due to differences in performance as fair but those due to luck as unfair and libertarians accept all inequalities as fair.
**Source:** Almas and others 2022.

Beliefs about the extent to which inequality is unfair matter more for attitudes towards redistribution than the actual level of income inequality, so beliefs have a direct bearing on support for different types of redistribution policy.[131]

Thus, recognizing that populations can be heterogeneous in preferences and beliefs and how these differences emerge from cultural processes is crucial to the design of institutions and policies, including, for instance, on tax compliance (spotlight 6.4).[132] In particular, it is essential in assessing what kind of interventions are more likely to trigger social tipping.[133]

### Understanding how enhancing agency and redressing polarization within countries can improve collective action across countries

A broader perspective on choice informed by these insights opens new vistas on how to advance cooperation and the provision of global public goods. And it helps show how risks associated with domestic patterns of political polarization may harm collective action across countries.

Interventions to enhance the provision of global public goods that are informed by the recognition that people are products of culture include the consideration of perceptions and aspirations when implementing policies or designing institutions. Different perceptions about how to interpret a noncooperative choice can result in cultural impediments to cooperation: when the choice is perceived as a mistake, it can lead to collective action in future interactions, but when it is perceived as an insult, it can result in the collapse of collective action.[134] Perceptions also matter when people infer the motives of others to make moral judgements,[135] and on perceptions about how (and in what ways) they are interdependent with others.[136] Breakdowns of cooperation in conflict are also shaped by this type of perception. There is evidence that the mental representation of payoffs that potential conflicting parties face rather than the actual payoffs determine not only how people think but also how they behave.[137] These perceptions are malleable to some extent and can be changed in ways that increase the propensity of players to pursue cooperation.[138]

Aspirations also matter because people act on what they believe is possible and desirable, and these beliefs are in part the result of social processes, shaped by narratives widely shared across society or within groups.[139] Aspirations, and the institutions and social norms associated with them, may have emerged as a result of cultural processes that made them suitable for some time in some contexts, but they may no longer be suitable for new contexts.[140] This mismatch acquires a novel dimension as we face the unprecedented challenges of the Anthropocene, in which it is unclear how processes of cultural variation and selection across societies that shaped in part how adaptive institutions and norms emerged would work when confronting planetary-scale challenges: they have to be addressed collectively and at a global scale because the relevant group is all of humanity.[141]

Such a mismatch can be characterized somewhat as reflecting what Karla Hoff and Allison Demeritt called an agency gap, which can be fuelled in part from a divergence between what societies believe is possible or probable and what is objectively possible.[142] To the extent that an agency gap is the result of widely shared beliefs, closing the gap will require more than providing information; it will also require mobilizing insights about the cultural determinants of the formation of shared beliefs.[143]

“Interventions to enhance the provision of global public goods that are informed by the recognition that people are products of culture include the consideration of perceptions and aspirations when implementing policies or designing institutions

Narrowing the agency gap is constrained by what is objectively possible but is malleable with respect to what people aspire to, which is sometimes articulated through narratives that have “political and psychological agency and can reinforce or challenge existing power relations and trajectories.”[144] This can take the form of what has been called the pursuit of emancipatory transformations,[145] which affirms the importance of enhancing not only people’s welfare but also their empowerment as agents of change.[146]

But it is one thing to recognize that perceptions and aspirations matter, and that broad recommendations such as reframing narratives can help close the agency gap, and quite another to see how to mobilize these insights. Here is where the concreteness of providing

global public goods can help, because global public goods correspond to a very specific way of addressing shared challenges when countries face interdependence. Global public goods, by their nature, correspond to non-zero-sum interactions and can thus be mobilized to overcome the psychology of zero-sum beliefs (one party's gain comes at the expense of the other party's losses). To be sure, many interactions across countries are zero sum, but pursuit of the provision of global public goods has the potential to open spaces for countries to interact that are not zero-sum.[147] Emphasizing the provision of global public goods can overcome three of the channels shown to elicit zero-sum beliefs (even in situations where the actual payoffs are not zero-sum): perceptions of threat, real or imagined resource scarcity and inhibited deliberation.[148] Zero-sum beliefs exacerbate conflict,[149] discourage cooperation[150] and suppress effort and economic development.[151]

"While diversity of beliefs and preferences can be harnessed for creativity and innovation, patterns of political polarization represent a major challenge for collective action

The provision of global public goods can mobilize the human ability of shared intentionality: "an understanding that individuals are solving a problem together and are committed to supporting each other."[152] In fact, understanding and sharing intentions have been argued to have evolved to account not only for joint actions and shared beliefs but also for the emergence of coordination on the need for giving reasons to justify those actions and beliefs.[153] Some evidence suggests that the pursuit of self-reliance (seeking to reduce interdependence) in confronting shared problems crowds out cooperation and exacerbates inequalities.[154] People are also able "to see the world from another individual's perspective and, specifically, to understand and formally represent another individual's knowledge states, beliefs and goals"[155] and even their emotional states, which is involved in empathy.[156] This can engender a proclivity for the pursuit of justice that, along with shared intentionally, can be a powerful driver for cooperation to enhance the provision of global public goods.[157]

The flip side is that these powerful potential drivers of cooperation often act within groups.[158] One manifestation of this "groupy" behaviour is the virtually universal higher levels of parochial (meaning within countries) cooperation than of universal cooperation.[159] While diversity of beliefs and preferences is not just a fact but something that can be harnessed for creativity and innovation,[160] patterns of political polarization (where no common factual foundation exists to undertake reasoned discussions and where groups alienate and even dehumanize each other) represent a major challenge for collective action (chapter 6).[161] How political polarization plays out domestically can be a central determinant of providing global public goods such as climate change mitigation and pandemic response.[162]

Zero-sum beliefs have been associated with political polarization in some countries.[163] For instance, along with international inequity in vaccine access, domestic attitudes towards vaccines determined the path of the Covid-19 pandemic, including in high-income countries.[164] Even when vaccines were free and plentiful, patterns of trust and political polarization shaped the course of the pandemic in many countries.[165] In some countries people's vaccine status identification is as polarizing as their other group identifications. More polarized attitudes towards vaccine status have been linked to greater resistance to vaccine uptake.[166]

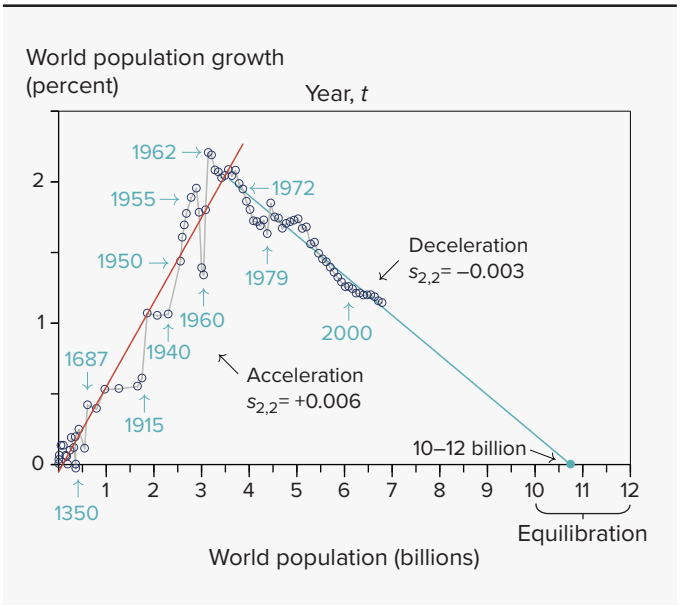
Political polarization matters also because the heterogeneity of preferences and beliefs and their cultural underpinning may prevent the mobilization of social norms towards more cooperative outcomes.[167] As the discussion above illustrates, and as demonstrated theoretically,[168] the distribution of social preferences (towards caring for the environment or aversion to inequality) can activate tipping, have no effect or even produce a backlash, depending on the effects that interventions have in different population groups and the reasons people within those groups adhere to social norms.

When behaviour conforming with a norm interacts with other motives, such as group identities, social tipping may not occur at all.[169] When belonging to a group is linked with salient identities, that can exaggerate the "othering" of other groups and blind members to the realization that everyone has multiple identities with different expressions and relevance at different times—people can lose sight of our shared humanity.[170] This is how in these situations behavioural markers that people rely on to signal group

affiliations may prevent social norms from tipping.[171] For example, if not being vaccinated against Covid-19 is a marker of belonging to a group, not only will behaviour not change when some members of the group are vaccinated, their being vaccinated can turn behaviours against vaccination to signal commitment and loyalty to the group.[172] A crucial aspect to consider, particularly in politically polarized contexts, is not only people's private beliefs but also their beliefs about what others think about certain issues, the perceptions they have about threats and how they believe that others think about them and how they will behave.[173]

The next two chapters consider these two challenges (narrowing the agency gap and redressing polarization) in more detail. It may seem that the current context of turbulence around the world is not conducive to meeting either challenge. Yet, as we move deeper into the Anthropocene, we may already be experiencing a major ecological discontinuity[174] characterized by a shift from uncontrolled population growth to controlled fertility (figure 4.5).[175] The transition to low fertility is complex and multifaceted and has recently been analysed from the perspective of cultural evolution (to consider factors that demographers designate as ideation).[176] Determinants of this transition include innovations in medicine and sanitation, empowerment of women, advances in education, shifts in social norms about the size of successful families, increasing attention to population growth, consciousness of planetary challenges and many other potential factors, all of them expressions at least in part of cultural factors.[177] Recognizing that we are in the new planetary context of the Anthropocene and in a novel ecological phase suggests a possibilist agenda: not optimism or pessimism but the possibility of consciously managing the self-inflicted problems that we are confronting on a global scale.[178] The provision of global public goods, which depends only on us, would be a way of acting on that possibilist agenda.

**Figure 4.5 The world is undergoing a major transition from accelerating to decelerating population growth**




**Note:** The graph plots per capita growth of the population as a percentage against the population level. The gray line connects data points in different years. The red line fits an ecological model of mutualistic interactions between humans and plants and animals in which the ecological parameter ($s_{2,2}$) is positive, signifying very rapid population growth. The blue line fits a model where the parameter has turned negative, signifying a deceleration in population growth, with a projected equilibrium in population at around 10–12 billion people some time in the next century.
**Source:** Lehman and others 2021.

SPOTLIGHT 4.1

# A technology-centred approach to climate change negotiations

**Scott Barrett,** *Columbia University and London School of Economics*

The world has been trying to limit climate change for more than 30 years. The first agreement, the United Nations Framework Convention on Climate Change (UNFCCC), was adopted in 1992. Under this framework, parties agree to cooperate to limit concentrations of greenhouse gases in the atmosphere to a level that would avoid "dangerous" climate change. Every country is a party to this agreement. All countries agree that they need to cooperate.

Since then, two other treaties have been adopted. The Kyoto Protocol of 1998 set binding emissions limits for Annex I countries[1] for 2008–2012, but these could not be enforced. The United States declined to participate. Canada participated initially but took no steps to meet its emissions limits and later withdrew in order to avoid a legal obligation to comply. In 2009 countries met in Copenhagen to broaden and deepen the Kyoto Protocol. More countries were expected to be bound by emissions limits, and previously negotiated emissions limits were to be tightened. However, countries were unable to agree on how to do this. Instead, they pivoted. First, they put a number on the UNFCCC's goal of avoiding dangerous climate change, specifying a 2°C target for mean global temperature rise. Second, they asked parties to pledge contributions towards meeting this common goal. Ultimately, this approach was codified in the Paris Agreement of 2015. That treaty strengthened the previous goal: countries are now to hold "the increase in the global average temperature to well below 2°C above preindustrial levels and [to pursue] efforts to limit the temperature increase to 1.5°C." The Paris Agreement also situated pledges in the context of naming and shaming, to encourage greater ambition. Unlike the Kyoto Protocol, all countries participate in the Paris Agreement. However, also unlike the Kyoto Protocol, pledge-making and achievement of pledges are explicitly voluntary.

Where has this approach gotten us? Concentrations of carbon dioxide have risen every year since negotiations began. More carbon dioxide has been emitted since the UNFCCC was adopted than in the previous 250 years. Carbon dioxide emissions reached an all-time high in 2022. The world is not on course to meet the goal countries have said they must meet.

Why? No phenomenon of this complexity has a single explanation, but one stands out, and it is surprisingly simple: countries are caught in a prisoner's dilemma. All countries recognize that they would all be better off if they all reduced their emissions, eventually to net zero. But each country fears that doing this would put its economy at risk. Each might be willing to reduce its emissions substantially if assured that others will reduce their emissions substantially and thus avoid dangerous climate change. However, when contributions cannot be enforced or are voluntary, this assurance eludes every country. The problem is not that every country does nothing; it is that every country does too little.[2]

How to do better? It is instructive to consider some things that have gone well.

## Successes

One success is the 99 percent drop in the price of solar photovoltaic cells since 1976. Public and private research and development account for 59 percent of the drop, economies of scale for 22 percent and learning by doing for 7 percent.[3] Research and development were particularly important early in the process; economies of scale became important later. The history of solar research and development can be traced from the first solar cell developed at Bell Labs in the United States in 1954 to further developments spurred by the Space Race; the US response (beginning with President Richard Nixon's Project Independence, a programme to make the United States energy independent by 1980) and Japan's response

(especially its Sunshine Program) to the energy crises of the 1970s; research in Australia in the 1980s; and the solar boom in Germany in the 2000s, helped by generous feed-in tariffs.[4] It took the combined efforts of multiple countries to get to today's situation, where costs are so low that, according to the International Energy Agency, solar photovoltaics are "becoming the lowest-cost option for electricity generation in most of the world."[5]

Another success is the decline in the price of lithium-ion batteries. Since commercialization began in 1991, the cost of this form of energy storage has fallen 97 percent. Public and private research and development account for 54 percent of the drop, economies of scale for 30 percent and learning by doing for 2 percent.[6] Most of these activities have been undertaken by the electronics industry (mobile phones, notebook computers, power tools and so on).[7] Advances in this technology, combined with policies to promote demand, have propelled a rapid increase in electric vehicle sales, particularly in China, the European Union and the United States. Globally, lithium-ion battery demand for electric vehicles increased 65 percent between 2021 and 2022.[8] Thanks to this technology, an increasing number of countries and vehicle manufacturers plan to phase out sales of internal combustion engines by 2035.

These developments (and others, such as the falling costs of wind turbines and light-emitting diode bulbs) took place outside the UNFCCC process and arguably had little to do with the climate negotiations. Negotiators have asked countries to reduce their emissions, an approach that falls into the trap of the prisoner's dilemma. Had countries focused more on changing the economics of new technologies, the outcome might have been different. Rather than ask countries to reduce their use of fossil fuels, why not focus on making alternative fuel sources cheaper than fossil fuels? Doing this practically guarantees the global spread of new technologies, reducing emissions everywhere.

## Tipping

Solar photovoltaics and battery-powered electric vehicles have spread (so far) without the help of a treaty. They are examples of cascade effects.[9] Once enough research and development have been undertaken to encourage uptake of a technology by one country, that country's production lowers costs for all, mainly through economies of scale, encouraging uptake by other countries. Their uptake in turn lowers costs further, encouraging even more countries to adopt the technology, and so on. Cascades generate positive feedback.

In some cases a single country may be unable to kick-start a cascade, but a critical mass of countries may be able exert the "big push" required for tipping. Network externalities often drive universal adoption. An example is ocean shipping of oil. Historically, most oil pollution in the seas resulted from the way oil was transported. After completing an oil delivery, a tanker would take on ballast water for the return journey. Before picking up its next load, the tanker would release its ballast water, mixed with oil residue, into the sea. This process, repeated over and over, was a major source of ocean pollution. To limit this pollution, the 1954 International Convention for the Prevention of Pollution of the Seas by Oil (OILPOL) required tankers to limit their release of ballast water to an area at least 50 miles from shore. Being a prisoner's dilemma, however, OILPOL like the Kyoto Protocol, was difficult to enforce.

In the 1970s a different approach was tried. The International Convention for the Prevention of Pollution from Ships (MARPOL) required that oil tankers separate the tanks that hold oil from the tanks that hold ballast water, necessitating tanker redesign. MARPOL's approach was more costly than OILPOL but easier to enforce.[10] Once enough ports denied entry to tankers of the old design, more tanker operators met the new standard, and as more tankers met the new standard, more ports allowed entry only to tankers that met the new standard to protect their coastlines. In this situation there was a tipping point for participation that guaranteed universal adherence.[11] Intuitively, the tipping point would need to be at least 50 percent of all shipping, and in practice, this turned out to be the threshold adopted for entry into force of the agreement mandating the new technology standard. According to the International Maritime Organization, "MARPOL has greatly contributed to a significant decrease in pollution from international shipping and applies to 99% of the world's merchant tonnage."[12]

The Kyoto Protocol wisely exempted emissions from international aviation and shipping, believing that these sources ought to be controlled through the International Civil Aviation Organization and the International Maritime Organization. These are essentially standards organizations, the appropriate bodies to negotiate emission reductions in their spheres of influence. By focusing on standards rather than emissions limits (which are, in any event, difficult to assign to individual countries), these organizations can stimulate positive feedback, causing a new standard to tip.

Suppose that the best alternative to bunker fuel turned out to be green ammonia, a fuel produced by combining nitrogen extracted from the air with hydrogen extracted from water, both processes powered by renewable energy. How to proceed? Ammonia would likely cost several times as much as heavy fuel oil. A switch to ammonia would also present technical challenges. It would require new engines, new onboard storage tanks (necessitating new ship designs) and new port facilities: in short, a new technology–fuel standard. A switch to ammonia clearly would not happen one country at a time. Vessel owners would not want to run their ships on ammonia unless a network of refuelling infrastructure were available, just as no country would want to build an ammonia fuel network unless assured that lots of ships would run on ammonia. Lock-in would be a barrier to switching if only one or a small number of countries switched. But as more ports switched to ammonia, more ship owners would want their vessels to run on ammonia, and as more ships ran on ammonia, more ports would want to switch. Tipping of a standard for green ammonia would resemble the experience with MARPOL.

Mission Innovation, a coalition of 22 countries working outside the UNFCCC process, has a plan to reduce emissions in shipping that obeys the logic sketched out above. A first goal is to undertake research and development to identify the best alternative to heavy fuel oil. A second goal is to facilitate the spread of this new technology–fuel standard. Again, suppose that the research and development undertaken in the first stage revealed ammonia to be the "winner." How to achieve the second goal of ensuring global spread of the new standard? Mission Innovation would aim to establish a fleet of at least 200 ships able to run on the new fuel; to build a "global port infrastructure to support vessels operating on zero-emission fuels so that by 2030, 10 large trade ports covering at least three continents supply zero-emission fuels";[13] and, finally, to scale up production of the new fuel so that it supplied at least 5 percent of the total market. It is unlikely that 200 ships, 10 large ports and a 5 percent share of the fuel market would suffice to tip the global market, but at least this initiative sees the logic of needing to change the system. Changing the system is the essence of a strategy that seeks to transform the prisoner's dilemma into a tipping game.[14] Once critical mass gets past the tipping point, such an approach generates positive feedback, leading to a global switch, as we saw with MARPOL.

## Trade

The approach pursued by the UNFCCC, focusing on emissions reductions, generates negative feedback. If one country (or group of countries) reduces its emissions unilaterally, comparative advantage in greenhouse gas–intensive sectors shifts to other countries, causing their emissions to increase—a phenomenon known as trade leakage. Also, if the emissions reductions are achieved by lowering fossil fuel use, global prices for these fuels will fall, causing other countries to increase their consumption and, thus, emissions. This negative feedback intensifies the incentive to free ride, which is inherent in the prisoner's dilemma.

Because of these trade-related concerns, domestic climate policies often exclude greenhouse gas–intensive industries from having to reduce their emissions—undermining unilateral efforts to reduce emissions. The European Union is planning to extend its emissions trading arrangements to previously protected industries in order to reduce emissions further. However, due to concerns about leakage, it is planning to replace the exclusions with industry-specific carbon border adjustment mechanisms—a move that may stimulate conflict. As happened previously when the European Union tried to extend its emissions trading system to international aviation, other powerful states may retaliate. Also, developing countries may protest that, by treating domestic production and imports alike in terms of emissions, border tax adjustments violate the principle of common

but differentiated responsibilities enshrined in the UNFCCC. Finally, correcting for leakage will not correct for free riding.

However, linking climate agreements to trade cooperation can help prevent free riding—and, in the process, prevent leakage. Trade agreements are easier to enforce than climate agreements. This is because trade is bilateral, whereas emissions reductions are a global public good. If a country violates a trade agreement, the countries harmed by the reduction in trade have a strong—almost built-in—incentive to retaliate. The fear of retaliation motivates countries to abide by their trade agreements. By contrast, if a country emits more than allowed by a climate treaty, other parties to the agreement harm only themselves by reciprocating—and so will not retaliate. Because trade agreements are easier to enforce, linking trade cooperation to cooperation in supplying a global public good may overcome free riding incentives.[15]

The prime example is the Montreal Protocol, which protects the stratospheric ozone layer. The treaty bans trade between parties and nonparties in chlorofluorocarbons (CFCs) and products containing CFCs and works as follows. If no other countries participated in the agreement, no country would want to participate because doing so would mean losing all gains from trade in CFCs in addition to losing out from free riding. However, if all other countries participated, any country would want to participate so long as the gains from trading with the rest of the world exceeded the gains from free riding. Intuitively, every country would have an incentive to participate provided enough others participated. Trade measures thus imply the same kind of tipping point as with MARPOL—a result that makes sense when considering that denying a vessel access to a port is equivalent to a trade ban. Thanks partly to the trade measure, the Montreal Protocol has been remarkably effective, preventing both leakage and free riding.[16]

In Kigali in 2016, the Montreal Protocol was amended to control hydrofluorocarbons (HFCs), a cousin of CFCs. Since HFCs are a powerful greenhouse gas and do not deplete the ozone layer, the Kigali Amendment is a climate treaty negotiated outside the UNFCCC process. Further, because Kigali incorporates the same trade measure as the Montreal Protocol, it represents the first climate treaty to incorporate a trade measure. The Kyoto Protocol was unable to control HFCs, but the Kigali Amendment will very likely do so, especially after its trade measure enters into force in 2029. Also, unlike unilateral policies to control for trade leakage, the Kigali Amendment incorporates a side payment mechanism to cover the incremental costs of developing countries' compliance with the treaty's control measures. The Kigali Amendment promotes cooperation in the same spirit as the UNFCCC, only by a different means.

## Way forward

For all its efforts the UNFCCC approach to limiting climate change has so far fallen short of achieving its goals. Fortunately, the Paris Agreement can be complemented by other agreements aimed at reducing emissions in particular sectors. Indeed, this has already happened. The Kigali Amendment was adopted less than a year after the Paris Agreement. Other developments, including the falling prices of solar photovoltaics and lithium-ion batteries and the aspirations of Mission Innovation, hint that more progress is possible. The key feature shared by all these efforts is their focus on interventions (technical standards, research and development, and trade measures) that can transform systems by achieving critical mass.[17] Surely, many more such possibilities remain to be discovered.

## NOTES

1. Annex I countries include the industrialized countries that were members of the Organisation for Economic Co-operation and Development in 1992, plus countries with economies in transition, including the Russian Federation, the Baltic states and several Central and Eastern European states.
2. See Barrett and Dannenberg (2016) for a laboratory experiment of the Paris Agreement showing that the process of "pledge and review" changes what players say (meaning their collective target and individual pledges) but not what they do (meaning their actual contributions to achieving their pledges and target).
3. Kavlak, McNerney and Trancik 2018.
4. Nemet 2019.
5. https://www.iea.org/reports/solar-pv.

6. Ziegler, Song and Trancik 2021.
7. Dugoua and Dumas 2023.
8. IEA 2023b.
9. Dixit 2003; Heal and Kunreuther 2010.
10. Mitchell 1994.
11. Barrett 2003a.
12. https://www.imo.org/en/ourwork/environment/pages/pollution-prevention.aspx.
13. https://explore.mission-innovation.net/mission/zero-emissions-shipping/.
14. Of the world's 10 biggest ports by volume, 7 are in China. China's participation in a strategy to change shipping is essential.
15. Barrett and Dannenberg 2022.
16. Barrett 2003a.
17. Barrett 2016.

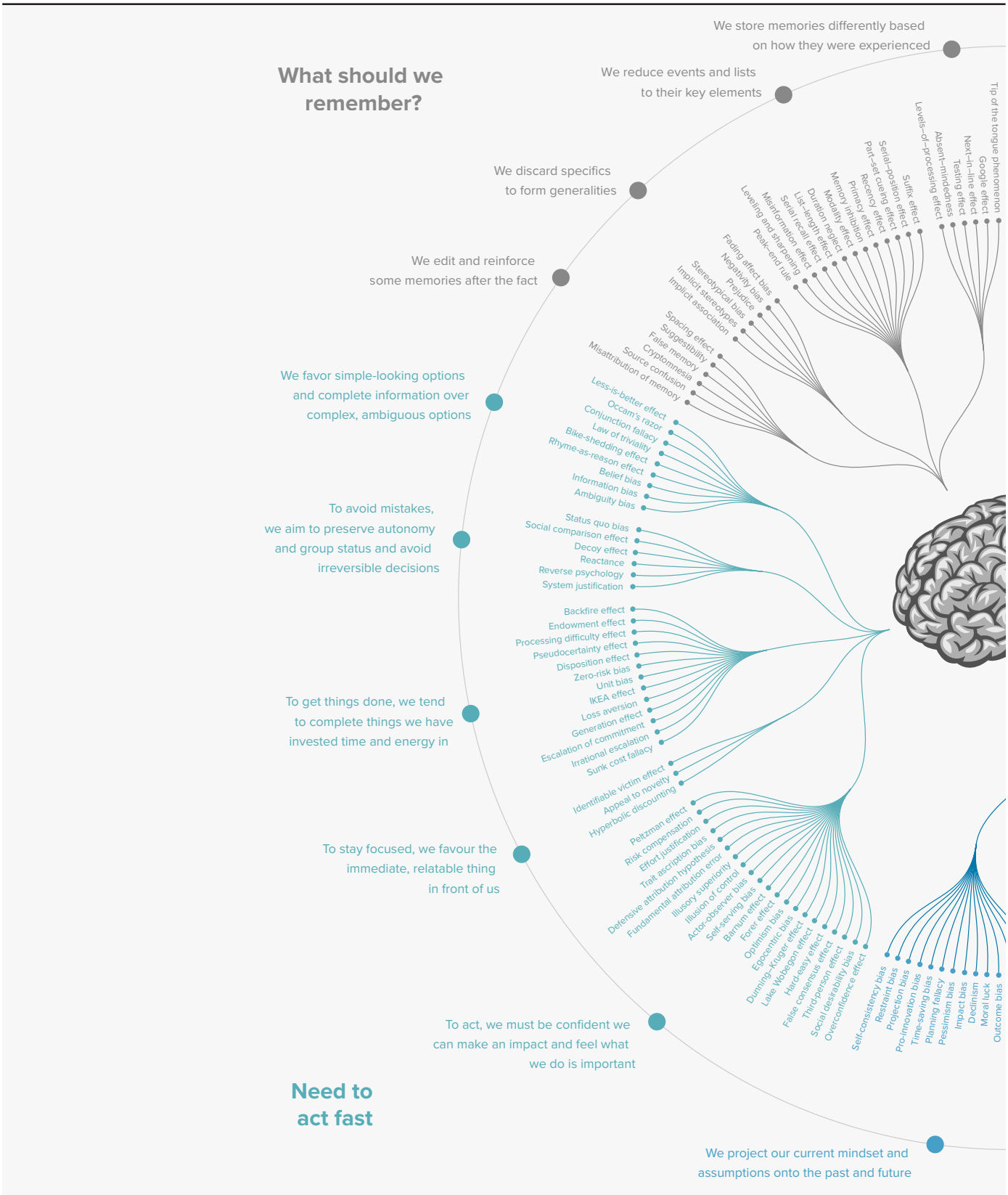
SPOTLIGHT 4.2

# Using insights from behavioural science: Watch out!

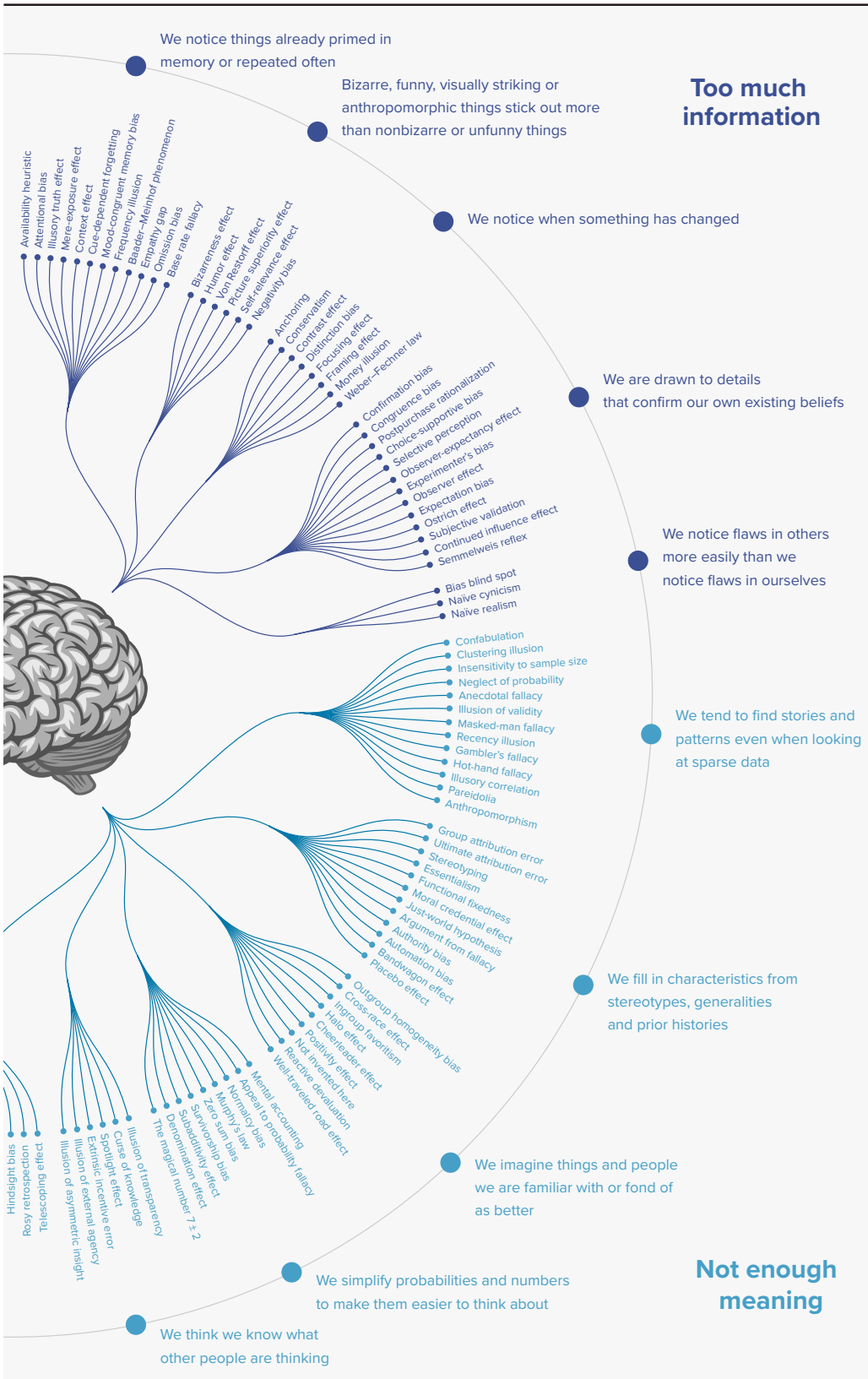
Caution in the use of behavioural insights is associated with challenges in replicating some findings.[1] Such concerns follow on the crisis of replicability that affected some psychology research in the 2010s, when several high-profile findings that garnered media and policy attention failed to be replicated in subsequent attempts.[2] In particular, studies over the past 20 years based on experiments failed to replicate at higher rates than nonexperimental studies.[3] A recent review found that only two-thirds of social science experiments reported in two top journals were replicated, and the average effect size was about half of that reported in the original studies.[4] One of the signature nudge interventions—making organ donations the default—failed to replicate and could even be counterproductive.[5] Several efforts have documented not only failures to replicate but also potential scientific misconduct.[6] Learning from these challenges, there is awareness that behavioural science will likely evolve to deliver more robust findings, be more cautious on claims based on statistical inference and address issues of more direct policy relevance.[7]

But insights from behavioural science confront another challenge. Given the proliferation of cognitive biases identified in the literature, even if findings are robust, it is challenging for interventions to address them all or to ensure that addressing one bias does not exacerbate some other bias. The cognitive bias codex (figure S4.2.1) may appear as little more than "a trove of plausible ad hoc modifications to rational choice models."[8] This challenge has motivated efforts to find a set of restricted causal mechanisms that could account for a large set of cognitive biases.[9] A better understanding of cognitive processes (and the limits of human cognition)[10] has inspired hypotheses about mechanisms that could account for several cognitive biases.[11] These include cognitive uncertainty[12] or an understanding of how people estimate probabilities through the selective recall of memories.[13] But even theories that held together different biases that have received strong empirical support[14] sometimes fail to be replicated.[15]

**Figure S4.2.1 The identification of 180 cognitive biases makes it hard to derive insights about how to change behaviour to enhance collective action**




**Source:** "The Cognitive Bias Codex - 180+ biases," designed by John Manoogian III based on categories and descriptions by Buster Benson, licensed with CC-by-SA-4.0 (https://commons.wikimedia.org/wiki/File:Cognitive_bias_codex_en.svg).


Too much information
We notice things already primed in memory or repeated often
Bizarre, funny, visually striking or anthropomorphic things stick out more than nonbizarre or unfunny things
We notice when something has changed
We are drawn to details that confirm our own existing beliefs
We notice flaws in others more easily than we notice flaws in ourselves
We tend to find stories and patterns even when looking at sparse data
We fill in characteristics from stereotypes, generalities and prior histories
We imagine things and people we are familiar with or fond of as better
We simplify probabilities and numbers to make them easier to think about
We think we know what other people are thinking
Not enough meaning
Availability heuristic
Attentional bias
Illusory truth effect
Mere-exposure effect
Context effect
Cue-dependent forgetting
Mood-congruent memory bias
Frequency illusion
Baader–Meinhof phenomenon
Empathy gap
Omission bias
Base rate fallacy
Bizarreness effect
Humor effect
Von Restorff effect
Picture superiority effect
Self-relevance effect
Negativity bias
Anchoring
Conservatism
Contrast effect
Distinction bias
Focusing effect
Framing effect
Money illusion
Weber–Fechner law
Confirmation bias
Congruence bias
Postpurchase rationalization
Choice-supportive bias
Selective perception
Observer-expectancy effect
Experimenter's bias
Observer effect
Expectation bias
Ostrich effect
Subjective validation
Continued influence effect
Semmelweis reflex
Bias blind spot
Naïve cynicism
Naïve realism
Confabulation
Clustering illusion
Insensitivity to sample size
Neglect of probability
Anecdotal fallacy
Illusion of validity
Masked-man fallacy
Recency illusion
Gambler's fallacy
Hot-hand fallacy
Illusory correlation
Pareidolia
Anthropomorphism
Group attribution error
Ultimate attribution error
Stereotyping
Essentialism
Functional fixedness
Moral credential effect
Just-world hypothesis
Argument from fallacy
Authority bias
Automation bias
Bandwagon effect
Placebo effect
Outgroup homogeneity bias
Cross-race effect
Ingroup favoritism
Halo effect
Cheerleader effect
Positivity effect
Not invented here
Reactive devaluation
Well-traveled road effect
Mental accounting
Appeal to probability fallacy
Normalcy bias
Murphy's law
Zero sum bias
Survivorship bias
Subadditivity effect
Denominator effect
The magical number 7 ± 2
Illusion of transparency
Curse of knowledge
Spotlight effect
Extrinsic incentive error
Illusion of external agency
Illusion of asymmetric insight
Telescoping effect
Rosy retrospection
Hindsight bias

## NOTES

1. Ijzerman and others 2020.
2. An early expression of concern was Simmons, Nelson and Simonsohn (2011). For reviews, see Nelson, Simmons and Simonsohn (2018) and Nosek and others (2022).
3. Youyou, Yang and Uzzi 2023.
4. Camerer and others 2016;' Camerer and others 2018; Yarkoni 2022. A recent review of multiple studies recommending interventions to increase happiness reveals very little support for several widely recommended policies (Folk and Dunn 2023).
5. Etheredge 2021.
6. Websites include http://datacolada.org/ and http://bps.stanford.edu/. The challenge has been widely reported in the media (see, for instance, Schelber 2023).
7. Hallsworth (2023) proposes a manifesto on how behavioural science needs to evolve to strengthen its empirical foundations and policy relevance. Duckworth and Milkman (2022) propose improvements in the conduct of meta-studies to enhance the validity of findings. Clark, Connor and Isch (2023) show that studies that fail to replicate are associated with declines in citations, thus the proposal by Zwaan and others (2018) that replication should become mainstream could enhance the validity and robustness of results. Box-Steffensmeier and others (2022) argue for the importance of cross-disciplinary learning. van Roekel and others (2023) propose improvements in the design of nudges so that they preserve autonomy, given that a strand of criticism of nudges is that they are paternalistic and curb people's ability to reason when making choices (these criticisms were reviewed in UNDP 2022a). Korbmacher and others (2023) document a series of positive structural, procedural and community changes in which the replicability crisis is turning to a credibility revolution.
8. Davis 2023, p. 476.
9. For instance, Stango and Zinman (2022) reduce 20 biases to 4 behavioural common factors. Goeree and Louis (2021) developed a model to integrate several findings from behavioural game theory.
10. Lieder and Griffiths 2020.
11. As explored, for instance, in behavioural game theory (Camerer, Ho and Chong 2015). Dube, MacArthur and Shah (2023) show how cognitive demands on policing can undermine officer decisionmaking. Enke (2020b) shows how people confronting complex decisions focus on what they see. Bordalo, Gennaioli and Shleifer (2022) draw the implication of the outsized influence of salient information on decisionmaking.
12. Enke and Graeber 2023.
13. Bordalo and others 2022.
14. Dean and Ortoleva 2019.
15. Chapman and others 2023.

SPOTLIGHT 4.3

# Cultural evolution and development policy

**Joseph Henrich,** *Harvard University, Departments of Human Evolutionary Biology and Psychology, Harvard's Kennedy School*

All approaches to policy design and economic development require assumptions about human nature, though these are often implicit, typically smuggled in without notice.[1] By attending closely to human evolutionary biology, the new interdisciplinary field of Cultural Evolution (CE) offers fresh insights into human behaviour, cultural differences, psychological changes, institutional effectiveness, technological innovation and economic outcomes.[2] Because of its historical and comparative approach, CE has explored a broad range of social phenomena, including religions,[3] witchcraft beliefs,[4] kinship systems,[5] collective rituals[6] and gender inequalities,[7] and considered their links to various economic, political, demographic, social and health outcomes.

Drawing on CE research, I shine a spotlight on the nature of human cooperation, the coevolution of institutions and cultural psychologies and the impact of shocks on people's psychology. Like economics, CE is built on a large body of formal mathematical models that act as mental prostheses for thinking about the learning and decisionmaking processes that underpin behaviour and how these give rise to sociological phenomena such as social norms, institutions, large-scale cooperation and ethnic groups.[8] However, unlike economics, CE is founded on evolutionarily plausible and empirically grounded assumptions about how humans actually learn and adapt rather than on notions of rational choice rooted in free-floating philosophical assertions.

## New evolutionary foundations

Taking an evolutionary perspective, CE theorists begin by asking a set of deep questions about our species. What kind of animal are we? What is the secret of our species' success? How are we different from other animals?

Decades of research point to a set of answers, but they are not the ones many people assume. Much of our nature is nurture, but nurture from selected members of our communities and peers as well as our families. We are a cultural species that has evolved genetically to rapidly, efficiently and often unconsciously acquire beliefs, ideas, heuristics, perceptions, motivations and much more from those around us.[9] Our life histories—gestation, infancy, childhood and so on—have evolved to permit us to adaptively calibrate aspects of our psychology, including our attention, preferences and perceptions, to the worlds we confront. Indeed, a growing body of research shows how, beginning in our first year of life, humans seem exquisitely well attuned to selectively attending to and learning from the people most likely to possess useful or adaptive information, often relying on cues of competence, skill, success and prestige to target our learning efforts.[10] We also assiduously attend to certain domains, such as those related to food, sex, reputation, animals, plants and social groups, and process these different domains in distinct ways.[11]

Over generations these selective learning processes and content filters generate, often without anyone realizing it, increasingly adaptive cultural packages of tools, know-how, beliefs, motivations and more. We have depended on the useful products of such cultural processes for so long that we have genetically evolved to rely on what we acquire from other people—culture—over our own experience or instincts. Many cultural products and practices, including our institutions, may appear "rational" but instead actually emerged through cultural evolution, often without anyone evaluating the costs and benefits of alternative options or even understanding how and why particular practices, institutions or heuristics emerged. Of course, our evolved learning abilities—like our instinctual tastes for fat, salt and sugar—can produce extravagant maladaptations, which include

deeply held commitments and beliefs that deviate wildly from reality. But that is the cost of being a cultural species.

In applying this approach to understanding human cooperation, researchers have focused on how cultural learning, operating within groups and over time, gives rise to social norms. Social norms are widely shared behavioural patterns typically sustained by reputational effects, punishment, signalling or other mechanisms that can incentivize individually costly behaviours or practices. Norms emerge spontaneously once people can learn both focal behaviours (such as sharing food) and the standards for judging others (for example, nonsharers are "bad"). Around the world both ethnography and experiments suggest that the social norms spread by cultural evolution may explain many widespread patterns of cooperative behaviour, from food sharing among hunter-gatherers to voluntary blood donations in modern urban centres.[12] Because humans have had to navigate social landscapes shaped by social norms for eons, we have genetically evolved a norm psychology that primes us to readily learn social rules, internalize these rules as behavioural heuristics or motivational preferences and react negatively to norm violators. Norm internalization may be a key aspect of what makes us behave fairly and altruistically in normative contexts.[13]

Behavioural economists stumbled upon these internalized normative motivations when they began conducting economic experiments such as the prisoner's dilemma or ultimatum game. And, of course, anthropologists established decades ago that game-related behaviours, driven by particular preferences or heuristics, are culturally transmitted[14] and vary predictably across human societies in ways patterned by cultural evolution.[15]

Rather than assuming institutions as if they descended from on high or were hammered out by rational actors in some Lockean social contract,[16] CE offers a natural way to theorize and understand the origins of institutions from the ground up. It proposes that informal institutions represent interlocking sets of social norms. Marriage institutions, for example, are formed by norms that specify such things as who pays whom to form the union (such as brideprice or dowry), where the couple lives after marriage (for example, with the groom's family) and how many spouses one can have at the same time (polygyny versus monogamy).[17] Formal institutions emerge when some of the norms or rules in a more comprehensive package are written down. This is part of the reason that formal institutions cannot be readily replicated by simply agreeing to follow the written elements of the institutions—many of the key constituents of any real institution are not written down.

## The oldest institution

Crucially, there is much more to human nature than simply our cultural learning abilities and our norm psychology. To see this, consider the oldest and most fundamental of human institutions—the family, or what anthropologists call kinship systems. These packages of social norms variously harness, extend or suppress aspects of our innate kin psychology.[18] Like other species, our kin psychology includes instincts for helping close relatives, avoiding inbreeding (such as sex with siblings) and sustaining pair-bonds. Cultural evolution exploits these aspects of our evolved psychology to build various social organizations or networks, including clans, kindreds, extended families and lineages, using norms that specify acceptable marriage partners (incest taboos), inheritance rules (of resources, leadership positions and identity), communal ownership, postmarital residence and shared culpability for crimes (termed corporate responsibility). By variously strengthening, weakening or modifying various kin bonds, cultural evolution can forge either corporate collectives capable of high levels of cooperation or sprawling social networks that offer people refuge when disasters strike.[19]

Historically, after the origins of food production more than 10,000 years ago, competition among groups with varying social norms drove changes in kin-based institutions that fostered intensive, tightly knit cooperative networks and larger scale cooperation. The shifts to polygynous clans and lineages during this period, particularly those based on patrilineal lines of descent, were so profound that they can be seen in the genome in the massive reduction in Y chromosome diversity after the emergence of agriculture but before the rise of states.[20]

To illustrate the power of kinship, consider a specific custom, the social norms specifying that adults can

have only one spouse at a time—normative monogamy. Most societies throughout human history—85 percent by some estimates—have permitted high-status men to take multiple wives.[21] Even in otherwise highly egalitarian hunter-gatherer societies, the most prestigious hunters, warriors, storytellers and shamans often took three to five wives. To explore why monogamous marriage emerged and spread in societies where massive wealth differences among men persist, cultural evolutionists have pointed out that polygynous marriage generates societal-level costs: it tends to create a large pool of low-status men who have little opportunity or access to the marriage and mating market. Faced with ending up as evolutionary zeroes, unless they can catapult themselves up a steep status hierarchy, men become more likely to take risks that result in crime, raiding, violence and rape.

Monogamous marriage, by contrast, creates a more equitable distribution of wives and children across the male status hierarchy, effectively draining the pool of low-status bachelors and, instead, enlisting them in an army of husbands and fathers, giving them a stake in the future. Interestingly, while in monogamous societies both marriage and fatherhood are associated with declines in men's testosterone levels, the same is not true of men in polygynous societies. Indeed, several lines of evidence suggest that, at least under some conditions, reducing polygynous marriage influences crime, domestic violence and gender inequality. The adoption of monogamous marriage is a fascinating case because it runs directly contrary to the interests of elite and powerful men, who usually have a disproportionate influence on laws and policy.[22]

Across traditional kinship practices, including norms related to polygyny, cousin marriage, inheritance and residence, ample evidence demonstrates the impact of kin-based institutions on important outcomes, including economic prosperity, trust, civic participation, innovation, corruption, child health, gender inequality, education investments and the effectiveness of democratic institutions. Duman Bahrami-Rad and colleagues, for example, show that measures of traditional kinship intensity predict global measures of economic prosperity based on nighttime satellite luminosity.[23] Indeed, focusing only within countries, their analyses show that crossing from an ethnic group with high kinship intensity (polygynous clans) into an ethnic group with low kinship intensity (monogamous nuclear families) corresponds to a substantial rise in luminosity/prosperity.

Of course, while kin-based institutions are notoriously resilient, policies can and have altered key social norms and changed how these institutions operate.[24] For example, using historical data for the United States, Ghosh, Hwang and Squires (2023) show how state laws prohibiting cousin marriage resulted in faster urbanization and more rapid income growth.[25] Similarly, illustrating potential pitfalls, a study of India shows how legal changes in 2005 that gave women equal inheritance rights caused a rise in arranged marriages to patrilineal cousins, which in turn resulted in a decline in both gender equality and women entering the labour market. In both cases the social and economic effects were inadvertent, though probably desirable to policymakers in the former case but undesirable in the latter.[26]

The study of kin-based institutions illustrates two important features of cultural evolution. First, understanding these institutions offers a clear example of why it is crucial to theorize about human nature—without such a framework it is difficult to fathom why people care so much about close relatives, why testosterone responds to the local mating environment (monogamy or polygyny) and why people internalize social norms (where do fairness preferences come from?). Concepts such as norms and institutions are not assumed into existence but instead are understood as arising through clearly defined evolutionary processes.

Second, cultural evolution shows how institutions can emerge without conscious social contracts or rational choice but still operate in functional ways, serving the interests of society or particular subgroups.[27] Indeed, like the proverbial fish that does not know it lives in water, most people do not understand how our institutions work. Normative monogamy offers an example of an institution that, operating over generations, dramatically influences societal social dynamics and important outcomes. Yet most people, including policymakers and legal scholars, do not recognize why or how it works or even realize that it "does" anything.[28] Here, cultural evolution offers a foundational understanding of kin-based institutions that

highlights an array of potential policy levers as well as potential pitfalls that typically go unrecognized.

## Markets, religion and intergroup competition

To understand the evolution of larger scale cooperation above the kin group, cultural evolution offers a multilevel perspective—supported by an armoury of formal models[29]—that analyses the impact of intergroup competition and conflict. The approach reveals how intense cooperation among smaller groups within societies, such as families, villages and ethnic groups, can undermine cooperation at higher levels such as in kingdoms, states and empires.[30] When smaller groups within societies command too much solidarity and loyalty, it gets harder to motivate people to pay taxes, fight wars, build canals and so on. This multilevel evolutionary perspective permits researchers to spot the fault lines where morality breaks down, cooperation plummets and conflict begins. This approach also underlines the challenges to achieving global-level cooperation.[31]

Beyond kin-based institutions, the social norms, beliefs and motivations that drive large-scale cooperation are influenced by many factors, including market institutions, religions and domesticated forms of intergroup competition. Focusing on markets, several lines of evidence indicate that greater market integration is associated with greater impersonal prosociality, including greater trust, fairness and cooperation with anonymous others. The idea, which traces back to the Enlightenment, proposes that by engaging with markets, people acquire and internalize norms that foster reciprocal and mutually beneficial transactions with strangers.[32] For example, using a global database of folktales, Enke shows that greater market integration is associated with greater moral universalism and trust in strangers, as captured by people's traditional stories.[33] Similarly, behavioural experiments in Ethiopia show that communities of Bale Oromo that are located closer to markets are more cooperative with anonymous others and consequently are better able to sustainably manage local forests.[34]

Cultural evolutionists have long argued that intergroup competition, operating over thousands of years, has shaped religions and rituals in ways that expand the sphere of cooperation and exchange, fostering the scaling up of human societies. Empirically, cultural evolution has explored the impact of different religions on family organization (kinship intensity), aspects of moral psychology, cooperation among strangers and economic outcomes.[35] For example, using both economic experiments and surveys, several studies show how stronger beliefs in powerful moralizing gods or universal karmic forces foster greater cooperation and fairness with anonymous others. This finding is particularly striking on realizing that a belief in such deities is not found in most human societies and emerged only during the last few thousand years. Similarly, global variation in people's commitment to world religions is correlated with key economic preferences, including generalized trust, altruism towards strangers and reciprocity with anonymous others.[36] Such psychological patterns converge with older research linking economic growth to religious beliefs about the afterlife.[37]

Finally, cultural evolution has also domesticated forms of intergroup competition within societies that galvanize higher trust and cooperation among strangers against the corrosive effects of self-interest, nepotism and cronyism. Cultural evolutionary theory suggests that competition among groups demands cooperation, resulting in the spread of motivations and practices that increase cooperation. Testing this idea, Francois and colleagues exploited a natural experiment in which changes in banking regulations increased competition among firms, mostly during the 1970s and 1980s.[38] They show that this policy change increased competition, which in turn drove trust gradually upward over many years. Supplementing this, the study used panel data for Germany to show that trust rose when individuals moved to a more competitive sector of the economy and declined when they moved to a less competitive sector. In the lab the team confirmed that increasing intergroup competition increased both people's willingness to cooperate with strangers and their inclination to state that "most people can be trusted" on the generalized trust question.

## Thinking, feeling and perceiving

Because CE proposes that human brains evolved genetically in worlds structured by changing institutions, languages and technologies, the field was

primed to recognize, study and eventually explain psychological differences across populations. Psychologists and economists typically assume that human minds are like digital computers—that the information-processing hardware is all fixed. However, it is increasingly clear that human brains evolved to ontogenetically adapt their information processing to the challenges that individuals face while growing up and, to a lesser degree, over the course of their lives. For example, recent work exploring the role of paddy rice agriculture, irrigation, ploughs, pastoralism, kin-based institutions and urbanization has sought to explain the variation around the world in moral psychology, conformity, holistic thinking, ingroup loyalty, normative tightness, nepotism, honour motivations, individualism, personality structure and impersonal prosociality (trust in strangers).[39] It is not just that different institutions create different incentives—as many economists have assumed—it is that people who grow up in different places come to process information differently. That is, they perceive, reason, feel and think differently.[40]

Such psychological variation implies that identical policies, laws and institutions will often have different outcomes due to underlying psychological differences. For example, in a field experiment conducted in Ghana, India and the Philippines, researchers randomly assigned workers to be paid using an individual piece rate, where they were paid according to how much they alone produced; a group piece rate, where they were paid according to the average productivity of their small working group; or a daily wage, where they were paid independent of their productivity.

Strikingly, the most profitable policy depended on the population. In the most individualistic country in this trio, India, both the individual and group piece rates generated roughly a 20 percent increase in average performance, which is about what would be found in the United States using an individual piece rate. In the Philippines the performance-enhancing effects of paying an individual piece rate were only about 10 percent (half that of India), but the effect of the group piece rate was not any better than simply paying a daily wage. In Ghana neither piece rate scheme generated any improvement in performance over the simple daily wage. The performance-enhancing policy depends on the cultural psychology that people bring into the labour market.

Indeed, using data from 11,702 firms around the world, analyses show that firms in more individualistic populations are more likely to rely on performance pay. Here, what might look like a failure to adopt the most effective management practices (that is, not using performance pay) might instead represent an appropriate calibration to the local cultural psychology. Such results, and numerous others, suggest that many insights from standard economic models are most applicable to societies with particular cultural psychologies.[41] CE offers an overarching framework for thinking about human behaviour, psychology and decisionmaking that seats individuals within their historical and cultural contexts, effectively organizing and explaining the potpourri of (mostly) culturally evolved heuristics and biases identified by behavioural scientists.

## Wars, hurricanes, earthquakes, epidemics and other shocks

Recognizing the central importance of shocks ranging from volcanic eruptions and plagues to wars and hurricanes, cultural evolutionists have examined how such events affect people's psychology and shape cultural change. A growing body of research demonstrates that shocks can powerfully affect people's sociality—bonding them more closely to their communities while also tightening their commitments to social norms. Using a variety of natural experiments, surveys, economic games, psychological measures (from text analysis) and naturalistic observations, researchers have shown that shocks strengthen cooperation within local groups, tighten social norms of all kinds, increase people's religious commitments and, perhaps oddly, shift them away from a universalistic morality. In Sierra Leone, for example, those most affected by the civil war, which had ended a decade prior, were more cooperative with their local ingroups and more religious but less inclined to cooperate with distant strangers.[42]

This is important because climate shocks shape morality and cooperation in ways that seem poorly suited to achieving the kind of global cooperation necessary to tackle problems such as climate change. To foster such large-scale cooperation, cultural evolution suggests there may be ways to scale up some of the processes that have galvanized cooperation over

the past 10 millennia. First, intergroup competition, whether among firms or countries, can be harnessed in more benign ways to increase cooperation.[43] Second, our evolved psychology of interdependence and ethnic psychologies can be tapped to create a pan-human sense of connection and a global identity that expands the moral sphere.[44] Third, given our powerful inclination to copy the most successful and determined prestigious nations, groups and individuals can foster greater cooperation by leading with costly prosocial acts that demonstrate the commitments they are seeking from others, not by waiting to see if others will cooperate.[45]

To conclude, equipped with a theoretically rich conception of human nature, the rapidly growing field of CE offers many new perspectives and approaches on how to think about and study cultural change, economic development and social policy.

## NOTES

1. Wilson 2019.
2. Boyd 2017; Henrich 2016, 2020; Laland 2017.
3. Gervais and others 2016; Watts and others 2015.
4. Carvalho and others 2023.
5. Enke 2019; Schulz and others 2019.
6. Xygalatas and others 2013.
7. Galor, Özak and Sarid 2018.
8. Boyd and Richerson 1985; Cavalli-Sforza and Feldman 1981.
9. Tomasello 1999.
10. Chudek and others 2013.
11. Henrich 2016.
12. Boyd and Richerson 2022.
13. Henrich and Muthukrishna 2021; House 2018; House and others 2020.
14. Baimel and others 2021; Cason and Mui 1998; Rand 2016; Salali, Juda and Henrich 2015.
15. Ensminger and Henrich 2014; Falk and others 2018; Henrich 2000; Henrich and others 2005; Lang and others 2019.
16. Acemoglu and Robinson 2012.
17. Henrich, Boyd and Richerson 2012.
18. McNamara and Henrich 2017.
19. Henrich 2020.
20. Zeng, Aw and Feldman 2018.
21. Henrich, Boyd and Richerson 2012.
22. Chaudhary and others 2015; Henrich 2020; Henrich, Boyd and Richerson 2012; Hudson and others 2023.
23. Akbari, Bahrami-Rad and Kimbrough 2019; Alesina and Giuliano 2013, 2015; Bahrami-Rad 2021; Bahrami-Rad and others 2022; Lowes 2022; Moscona, Nunn and Robinson 2017, 2020; Schulz 2022; Schulz and others 2019.
24. Bau 2021.
25. Ghosh, Hwang and Squires 2023.
26. Bahrami-Rad 2021.
27. Henrich 2020; Turchin 2009.
28. Henrich 2020; Henrich, Boyd and Richerson 2012; Hudson and others 2023.
29. Boyd and Richerson 1990, 2002, 2011; Turchin 2015.
30. Turchin 2013; Wilson and others 2023.
31. Desmet, Ortuño-Ortín and Wacziarg 2017; Handley and Mathew 2020; White, Muthukrishna and Norenzayan 2021; Wilson and others 2023.
32. Hirschman 1982.
33. Enke 2023a.
34. Baldassarri 2020; Enke 2023a; Henrich and others 2005; Henrich and others 2010; Kosfeld and Rustagi 2015; Rustagi, Engel and Kosfeld 2010; Voors and others 2012; White, Muthukrishna and Norenzayan 2021.
35. Schulz and others 2019.
36. Atkinson and Bourrat 2011; Caicedo, Dohmen and Pondorfer 2023; Gervais and others 2016; Lang and others 2019; Purzycki and others 2016; White and others 2019.
37. Barro and McCleary 2003.
38. Francois, Fujiwara and van Ypersele 2018.
39. Atari and others 2017; Buggle 2017; Enke 2019; Falk and others 2018; Gelfand and others 2011; Lukaszewski and others 2017; Schulz and others 2019; Talhelm 2020.
40. Apicella, Norenzayan and Henrich 2020; Henrich and others 2022.
41. Bandiera, Barankay and Rasul 2011; Medvedev and others 2024.
42. Bauer and others 2014; Bauer and others 2016; Gelfand and others 2011; Henrich and others 2019; Rao and others 2011; Sinding Bentzen 2019; Winkler 2021.
43. Francois, Fujiwara and van Ypersele 2018; Wilson and others 2023.
44. Handley and Mathew 2020; Moya 2016; White, Muthukrishna and Norenzayan 2021.
45. Chudek and others 2012; Henrich 2009; Henrich and Gil-White 2001; Henrich, Chudek and Boyd 2015; Kraft-Todd and others 2018.

SPOTLIGHT 4.4

# The role of trust and norms in tax compliance in Africa

**Odd-Helge Fjeldstad and Ingrid Hoem Sjursen,** *Chr. Michelsen Institute, Norway*

The tax system is a key formal institution with a unique role in the social contract between people and governments, as an essential source of revenue for governments to fund public services and programmes that benefit the community. It also provides an important entry point to explore how people engage with institutions across different contexts and the role of culture, beliefs, norms and perceptions in determining issues such as compliance with policies. This spotlight synthesizes findings from recent research on determinants of tax compliance and evasion, with a focus on developing countries.

Mobilizing domestic revenue is crucial for developing countries to achieve the Sustainable Development Goals. However, tax evasion is a major challenge in many countries. Research and policymakers have generally focused on law-based compliance and the role of formal rules and institutions such as audits and penalties to reduce tax evasion—often referred to as enforced compliance.[1] More recently, voluntary compliance[2]—informal norms and beliefs motivating taxpayers' compliance, particularly trust and norms—have received more attention.[3] Voluntary compliance is likely to be particularly important in countries where enforcement capacity is weak.[4] This spotlight starts with a short theoretical background on how a deeper understanding of trust and norms can enhance our understanding of voluntary compliance. It then examines how these factors vary across different contexts, taxpayers and tax bases and how these variations affect voluntary compliance. The last section discusses policy implications.

## Deeper knowledge of trust and norms can enhance our understanding of tax compliance

Trust (a person's belief that another person or institution will act consistently with their expectations of positive behaviour)[5] fosters social and economic progress.[6] Theoretical work emphasizes the importance of trust in the government and in the tax administration, as well as for voluntary tax compliance. Kirchler, Hoelzl and Wahl (2008) develop a theoretical framework in which trust in tax authorities and the power of authorities are the main determinants of tax compliance, where trust fosters voluntary compliance and power leads to enforced compliance. When taxpayers trust the tax administration and perceive it as benevolent and working beneficially for the common good, taxpayers may feel obliged to adhere to decisions, policies and rules, even in the absence of powerful administration and enforcement.[7] Prichard and others (2019) develop a conceptual framework for tax reform and compliance that highlights four key drivers of trust: fairness (the tax system is fairly designed and administered), equity (burdens are equitably distributed and everyone pays their share), reciprocity (tax revenue is used for public goods and services) and accountability (governments are accountable to taxpayers). While fairness and equity are features of the tax system, reciprocity and accountability relate to broader governance issues. The equity dimension entails that in addition to trust in the tax authority, trust in fellow citizens may be an important determinant of tax compliance.

Both personal and social norms have been argued to be important determinants of tax compliance (table S4.4.1).[8] Social norms may be important to tax compliance because people care about how they are perceived by others and the social sanctions and rewards associated with these perceptions[9] or because they want to behave as others do. Importantly, personal and social norms can be misaligned, and people may not always act according to their own personal norms.[10] Several studies have identified the phenomenon of pluralistic ignorance, a situation in which most group members personally reject a norm but believe that most others accept it.[11] When pluralistic ignorance exists, providing information about the views of others has been shown

**Table S4.4.1 Types of norms and examples**

| Personal norm or attitude (Moral norm) | Social norm ("a rule of behavior such that individuals prefer to conform to it on the condition that they believe that (a) most people in their reference network conform to it (empirical expectation), and (b) they ought to conform to it (normative expectation)"; Bicchieri 2016, p. 35) | |
|---|---|---|
| | Descriptive norm (Empirical expectation) | Injunctive norm (Normative expectation) |
| What I believe is the right thing to do | What I believe others do | What I believe most people think I should do |

**Source:** Bicchieri 2016; Cialdini, Kallgren and Reno 1991.

to change both tax behaviour[12] and behaviour in other areas.[13] Thus, to understand taxpayer behaviour, it is important to identify and analyse the personal and social norms associated with tax compliance and to investigate the various factors that influence personal and social norms. Differentiating between personal and social norms is a prerequisite for designing efficient policies to enhance desirable outcomes.[14] Empirically, a large literature of field and lab experiments shows that personal and social norms influence each other and that both motivate behaviour but that social norms affect behaviour more than personal norms do.[15]

### *Trust and norms can vary across different contexts and affect tax compliance*

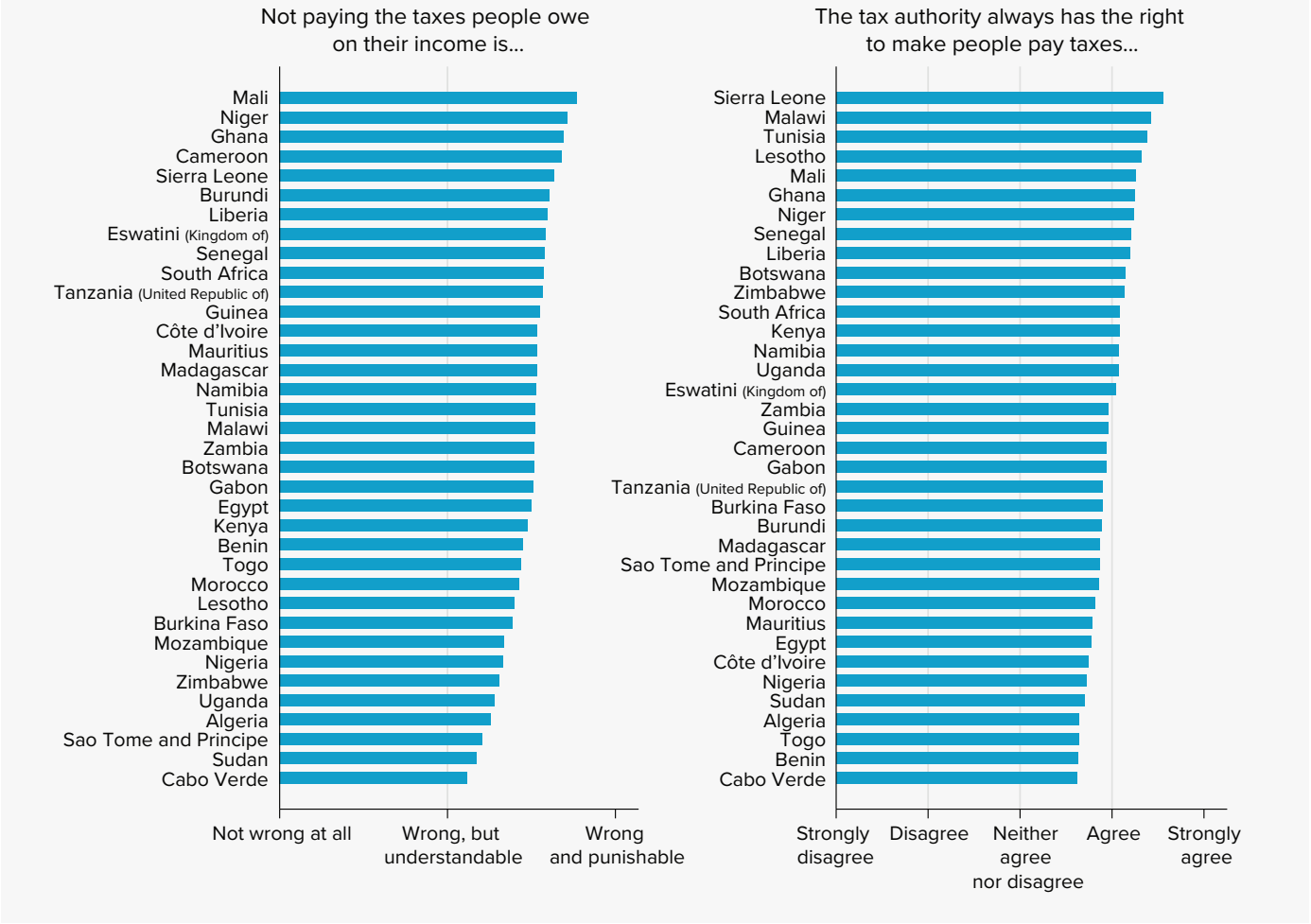
Tax compliance is challenging to measure because individuals are typically trying to hide noncompliant behaviour and attitudes.[16] Empirical investigations of determinants of voluntary compliance have commonly used survey questions from large databases, such as Afrobarometer and the World Values Survey, asking respondents about their views of whether not paying tax is wrong and punishable/justifiable or whether the tax authority has the right to make people pay taxes (figure S4.4.1).[17] In all countries the average respondent thinks that not paying taxes on income is at least "wrong, but understandable" and is closer to agreeing than disagreeing with the statement that the tax authority always has the right to make people pay taxes—but there is substantial variation across countries.

Studies based on such survey measures show that within countries voluntary compliance is positively correlated with a stronger feeling of national identity,[18] trust in the tax authority[19] and perceived fairness in how the government treats the respondent's own ethnic group,[20] which according to the framework of Prichard and others (2019) is an important driver of trust. Furthermore, there is a positive correlation between voluntary compliance and the perceived social norm for tax compliance, as well as satisfaction with provision of public services.[21] However, there are also substantial differences in correlates of voluntary compliance among Kenya, United Republic of Tanzania, Uganda and South Africa.[22] While these studies provide interesting insights into correlates of voluntary compliance, they do not offer causal evidence or explanations for the mechanisms through which the determinants affect voluntary compliance.

### *The weight of history in shaping trust and norms today*

To better understand the causal mechanisms behind variations in voluntary compliance, one strand of the literature studies the effect of historical roots and cultural heritage on voluntary tax compliance.[23] Cultural heritage is passed on from one generation to the next and coupled with the country or ethnic group of origin. And it is well documented that it can affect people's trust in others—for instance, trust in people from the same ethnic group or (dis)trust in people from other ethnic groups, as well as trust in public institutions.[24] For instance, evidence suggests that trust is an important causal mechanism in the negative relationship between economic development today in parts of Africa and the slave trade: individuals who belong to ethnic groups that were more exposed to slave trade are less trusting in their relatives, neighbours, others of the same ethnicity and local government.[25] Moreover, the individual variation in trust in public institutions and neighbourhood caused by differential exposure to the slave trade also explains variations in voluntary tax compliance in several

**Figure S4.4.1 Most people in African countries think that not paying taxes on income is at least "wrong, but understandable" and are closer to agreeing than to disagreeing that the tax authority always has the right to make people pay taxes**




**Note:** The survey question for the left figure was "Please tell me whether the following is not wrong at all; wrong, but understandable; or wrong and punishable: Not paying the taxes they owe on their income," and the survey question for the right figure was "Please tell me whether you disagree or agree: The tax department always has the right to make people pay taxes."
**Source:** Based on the results of Afrobarometer Round 6, 2014/2015 (https://www.afrobarometer.org/, accessed 25 January 2024).

countries: more trusting individuals have a higher voluntary compliance.[26]

A study in Uganda finds that history also plays a role in that people in historically centralized parts of Uganda have mistrust towards the central government and public institutions but may be willing to follow rules and pay taxes when they live in a setting with higher interpersonal trust.[27] Trust affects voluntary tax compliance, and trust is affected by group heterogeneity shaped by history. Thus, historical events and organization of societies continue to shape present voluntary tax compliance through trust and social norms. This finding relates to results in the broader literature in institutional economics that history can matter for present-day outcomes through the evolution and persistence of early institutions.[28]

## How trust and norms inform challenges with tax compliance

### *Opportunities for tax evasion by self-employed individuals*

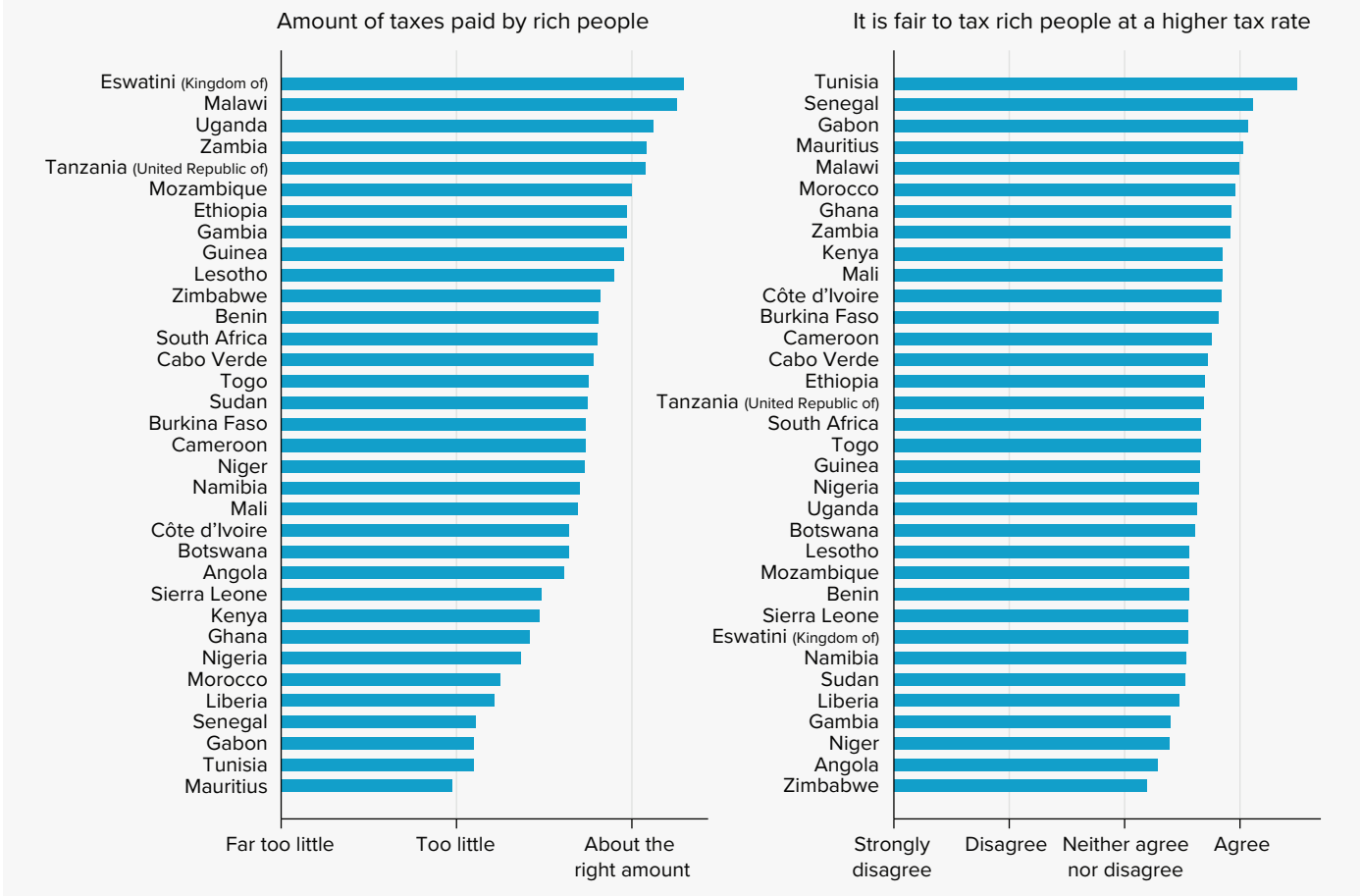
Self-employed professionals have more opportunities than salaried workers to minimize their reported

incomes—because more of their income is self-reported as opposed to reported by a third party[29]—and are more likely to take advantage of these opportunities.[30] Opportunities for tax evasion may affect people's voluntary tax compliance. Research shows that self-employed individuals have less favourable views on taxes and the tax authorities than other taxpayers.[31] Tax evasion is also found to be high among many self-employed individuals.[32]

### *Taxing the rich: Noble objectives, unrealistic expectations?*

Some studies argue that "the weakness of taxes on the wealthy not only affects revenue but also risks undermining broader trust in the tax system and weakening the social contract."[33] Thus, it is argued, "taxing the wealthy more effectively is critical not only to increasing revenue, but also to building trust in the tax system, thereby unlocking more sustained political support for taxation and the achievement of longer-term gains." However, redistribution through taxation is not a salient election issue in most African countries,[34] nor is it a strong priority of their citizens.[35] In most countries the average response to the Afrobarometer survey question on the amount of taxes that rich people are required to pay is closer to "about the right amount" than to "too little," and while the average respondent in all countries is closer to agreeing than disagreeing with the statement that rich people should be taxed at a higher rate to help poor people, the support for the statement is relatively weak in many countries (figure S4.4.2).

**Figure S4.4.2 Redistribution through taxation is not a salient election issue in most African countries, nor is it a strong priority of their citizens**




**Note:** The survey question for the figure on the left was "Do you think that the amount of taxes that rich people in [COUNTRY] are required to pay is too little, too much, or about the right amount?" and the survey question for the figure on the right was "Do you agree or disagree with the following statement: It is fair to tax rich people at a higher rate than ordinary people in order to help pay for government programmes to benefit the poor."
**Source:** Based on the results of Afrobarometer Round 8, 2019/2021 (https://www.afrobarometer.org/, accessed 25 January 2024).

Redistributive coalition building in ethnically diverse societies may be especially difficult,[36] so that any push for a wider redistributive agenda to benefit the poor tends to be weak.[37] While we sympathize with the argument that "the time has come to tax the rich," the focus of many African governments is to increase revenue by broadening the tax base to incorporate larger segments of individuals and firms in the tax net. The wealthy elites will probably be affected little by these reforms. This is reflected in what Mick Moore refers to as tax administrations' obsession to register new tax taxpayers, the majority of which are small-scale businesses and poor individuals.[38] This approach is associated with the idea that the major source of uncollected revenue in Sub-Saharan Africa is the informal sector.[39] A policy of taxing the very rich is not easy to implement.[40]

### *Corporate taxpayers: Trust and a predictable tax system*

Medium and large firms account for most of the tax revenue in many low- and lower-middle-income countries. Their voluntary compliance is likely to be influenced by different factors than individuals and small firms and needs to be conceptualized differently.[41] Voluntary compliance by firms is likely to be driven by self-interest to a larger extent than voluntary compliance by individuals.[42] Predictability is a critical concern of corporate taxpayers and enhances trust in a way that can allow firms to properly budget and make realistic plans for the future.[43] It also ensures that firms will be treated like their competitors. Questions about fairness and equity are often important for corporations because they affect market competition, profitability and the predictability of their operations.[44] For instance, are other firms in the same sector bearing equivalent tax burdens? Firms also are more likely to be compliant when they believe the government is funding services and activities that benefit them and when they have a voice in shaping those decisions.[45] Thus, improving the predictability and fairness of tax enforcement can foster voluntary compliance and support for reform for corporations.[46]

### *Taxing the informal sector*

A large share of economic activity in poor countries takes place in the informal sector, which is hard to tax.[47] Until recently, tax administrations tended to give it little priority because returns to effort may be low in cash terms, and collection is likely to be difficult. From the economic and administrative perspectives, it makes sense not to tax multitudes of poor people. The value-added tax system generally exempts basic goods that are consumed heavily by poor people, and the income tax code generally excludes individuals and entities with incomes below a certain threshold. However, in recent years several national revenue agencies have introduced special presumptive taxes directed at the informal economy that are based on workers' presumed rather than actual income, given the type of work they perform.[48]

A wider tax net is not always a good thing, but the possibility that tax reforms are driven by a calculus that emphasizes the advantages of excluding marginal payers must be a cause of concern.[49] This would be less of a problem if the actual tax burdens in poor countries were fairly and effectively distributed, but they are not. In particular, they often fall heavily on a small number of registered, formal companies.

Evidence suggests that the relationship between firm size and evasion is negative or U-shaped, implying that small firms are more likely to evade taxation.[50] This evasion may lead to unfair competition, which can undermine trust and negatively affect the voluntary tax compliance of medium firms.[51] Thus, one argument for improving taxation of small and medium enterprises is that it is important for ensuring equity and improving voluntary compliance. It thus makes sense to question the arguments for excluding smaller taxpayers from the tax net on pure efficiency grounds and to explore the potential political and revenue advantages of widening that net, while also carefully considering the administrative implications of doing so.

## Policy levers to address tax evasion: Beyond formal laws and regulations

Findings from the research reviewed above show that history, ethnic diversity and how tax revenue is spent may substantially affect people's voluntary tax compliance and trust in government and other citizens. Voluntary tax compliance is also likely to differ between segments of taxpayers (for example, between

individuals and businesses), between different taxes (for example, between direct and indirect taxes) and in how taxes are enforced. A general conclusion from this literature is that policies aiming to improve attitudes towards taxes in Africa should pay attention to strengthening the general environment of trust.[52] This is linked to a political economy approach that takes the historical, cultural and political contexts seriously, combined with conventional economic thinking.[53] Thus, it is important to move away from a purely technocratic approach when addressing tax evasion. Advice on tax policy, including methods of auditing and better tax design are valuable but must be located in a wider and case-by-case context, especially given the characteristics of many African countries.

A first step to addressing deep-rooted tax evasion norms is understanding how things actually function in the specific context, independently of how we would expect the tax system to perform according to good governance. This calls for more robust analysis of country and local contexts and institutions, particularly trust in tax authorities and social norms for tax compliance. Improving voluntary tax compliance furthermore requires thoroughly analysing different segments of taxpayers and revenue administrations, as well as their environment, to understand key players' norms and incentives.

This analysis leads to a two-pronged approach to reform. The first prong relates to developing policy instruments that are directed at both the incentives and opportunities for evasion. Unless taxpayers recognize that the penalties for being caught are much more severe than the potential gains, they will continue to take risk evading taxes. This, of course, requires enforcing the rules, which depends on the willingness at the top to reduce tax evasion. The second prong must go beyond legal and regulatory reform to address the root causes of tax evasion. Many efforts to adopt stricter rules for tax administration have failed because informal practices have continued. Changing social norms and mindsets is much more difficult than bringing in new regulations in part because social norms are deep rooted. Successful reforms are not achieved overnight. Reformers must keep this in mind and not be discouraged when they face challenges in implementing their reforms.

Social norms can be persistent across generations, economic development and political regimes.[54] But when they change, it can happen quickly—for instance, when new public information becomes available.[55] Behavioural tipping points—that is, when enough people have strong attitudes against an existing social norm (or towards a new one)—are decisive for norm change. In situations where the social norms for tax compliance are misperceived (underestimated), providing factual information about others' views may enhance compliance.[56]

Education can play a role when designed to help taxpayers understand the importance of paying taxes and how to do so. A wide range of taxpayer outreach and education activities exist across countries.[57] For instance, the Tanzania Revenue Authority is working with secondary schools to mainstream tax education into the curriculum. Government taxpayer education and outreach programmes generally often appeal to state-building narratives. Such programmes are valuable, but they must move beyond the frequent emphasis on why people should pay taxes towards emphasizing who pays taxes, how to pay them and what taxpayers receive in return.[58]

An essential component of building trust is the government's ability to demonstrate that tax revenue results in public services and broader benefits for taxpayers.[59] When governments can demonstrate those connections, it is possible to build meaningful popular support for more effective taxation and compliance.[60] This, combined with more transparent and predictable tax systems, is likely to result in more positive attitudes towards taxation in Africa and popular support for more effective taxation.

Just as improved service delivery is likely to be critical to encouraging voluntary compliance, so too is there an opportunity for more sustained investment in building trust with taxpayers.[61] A starting point for such trust building lies in improving the basic fairness of tax systems. Although discussions of building voluntary tax compliance often centre on improving the provision of public services, improvements in fairness may be important.[62] Such improvements are also much more directly under the control of tax administrations, which may be pursuing reform and seeking to build voluntary or quasi-voluntary compliance. Perceived corruption in tax authorities remains a major barrier to improving trust and voluntary compliance.[63]

## NOTES

1. Early research includes Allingham and Sandmo (1972); for a recent review, see Slemrod (2019).
2. A related term for voluntary compliance is "tax morale." We prefer "voluntary compliance" because it better captures aspects of taxpayer motivation that we consider important.
3. Besley 2020; Besley, Jensen and Persson 2023; Luttmer and Singhal 2014; Prichard and others 2019.
4. Kirchler, Hoelzl and Wahl 2008.
5. OECD 2017.
6. Algan and Cahuc 2014.
7. Kirchler, Hoelzl and Wahl 2008.
8. Kirchler, Hoelzl and Wahl 2008.
9. Besley, Jensen and Persson 2023; Luttmer and Singhal 2014.
10. Köbis, Jackson and Carter 2020; Onu 2016.
11. Miller and McFarland 1987.
12. Wenzel 2005.
13. Bursztyn, González and Yanagizawa-Drott 2020.
14. Bursztyn and Jensen 2017; Hallsworth and others 2017.
15. Bicchieri 2016. In the literature on tax compliance, see, for instance, Antinyan and Asatryan (2020), Bott and others (2020), d'Adda and others (2020), Dwenger and others (2016), Hallsworth and others (2017) and Slemrod (2019).
16. Ali, Fjeldstad and Sjursen 2014.
17. See Prichard (2022) for an overview of survey questions used in cross-country studies on voluntary compliance and a discussion of the weaknesses and limitations of these measures. Afrobarometer Round 6 was chosen because it is the most recent survey that includes both of the questions reported in the figure.
18. Besley and Mueller 2021; Blimpo and others 2018.
19. Besley and Mueller 2021.
20. Ali, Fjeldstad and Sjursen 2014; Sacks 2012.
21. Ali, Fjeldstad and Sjursen 2014; Blimpo and others 2018.
22. Ali, Fjeldstad and Sjursen 2014. See Fjeldstad, Schulz-Herzenberg and Hoem Sjursen (2012) for a broader review of correlates of voluntary tax compliance.
23. See Nunn (2020) for a review of the research on the historical roots of economic development more broadly.
24. Dinesen 2011; Kouamé 2021; Uslaner 2008; Woolcock and Narayan 2000.
25. Nunn 2008; Nunn and Wantchekon 2011.
26. Kouamé 2021.
27. Ali and Fjeldstad 2023. A state's organized power to uphold authority implies that it can uniformly apply policies throughout a given territory, such as extracting labour, enforcing the law and demanding taxes (Schraeder 2000).
28. Nunn 2009.
29. Dom and others 2022; Kleven and others 2011.
30. Engström and Holmlund 2009; Saez 2010.
31. Kogler and Kirchler 2020.
32. Chetty, Friedman and Saez 2013.
33. Dom and others 2022, p. 60.
34. Bleck and Van de Walle 2019.
35. Except, perhaps, in Ghana (Bleck and Van de Walle 2019). Inequality per se is not mentioned as a priority of citizens in a large recent Afrobarometer survey of 34 African countries (Coulibaly, Silwé and Logan 2018). Poor access to public services is a major concern, however—one that clearly has equity implications.
36. Mazrui 2008.
37. Bolch, Ceriani and López-Calva 2022.
38. Moore 2023.
39. Moore 2023. For instance, in a survey of 26 national tax administrations for the 2018 African Tax Outlook, 15 reported one or more special programmes or initiatives to deal with the informal sector (ATAF 2018). By contrast, only 4 had special sections for high-net-worth individuals.
40. Harrington 2016.
41. Alm and McClellan 2012; OECD 2019; Prichard and others 2019; Slemrod 2019.
42. Prichard and others 2019.
43. Campos, Lien and Pradhan 1999; World Bank 2018.
44. Alm and McClellan 2012; OECD 2019; Prichard and others 2019.
45. Prichard 2015.
46. Dom and others 2022.
47. Bird and Wallace, 2003. The concept the informal sector is disputed. When the term is used in relation to taxes in Africa, it generally refers to unregistered, small-scale economic operators (Moore 2023). The size of the informal economy is difficult to estimate. Estimates for low- and middle-income countries suggest that informality accounts for 30–70 percent of GDP and 20–80 percent of the labour force (Ulyssea 2020; Ulyssea, Bobba and Gadenne 2023). Estimates from West Africa suggest that more than 80 percent of total employment is informal and up to 60 percent of GDP is produced by informal activities (Benjamin, Mbaye and Diop 2012).
48. Dube and Casale 2016; Joshi, Prichard and Heady 2014; van den Boogaard, Prichard and Jibao 2018.
49. Ali, Fjeldstad and Sjursen 2014; Fjeldstad and Moore 2008.
50. Abdixhiku and others 2017; Cowell 2003; Hanlon, Mills and Slemrod 2007; Slemrod 2004.
51. Torgler and Schneider 2007.
52. Kouamé 2021.
53. Besley and Mueller 2021.
54. Jackson and Köbis 2018.
55. UNDP 2019.
56. Wenzel 2005.
57. Dom and others 2022.
58. Dom and others 2022.
59. Ali, Fjeldstad and Sjursen 2014; Bird 2011.
60. Dom and others 2022; Sanogo 2019.
61. Dom and others 2022.
62. Kogler, Muehlbacher and Kirchler 2015.
63. Aiko and Logan 2014; Fjeldstad 2006.