Activating cultural evolution for good when people differ from each other

Charles Efferson[1,*], Sonja Vogt[1,*], and Lukas von Flüe[1,*]

[1]Faculty of Business and Economics, University of Lausanne

[*]Address correspondence to *charles.efferson@unil.ch*, *sonja.vogt@unil.ch*, and *lukas.vonfluee@unil.ch*.

Keywords: applied cultural evolution, conformity, coordination, social tipping, behaviour change

Abstract

Interventions designed to promote behaviour change operate through at least two channels. First, an intervention can have a direct effect in that some people directly exposed to the intervention make a choice as a result of exposure. Second, an intervention can have an indirect effect in that some people make choices because they observe other people whose choices have already been influenced by the intervention. We surveyed the literature and found that this second cultural evolutionary channel has been offered as a way to amplify the effects of interventions across a spectacular array of domains related to health, human rights, and the environment. We argue, however, that ordinary forms of heterogeneity and the intrinsic complexity of cultural evolutionary processes introduce a number of ethical and practical challenges. For example, when people have different preferences, social planners should typically expect a trade-off between the direct and indirect effects of an intervention. Increasing one means decreasing the other, and social planners may often lack the information they need to resolve the trade-off effectively. In spite of these challenges, however, we also argue that narratives designed to be both educational and entertaining could play a powerful supporting role given our limited understanding of how to active cultural evolution for good.

1 Introduction

Imagine a well-intentioned social planner who wants to initiate behaviour change for good. We discuss below what "for good" may or may not mean, but for now assume this is a solved problem. Our representative social planner could be a policy maker, a business leader, or more broadly a person or organisation with influence or authority. The social planner has various tools she can use to induce people to change their behaviour. At the unobtrusive end of the spectrum, she can use subtle nudges that exploit psychological biases (Thaler & Sunstein 2021). Organ donations, for example, are dramatically higher when people have to opt out of donating than when people have to opt in (Johnson & Goldstein 2003, Davidai et al. 2012).

For some challenges, however, with climate change a clear example, nudges will at most provide useful but small contributions to an overall solution (Liebe et al. 2021, Thaler & Sunstein 2021). In these cases, the social planner might also turn to obtrusive approaches. For example, she can simply mandate a specific behaviour if she has the authority. Depending on how people perceive such mandates, and depending on how acute the enforcement problem is, outcomes can be quite counterproductive. Some people might even deliberately thwart the mandate (Thomas 2000, Camilotti 2016, Karlsen et al. 2019, Arnot et al. 2020, Moya et al. 2020). All in all, interventions from nudges to mandates can have heterogeneous effects among those directly exposed to the intervention. These heterogeneous direct effects depend in potentially complex ways on the nature of the intervention, the decision-making domain, heterogeneous characteristics of the individuals exposed to the intervention, and characteristics of the society in which those individuals live (Holtsmark & Skonhoft 2014, Pritchett & Sandefur 2014, Vivalt 2015, Vogt et al. 2016).

An intervention can also have indirect effects because of social interactions within the population. Because people learn from and influence each other, an individual who changes behaviour can lead others to do the same. More broadly, when individuals learn from and influence each other, endogenous cultural evolution typically follows at the population level (Boyd & Richerson 1985). If the social planner understands this process, she can implement an intervention that biases subsequent cultural evolution in favour of the change she would like to see (Wilson et al. 2014). To caricature the idea slightly, the social planner aims to activate a cultural evolutionary process, step back, and watch the process run its course.

We examine this idea with a focus on heterogeneity. We expect people to respond to interventions in heterogeneous ways, and this kind of heterogeneity will shape the direct effects of an intervention. Various experimental literatures on social influence and cultural evolution also show that people vary in terms of how they respond to information about the choices of others (Mesoudi et al. 2016, Kendal et al. 2018). This kind of heterogeneity should control the indirect effects of an intervention. Assuming the social planner wants to maximise

the total behaviour change, she will need to understand how these two ordinary forms of heterogeneity shape the interaction between the direct and indirect effects. Nonetheless, we argue below that the inherent complexity of the problem currently limits our understanding of how to do so. We also suggest that entertaining narratives could be an especially useful tool given our relative ignorance about the interplay between interventions and the cultural evolutionary dynamics they activate.

2 A spectacular array of domains

The notion that social influence and social learning can help achieve social goals has been widely influential. To get a preliminary grip on how widespread the idea is, we searched six databases (EconLit, Embase, PsycInfo, PubMed, Scopus, Web of Science). Full details are available online (`https://doi.org/10.17605/OSF.IO/6RDNF`). We searched for papers that include terms related to social learning (e.g. "social influence") and behaviour change (e.g. "behaviour") and interventions (e.g. "impact"). We excluded papers about non-human animals. As an additional step, we took relevant papers we were familiar with and conducted a search for other papers citing or cited by the original papers. This exercise produced a new set of papers that overlapped partially with the first set. We then searched all papers in the union of the two sets for those papers that explicitly discuss social influence and cultural change in policy-relevant domains. The end result was over 200 papers.

To organise these papers, we developed a table that reflects three important considerations. The full table is available online (`https://doi.org/10.17605/OSF.IO/6RDNF`), and an abbreviated version appears here as Table 1. First, we group papers based on the domain under study. Domains range widely from studies on the spread of accurate versus inaccurate information to studies about decision making related to health, sustainability, finance, or politics.

Second, many studies seem to fall on one side of a fuzzy divide between approaches that

emphasise social network structure versus those that emphasise social learning biases. Some papers focus on how social network structure influences the diffusion of information, opinions, and behaviours. These papers examine the complex effects of networks, but many of these papers also downplay or ignore the numerous ways in which agents can respond to social information distributed on a given network. Other papers, in contrast, may take a limited view of social networks, while at the same time emphasising the myriad psychological biases that can structure social learning and cultural transmission given a network. This fuzzy divide, to the extent that it separates distinct research traditions, is unfortunate. Existing theory indicates that network structure and social learning interact strongly; we cannot understand how one affects culture without the other (Barkoczi & Galesic 2016, Derex & Boyd 2016, Efferson, Vogt & Fehr 2020).

Third, we highlight papers that discuss the link between heterogeneity and social tipping. The most basic view of activating cultural evolution for good rests on the fact that one person's choices can affect other people's choices. If a social planner can get some people to change from a harmful behaviour to a beneficial alternative, the influence these people have may lead others to follow. Social tipping is a more precise idea. The claim is that social learning generates multiple locally stable equilibria at the population level and associated path-dependent dynamics. Some mix of a conformist psychology and coordination incentives, for example, is a common hypothesis about underlying mechanisms that could do this.

With this backdrop, the social planner favours a specific equilibrium that is not consistent with people's current values, norms, or behaviours. However, because the social planner assumes her preferred outcome is an equilibrium, she does not need an never-ending despotic intervention to generate behaviour change. Instead, her task is to dislodge the population from its current state and tip it into the basin of attraction for her preferred equilibrium, at which point endogenous cultural evolution takes over. This is the basic theory of change based on social tipping. The potential of the idea is consistent with both experimental studies on tipping (Centola et al. 2018, Andreoni et al. 2021) and observational studies that reveal the

following pattern. Cultural evolutionary processes can sustain distinct cultures at remarkably local spatial scales (Eugster et al. 2017, Lowes et al. 2017), which suggests multiple equilibria. Moreover, cultural traditions can change rapidly after long periods of stability, which suggests tipping from one equilibrium to another (Mackie 1996, Young 2015).

Aside from demonstrating the potential of social tipping, recent research has turned to unpacking the mechanisms that influence tipping (Nyborg et al. 2016, Efferson, Vogt & Fehr 2020, Andreoni et al. 2021). Individual heterogeneity figures prominently. In this chapter we dwell on why and how heterogeneity can be so important. We first define direct and indirect effects and then outline the logic of tipping in the simplest case with no heterogeneity of any kind. We then turn to the simplest form of heterogeneity. The social planner is different from people in the population, but everyone in the population is the same. As we will see, this introduces ethical subtleties, but it does not change the generic potential for tipping. Finally, we allow people to differ from each other, which expands the range of possible outcomes and complicates the social planner's task considerably.

Table 1: Overview of applied cultural evolution

Domains	**General vs. detailed notions of social influence**	**Tipping dynamics and individual heterogeneity**
Important fields of application include the cultural evolution of traditions that harm women and girls (Vogt et al. 2016, Efferson, Vogt & Fehr 2020, Schief et al. 2021), the diffusion of behaviours related to smoking (Christakis & Fowler 2008), contraceptive use (Alvergne & Stevens 2021), vaccination (Salali et al. 2022), the diffusion of pro-environmental norms and behaviour (Allcott 2011, Castilla-Rho et al. 2017), the spread of microfinance (Banerjee et al. 2013), the wisdom of crowds (Golub & Jackson 2010), and the diffusion of misinformation (Acerbi 2019, Sulik et al. 2021).	A fuzzy dichotomy seems to exist between either focusing on social networks while simplifying the countless ways in which psychological biases can structure the response to social information or focusing on the psychology of biased social learning while simplifying network formation. Specifically, studies that focus on the structure of social ties often posit simplified notions of social learning (Centola 2011, Christakis & Fowler 2013), while another strand of literature emphasises the importance of social learning biases (Henrich 2001, Smaldino et al. 2017, Efferson, McKay & Fehr 2020, Sulik et al. 2021). With some exceptions (Boyd & Richerson 1987, Efferson, Lalive & Fehr 2008, Derex & Boyd 2016), this latter strand of literature has highlighted the psychology of biased social learning given a network to the exclusion of other mechanisms (Singh et al. 2021), including the psychology of network formation. Recent work, however, suggests that we should expect strong interactions between network structure and social learning biases (Barkoczi & Galesic 2016), with downstream effects on optimal policies for social planners attempting to activate a cultural evolutionary process (Efferson, Vogt & Fehr 2020).	Social tipping can be crucial with respect to climate change (Nyborg et al. 2016, Farmer et al. 2019, Otto et al. 2020) or the cultural evolution of traditions that harm women and girls (Efferson et al. 2015, Cloward 2016, Novak 2020). Importantly, studies show that heterogeneous individual characteristics, social learning, and networks may often disrupt the potential for social tipping (Hayford 2005, Castilla-Rho et al. 2017, Efferson, Vogt & Fehr 2020). Tipping may or may not occur in ways that depend on the structure of the heterogeneity, the institutional setting, and the precise information available to decision makers (Andreoni et al. 2021). Models suggest that interventions targeting people amenable to change may often represent the worst strategy for activating endogenous cultural change (Efferson, Vogt & Fehr 2020). Ultimately, for social planners who want to influence economic and political decision making in a population, a joint approach to networks, social learning biases, and heterogeneity may prove necessary (cf. Derex & Boyd 2016, Efferson, Vogt & Fehr 2020).

3 Activating endogenous cultural change

In the simplest case, two behaviours are possible. One behaviour, U, is undesirable from the social planner's perspective. The other, D, is desirable. When the social planner enters the scene, most people are choosing U because of some past cultural evolutionary process. The social planner has an intervention that promotes changing from U to D. The social planner intervenes and reaches some people, but not everyone. The intervention could be, for instance, new subsidies that incentivise switching from gas to electric vehicles (Holtsmark & Skonhoft 2014), a television show dramatising choices that affect the risk of HIV (Banerjee et al. 2019), or a new wage scheme designed to reduce corruption (Soraperra et al. 2019). Whatever the details, some people may choose D as a direct consequence of being exposed to the intervention. Choices of this sort constitute the **direct** effect of the intervention. Other people may choose U or D because they observe the choices of others. This second mechanism generates the **indirect** effect. The indirect effect pertains to people who either were not exposed to the intervention or were exposed but did not respond. Regardless, these people choose because of social interactions within the population after the intervention.

Indirect effects can be consistent or inconsistent with the social planner's goals (Efferson, Vogt & Fehr 2020). Consider a population in which U is eating beef, and initially everyone does this. The social planner rolls out a media campaign showing how switching from beef to plant-based alternatives, D, is good for the heart and the environment. She reaches half of the population, and 3/5 of those exposed abandon beef because of the campaign. The direct effect is $(1/2)(3/5) = 3/10$. Assume further that a third of those who initially abandoned beef give up and switch back after experiencing social pressure at grill parties with friends. If no one adopts plant-based alternatives after seeing others do so, the indirect effect is $(-1/3)(3/10) = -1/10$, and the total effect of the intervention is $1/5$. Imagine instead that, aside from our inconstant vegetarians who cave at grill parties, 4/7 of the people who did not initially abandon beef because of the campaign do so later after observing others eating

veggie burgers. The indirect effect is $-1/10 + (4/7)(7/10) = 3/10$, and the total effect is $3/5$.

From the social planner's perspective, the ideal scenario is a small intervention that activates a large, net positive, indirect effect. If this happens, endogenous cultural evolution dramatically amplifies the effects of the intervention. This amplifying dynamic promises to help with important challenges. A large positive indirect effect means that social processes within the population produce most of the behaviour change, which implies the social planner is using limited resources efficiently. In addition, in ethical terms, leaning heavily on endogenous mechanisms might mitigate concerns about the social planner imposing her values on others. The social planner does not aggress so much as create a situation conducive to people deciding to change behaviour for themselves. In this sense, positive indirect effects are a cultural evolutionary analogue to nudging (Thaler & Sunstein 2021). To examine these ideas, we turn to female genital cutting, a domain in which the discussion has been remarkably explicit about the value of indirect effects (Mackie & LeJeune 2009, Cloward 2016), and we examine the role of heterogeneity as we do so.

4 The social planner's fundamental tension

The social planner who wants to activate cultural evolution faces a fundamental tension, and female genital cutting reflects the tension especially clearly (Shell-Duncan & Hernland 2000, Gruenbaum 2001, Boddy 2007, Cloward 2016, Lawson & Gibson 2022). Some cultural groups practise cutting. People who prioritise tolerance of cultural differences might conclude, especially if they have cultural backgrounds without cutting, to stay out of any debates about cutting. People who prioritise protecting universal human rights, particularly those related to gender equality, might decide to advocate for an end to cutting (Shell-Duncan 2008). People who place an equally high value on intercultural tolerance, equality, and human rights are torn between incompatible values (Cloward 2016). To put the matter bluntly, to what extent is a social planner who wants an end to cutting correct, and to what extent does

she impose her own culturally evolved preferences on others? One's answer to this question will determine if one views the social planner's efforts as socially beneficial or just another example of cultural imperialism.

Specifically, a longstanding interpretation of female genital cutting is that families need to coordinate their choices to ensure good marriage prospects for their daughters (Mackie 1996, Mackie & LeJeune 2009, Vogt et al. 2016). Coordination incentives ensure that both a tradition of cutting (U) and a tradition of not cutting (D) are equilibria. If a population is in the cutting equilibrium, individual families cannot afford to deviate, and so cutting persists. The social planner's task is to convince sufficiently many families to abandon cutting in a short period of time. Doing so should tip the population into the basin of attraction for not cutting, and the prediction is that everyone will eventually abandon cutting without much additional effort on the part of the social planner.

Assuming this is feasible, what could a social planner mean when she says she is pushing for beneficial tipping? Table 2 presents two views of a coordination game. First, decision makers in the population actually play the game, and they face homogeneous incentives to coordinate. Everyone has the same preferences, represented by a, b, c, d, with $a > d$ and $b > c$. One can easily verify that $(b - c)/(a - d + b - c)$ is the tipping point. Specifically, conditional on one's partner choosing U, $b - c$ is the cost of miscoordinating. Equivalently, one could have chosen U and received b. If instead one chose D and received c, $b - c$ is the loss associated with this decision. Analogously, conditional on one's partner choosing D, $a - d$ is the cost of miscoordinating by choosing U instead of D. Under simplifying assumptions about how people form beliefs about the choices of potential partners (Mäs & Nax 2016), the tipping point is the loss from choosing D and miscoordinating relative to the sum of both types of miscoordination cost. If the tipping point is large, decision makers have preferences such that the cost of choosing D and miscoordinating is bigger than the cost of choosing U and miscoordinating. If the tipping point is small, the opposite holds. In terms of aggregate dynamics, if the proportion choosing D is larger than the tipping point

Decision makers			Social planner		
	D	U		D	U
D	a	c	D	a_P	c_P
U	d	b	U	d_P	b_P

Table 2: Row player payoffs for two views of a coordination game. Decision makers in the population actually play the game, and they face homogeneous incentives to coordinate. The social planner does not necessarily play the game herself, but she has a view of the game that reflects a mix of the incentives people face and her own preferences about how she would like people to behave.

after an intervention, everyone should eventually choose D. If the proportion is less than the tipping point, any direct effects from the intervention are temporary, and the population should eventually converge on everyone choosing U.

The social planner has her own view of the game that may be more or less congruent with those of the people. We distinguish the preferences of the social planner from those of the people with the subscript P (Table 2). The social planner's view reflects the coordination incentives she knows people face, i.e. $a_P > d_P$ and $b_P > c_P$. In this way, the social planner's view is congruent with that of decision makers. In addition, however, the social planner favours D over U. Specifically, the social planner prefers that people coordinate on D instead of U, i.e. $a_P > b_P$, and she views the cost of choosing D when a partner plays U to be less than the cost of choosing U when a partner plays D, i.e. $b_P - c_P < a_P - d_P$. We assume the social planner has these preferences regardless of the preferences that hold among decision makers. This is simply another way of saying the social planner wants people to choose D instead of U.

If $a > b$ and $b - c < a - d$, decision makers and the social planner have fully congruent preferences. Both view D as payoff dominant and risk dominant (Harsanyi & Selten 1988). Intuitively, everyone prefers coordinating on D over coordinating on U (i.e. D payoff dominant), and everyone views the cost of miscoordinating when playing D to be less than the cost of miscoordinating when playing U (i.e. D risk dominant). The tipping point is less than

a half, and an intervention with a relatively small direct effect is theoretically sufficient to initiate a cultural evolutionary trajectory that is unambiguously for good. In this scenario, the cutting population is trapped in an unequivocally harmful equilibrium, and the social planner helps people help themselves by disrupting this equilibrium. The puzzle here is that empirical evidence shows that populations tend to converge on risk-dominant equilibria on their own. So, why would the population be stuck in the other equilibrium in the first place (Platteau et al. 2018)?

Alternatively, consider $a > b$ but $b - c > a - d$. Decision makers agree with the social planner that coordinating on D is better than coordinating on U. They disagree, however, about the relative costs of miscoordinating. Decision makers consider choosing D when their partners play U to be a costlier mistake than choosing U when their partners play D. U is thus risk dominant. Equivalently, the tipping point is larger than a half, and thus the social planner needs a large intervention. Moreover, when payoff dominance and risk dominance favour different equilibria, which is the case here, experimental results show that risk dominance tends to control final outcomes (Weber 2006) unless special conditions are in place (Devetag & Ortmann 2007, Riedl et al. 2016).

In this case, is behaviour change for good from the people's perspective? Yes and no. Yes, like the social planner, people favour coordinating on D over coordinating on U. No, unlike the social planner, they consider choosing D and miscoordinating to be a costlier mistake than choosing U and miscoordinating. Insofar as experimental evidence provides guidance, it suggests that this kind of cost asymmetry should exert a strong pull on cultural evolution in favour of U (Weber 2006, Efferson, McKay & Fehr 2020). The social planner, by extension, requires an aggressive intervention, and any transition to a new equilibrium will likely take longer and cause more pain along the way than the social planner's view acknowledges.

Two other cases remain. If $a < b$ but $b - c < a - d$, then U is payoff-dominant, but D is risk dominant. The tipping point is less than a half, the social planner requires a relatively small intervention, and any transition to the social planner's preferred equilibrium is likely

to be relatively easy. Ex post, however, assuming people's preferences do not change, the population will be in an equilibrium the social planner likes, but the people do not. Finally, if $a < b$ and $b - c > a - d$, then U is both payoff-dominant and risk dominant. The tipping point is relatively large, and the social planner requires a correspondingly large intervention. If she manages to push the population across the tipping point, she initiates a cultural transition, likely protracted and difficult, to an equilibrium people dislike.

Importantly, in all four of the hypothetical scenarios discussed here, a tipping point exists. Consequently, the social planner has an opportunity to activate a self-reinforcing cultural evolutionary process in which the population transitions from cutting to not cutting. In only one of the scenarios, however, is such a transition unequivocally "for good" from the people's perspective. Importantly, we have ignored many considerations. For example, cutting is a choice that adults make for children. Thus, the people making choices do not deal with most of the consequences that follow from those choices, a situation that typically leads to socially inefficient outcomes (Bowles 2009). Moreover, the people who do face these consequences are especially vulnerable when the choices are made for them. Finally, we have ignored the possibility that people in a cutting society, once they have experienced the non-cutting counterfactual, change their preferences. By ignoring these issues, we do not mean to suggest they are unimportant. Rather, we have attempted to outline a taxonomy of the basic ways a social planner's preferences may or may not cohere with those of the people over whom she has influence.

We will not presume to settle the ethics of the four scenarios above. We would, however, like to highlight the following. Whatever one's ethical judgements, the core practical question about the potential for tipping is settled in all four cases. Although the tipping point varies based on people's preferences, it always exists as a direct consequence of the assumption that the people in the population are all the same. However, as soon as we accept the trivial possibility that people may differ, even the practical question is no longer settled.

5 Ordinary heterogeneity and the individual-population disconnect

How does heterogeneity at the individual level affect cultural evolution at the aggregate level and by extension the scope for a social planner to recruit cultural evolutionary processes? As we discuss below, the inherent complexity of the problem almost ensures that for now we do not have a comprehensive answer to this question. That said, theory and empirical work are at least relatively well-developed when decision making involves conformity and coordination. Consider the following simple model (Granovetter 1978). The population proportion choosing D at time t is q_t. Each decision maker chooses D with certainty at time $t+1$ if q_t is sufficiently large, or equivalently if D was sufficiently common in the recent past. Otherwise, the decision maker chooses U with certainty. People are heterogeneous in the precise sense that they vary in terms of the q_t values that induce them to switch between D and U (Fig. 1). These heterogeneous switching values are often called "thresholds", and each individual can have a unique threshold.

This model has at least two interpretations at the individual level. First, everyone faces incentives to coordinate their choices, but they have heterogeneous preferences (Andreoni et al. 2021, Efferson 2021). Specifically, people pair off randomly to play a coordination game. Each person's threshold is an indifference point. If she believes the proportion of potential game partners choosing D is equal to her threshold, she is indifferent between D and U. Further assume that everyone takes q_t as their belief about the probability a partner will play D in $t+1$, everyone chooses the option with the highest expected value given this belief, and everyone chooses D when indifferent. This simple model of myopic belief formation and associated decision making is consistent with an extraordinary number of observed choices in coordination game experiments (Mäs & Nax 2016). For present purposes, the key point is that everyone wants to coordinate, but they also vary in a completely ordinary way.

The second interpretation is that everyone exhibits an extreme conformist bias (Boyd

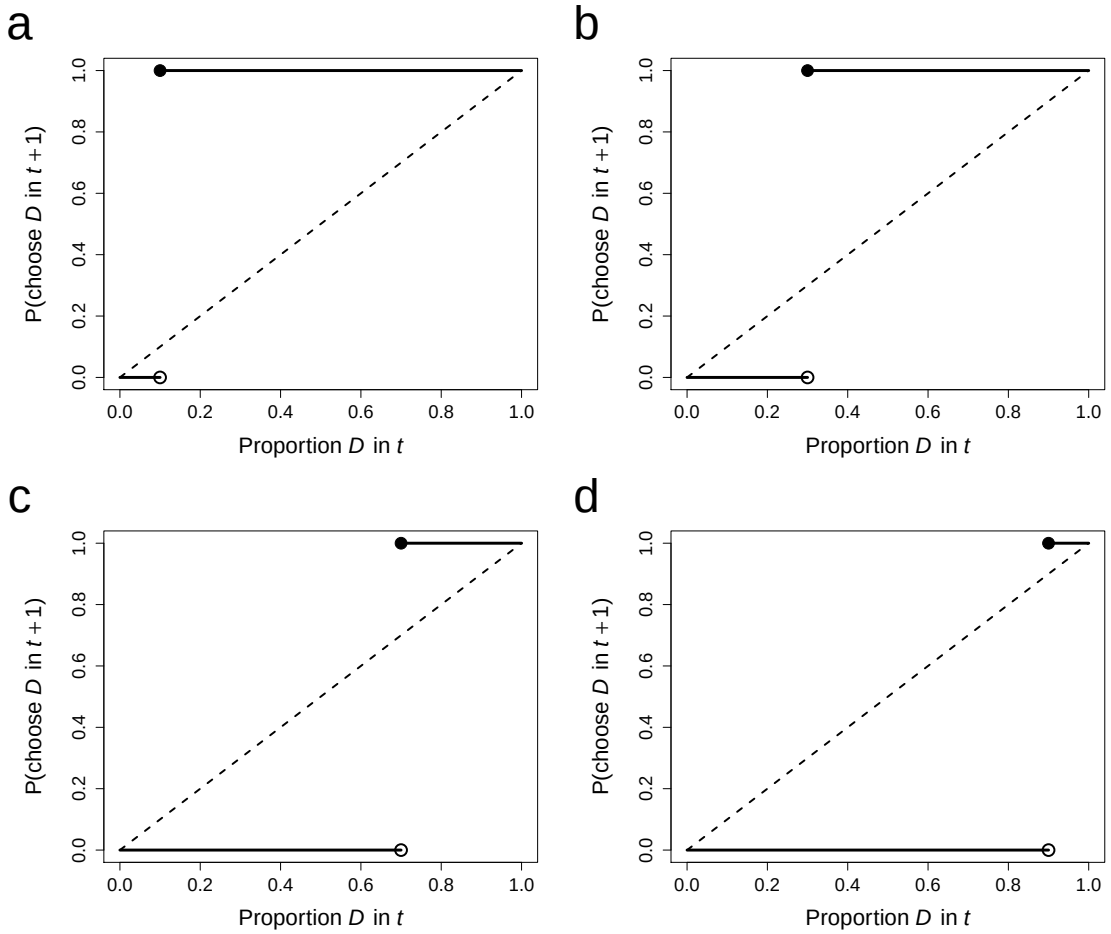

Figure 1: An example of four hypothetical individuals with heterogeneous responses to choice frequencies. **a** - **d**, Each panel shows a single individual's social learning strategy as a function mapping the population proportion choosing D in t to choice probabilities for the individual in $t+1$. We can interpret these functions as representing people with heterogeneous preferences playing a coordination game or as extreme conformists with heterogeneous content biases. Individuals **a** and **b**, for example, prefer coordinating on D over coordinating on U, or equivalently they have content biases favouring D. Individuals **c** and **d** have the opposite preferences and content biases. The diagonal, which represents unbiased social learning (Boyd & Richerson 1985), is shown for reference.

& Richerson 1985, Efferson, Lalive, Richerson, McElreath & Lubell 2008, Morgan & Laland 2012, Muthukrishna et al. 2016) mixed with heterogeneous content biases (Morin 2016, Kendal et al. 2018). Conformity is a disproportionate tendency to follow the majority (Boyd

& Richerson 1985, Efferson, Lalive, Richerson, McElreath & Lubell 2008). If $q_t > 0.5$, a decision maker chooses D with some probability greater than q_t. If $q_t < 0.5$, the decision maker chooses D with some probability less than q_t. These disproportionate choice probabilities create the multiple equilibria associated with conformity. The threshold model is a model of extreme conformity because choice probabilities increase in q_t and only take the disproportionate values of 0 and 1. That said, the model is not a traditional model of conformity because it generalises the notion of a majority to account for idiosyncratic content biases. If Sonja has a content bias that favours U, she may choose D if and only if $q_t \geq 0.75$. If Lukas has a content bias that favours D, he may choose D if and only if $q_t \geq 0.32$. Individuals do not simply compare q_t to a uniform benchmark of 0.5. Instead, they compare q_t to their respective threshold values.

Whichever interpretation one prefers, cultural evolution is the same in both cases, and it hinges on the structure of heterogeneity. If F is the cumulative distribution function for threshold values, culture evolves (Granovetter 1978) as $q_{t+1} = F(q_t)$. Depending on the shape of F, all sorts of dynamics can follow. For example, if thresholds are symmetrically and unimodally distributed around 0.5 (Fig. 2a), content biases exist, but they are unsystematic and relatively weak. They are unsystematic because the average threshold is 0.5, and they are relatively weak because most thresholds values are relatively close to 0.5. Culture evolves as we expect when conformity is important. The population converges, depending on where it starts, on either everyone choosing D or everyone choosing U (Fig. 2b). This is the traditional view of conformity as a mechanism that stabilises cultural differences between groups (Boyd & Richerson 1985), a form of population structure hypothesised to have far-reaching consequences for the evolution of human social behaviour (Boyd et al. 2011, White et al. 2021).

In stark contrast, if thresholds are symmetrically and bimodally distributed around 0.5 (Fig. 2c), the population converges on a stable mix of choices regardless of initial conditions (Fig. 2d). Content biases are unsystematic but groupish. The average threshold is 0.5, but

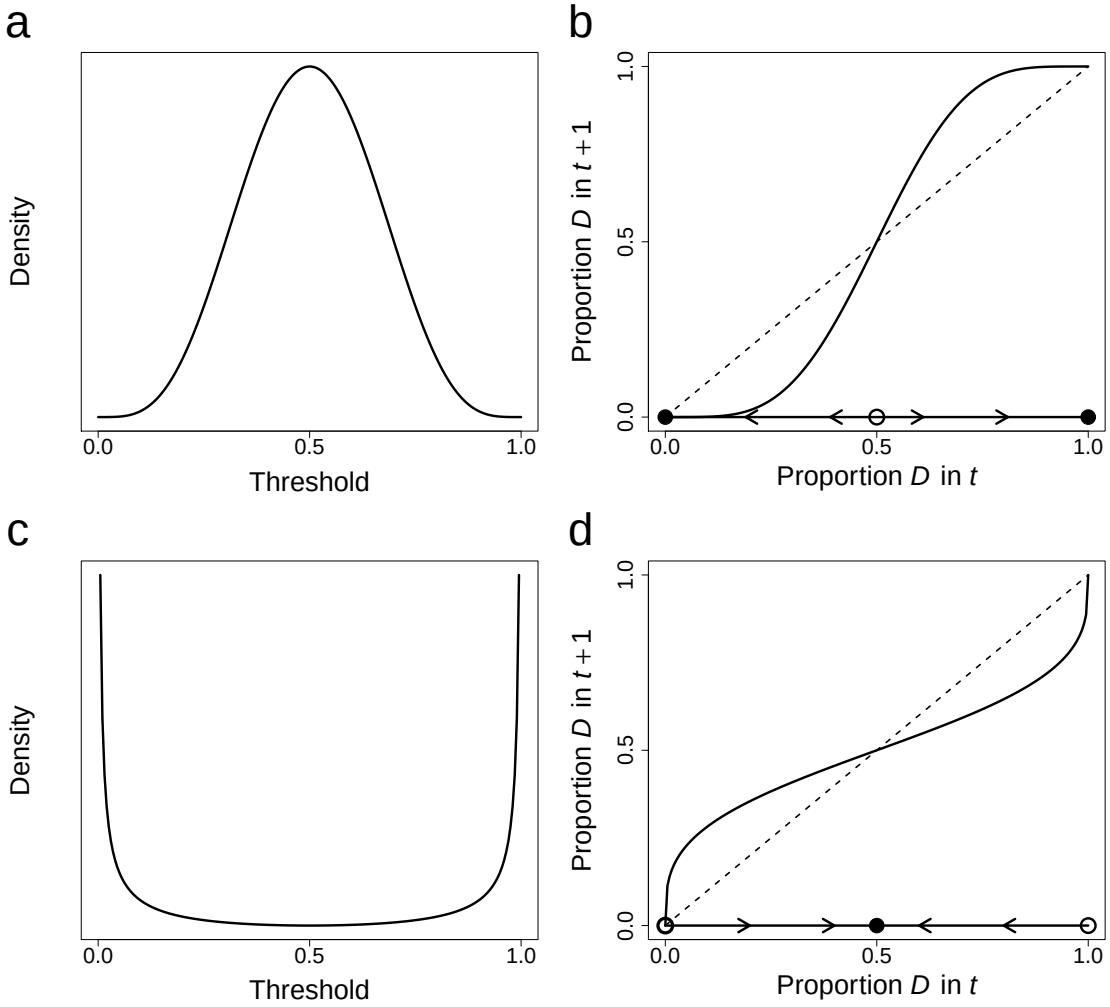

Figure 2: **a**, An unbiased, symmetric, and unimodal distribution of thresholds. **b**, The cumulative distribution function associated with **a** provides a handy graphical approach (e.g. cobwebbing) for characterising the steady states of the cultural evolutionary system. The line along the bottom with arrows and points is the phase space for the system, with solid points showing locally stable equilibria and the open point the unstable equilibrium (i.e. tipping point). **c**, An unbiased, symmetric, and bimodal distribution of thresholds. **d**, The cumulative distribution function and phase space associated with **c**. Notice the two distributions produce the same equilibria, but the stability properties of the equilibria are mirror images of each other. The social planner can rely on conformity and coordination to support behaviour change in the first case (**a**,**b**) but not the second (**c**,**d**).

the distribution of thresholds suggests two distinct groups. This groupish quality effectively destroys the multiple stable equilibria typically associated with conformity and coordination. By extension, it destroys the theoretical potential for a social planner to tip the population from one equilibrium to another. This idea is consistent with theoretical and empirical studies suggesting that polarized opinions, oppositional identities, and outgroup aversion undermine the scope for a social planner to recruit cultural evolutionary processes to promote widespread behaviour change (Smaldino et al. 2017, Efferson, Vogt & Fehr 2020, Moya et al. 2020, Smaldino & Jones 2021, Ehret et al. 2022).

Although heterogeneity can take many forms other than the simple examples here, these examples are sufficient to make a crucial point. Under ordinary heterogeneity, the link between social learning at the individual level and cultural evolution at the population level can become complex and counterintuitive. Even if everyone is an extreme conformist, the population can still converge on maximum heterogeneity, exactly the opposite of what we associate with conformity.

For our social planner, the practical implication is that conformity at the individual level does not imply behavioural homogeneity and the potential for social tipping at the population level. Moreover, behavioural heterogeneity at the population level does not imply that decision making is not conformist at the individual level. This potential disconnect between the individual and aggregate levels has proven important in relation to programmes promoting the abandonment of female genital cutting. For several years many programmes have been explicitly organised around the assumption that cutting is subject to conformity and coordination incentives (Cloward 2016). The social planner thus faces an opportunity to tip a population from the cutting to the non-cutting equilibrium.

Recent empirical studies, however, have found that attitudes and practices related to cutting are highly variable at surprisingly local scales (Bellemare et al. 2015, Efferson et al. 2015, Cloward 2016, Vogt et al. 2016, 2017, Platteau et al. 2018). This finding suggests that cutting and non-cutting families are in regular contact with each other, which is difficult to reconcile

with distinct cutting and non-cutting equilibria. A naive interpretation would lead one to conclude that families do not care about behaving like those around them, but additional evidence clearly indicates this conclusion is wrong (Hayford 2005, Efferson et al. 2015, Novak 2020). People do care about behaving like those around them, but they also seem to have heterogeneous thresholds (Hayford 2005, Novak 2020) with, perhaps unsurprisingly, ethnicity playing some moderating role (Howard & Gibson 2017). In this sense, recent research on cutting demonstrates how the questions the social planner must ask are potentially more nuanced than simply asking, do people want to coordinate and conform?

Even if the social planner knows people want to coordinate, she still must ask if tipping is possible. If possible, she must also ask how many people to target with her intervention and, because people vary, which specific people to target. If she knows people's preferences, she must decide whether to target people who are relatively amenable to adopting her preferred behaviour, people who are relatively resistant, or people in the middle (Efferson, Vogt & Fehr 2020). Relatedly, if the social planner does not exactly know people's preferences, she might have observable proxy variables like age or gender that correlate with preferences (Schief et al. 2021). She must then decide whom to target based on these proxies. Does she target men in their twenties, middle-aged parents, or 80-year-old women?

Recall that we expect at least two forms of heterogeneity to affect outcomes. Heterogeneity in response to the intervention should affect the direct effect, and heterogeneity in social influence and social learning should affect the indirect effect given a direct effect. Although we know of no systematic empirical work on how the social planner's choices integrate these two forms of heterogeneity to create a total effect, the threshold model readily accommodates relevant modifications (Efferson, Vogt & Fehr 2020).

Assume the probability of choosing D because of exposure to the intervention is some function of thresholds. A natural assumption is that the probability of choosing D declines as thresholds increase; relatively resistant people are relatively unlikely to respond to the intervention. With this in place, the social planner must choose whom to target. Here

we consider two extreme possibilities, the most amenable segment and the most resistant segment.

This simple setting captures a basic trade-off for the social planner. Assume the social planner implements an intervention of size ϕ, where ϕ is the proportion of the population exposed to the intervention. Targeting the most amenable ϕ individuals maximises the direct effect because relatively amenable individuals are relatively likely to respond to the intervention. This strategy, however, minimises the indirect effect because the individuals not targeted make up the most resistant $1 - \phi$ proportion of the population. Targeting the most resistant ϕ does the opposite; it minimises the direct effect but maximises the indirect effect given the direct effect. The social planner wants to maximise the total effect, and this requires her to resolve the trade-off between the direct and indirect effects in the best way. How to do so depends on the details (Efferson, Vogt & Fehr 2020, Schimmelpfennig et al. 2021).

6 Potential complexity and edutainment as indiscriminate motivator

Although we have concentrated on the complexity of activating cultural change under frequency-dependent social learning in heterogeneous populations, we suspect that the potential for complexity is far more extensive than this. We suspect this simply because the translation from social learning to cultural evolution involves incredible scope for complexity. To illustrate, in the box we describe a simple learning problem in a population of 10 individuals.

Despite the frankly unacceptable simplicity assumed for this learning problem, a single decision maker has at her disposal 2^{512} possible strategies, a number vastly larger than the number of atoms in the universe. If we allow decision makers to exhibit different strategies from each other, which empirical evidence indicates we should (Mesoudi et al. 2016, Kendal et al. 2018), the number of configurations our humble population of 10 can take explodes further. If we allow learning based on two recent points in time instead of just one, problems

Individual i has two choice options, $\{0, 1\}$. At time t, each individual faces one of two situations, also $\{0, 1\}$. Choice 0 ($C_{it} = 0$) is best if the individual is facing situation 0 ($S_{it} = 0$), and choice 1 ($C_{it} = 1$) is best if facing situation 1 ($S_{it} = 1$). The population consists of 10 individuals divided into two groups of five. Everyone in a group is facing the same situation, but the two groups may or may not face different situations. Individuals can never know with certainty which situations they face, but they can improve their chances of making the best choices by learning. To learn, at the beginning of t, each individual observes the following.

(1) The choice of each ingroup member in $t - 1$, which includes one's own choice.

(2) The choice in $t - 1$ generating the highest ingroup payoff and the choice generating the lowest.

(3) The choice of each outgroup member in $t - 1$.

(4) The choice in $t - 1$ generating the highest outgroup payoff and the choice generating the lowest.

(5) Private information about the individual's current situation, which takes one of two possible values.

A strategy is a function, f, from the set of possible observations, defined jointly over (1) - (5), to probability distributions over choice options. Let the set of possible observations be Ω. The set Ω includes 512 possibilities. Thus, if we limit attention to deterministic strategies that specify a choice with certainty, $f : \Omega \to \{0, 1\}$, this setting admits 2^{512} possible strategies.

with more than two choice options, and populations larger than 10, the potential complexity becomes quite staggering. Complexity increases further still if we treat social learning as a two-step process. First, choose social ties to form a network. Second, respond to information distributed on the resulting network. Even though we ignore all of these obvious and important sources of complexity in our stylised problem, we are still left with a vast space of 2^{512} strategies that could in principle shape how learning unfolds.

We know that the distribution of learning strategies in a population drives cultural evolution. What this actually means will often be a challenging question because the set of possibilities will typically be genuinely, beyond astronomically huge. Although the study of social learning and gene-culture coevolution has experienced an impressive transition from

theory to theory plus empiricism in the last 15 years or so, most of what could be happening in evolving cultures remains unexamined and even unimagined. For now at least, this intrinsic complexity limits our understanding of cultural evolutionary processes in general and our understanding of how to activate cultural evolution for good specifically.

That said, taking this potential complexity as given, we would like to suggest the potential value of edutainment for the social planner. "Edutainment" refers to a communications strategy that merges entertainment and education. A television show, a novel, or a bard's tale around the campfire could all qualify. The key idea is that people join for the story and the entertainment it provides, but along the way they encounter values, ideas, and behaviours that are new and packaged in a compelling way (DellaVigna & La Ferrara 2015). The hope is that the social planner can use edutainment to induce a shift in values and norms more effectively than with an approach that simply provides information (Singhal & Rogers 2003). For example, instead of simply informing people that certain behaviours affect the risk of HIV, dramatise the behaviours in a show on MTV (Banerjee et al. 2019).

Edutainment has the potential to solve two important problems for the social planner. First, it can limit biased exposure to interventions. In particular, the demand for entertainment seems to be ubiquitous (DellaVigna & La Ferrara 2015), and modern technologies allow us to disseminate media at marginal costs close to zero (La Ferrara 2016). Both of these characteristics imply scope for the social planner to reach a broad and diverse cross section of society. Existing theory suggests this should be valuable because biased targets in general, but especially amenable targets, undermine the total effect of an intervention by limiting either the direct effect or the indirect effect (Efferson, Vogt & Fehr 2020, Gavrilets 2020). Crucially, however, in a world where social media algorithms feed us material consistent with our past choices, we risk losing entertainment's indiscriminate appeal.

Second, edutainment can prepare the population for change in terms of values and preferences. Existing theory suggests that the social planner can make choices that influence indirect effects, but the pre-existing values and preferences in a population are often far more

important (Efferson, Vogt & Fehr 2020). This further suggests the potential value of using edutainment to provoke an initial shift in preferences, perhaps with a subsequent intervention designed specifically for behaviour change via indirect effects.

What is the evidence that edutainment causes changes in values, preferences, and even behaviour? Early studies suffered from potentially serious selection bias, and thus they offered little scope for causal inference (e.g. Johnson et al. 2002). Starting in the 2000's, however, a number of studies developed clever identification strategies based on quasi-random variation in access to cable television (e.g. DellaVigna & Kaplan 2007, Jensen & Oster 2009, La Ferrara et al. 2012) and radio (e.g. DellaVigna et al. 2014). We now have evidence for edutainment positively affecting various development outcomes related to conflict resolution, family planning, gender norms, domestic violence, education, and financial literacy (Paluck 2012, DellaVigna & La Ferrara 2015, La Ferrara 2016). More recently, research has turned to identifying what exactly makes a compelling story (Vogt et al. 2016, Banerjee et al. 2019, Berl et al. 2021, Singh et al. 2021), and what conditions amplify the effects of exposure to edutainment (Efferson & Vogt 2018).

7 Conclusion

The appeal of activating cultural evolution for good rests squarely on the possibility that endogenous mechanisms can do some, even most of the work that leads to behaviour change. This possibility implies that a social planner with a light touch can achieve impressive results. We have proof of concept for several relevant ideas. First, some people want to conform at least some of the time (Muthukrishna et al. 2016). Second, whatever the exact underlying mechanisms, cultural evolutionary processes can create local homogeneity in behaviour (Young & Burke 2001) coupled with global heterogeneity (Eugster et al. 2017), a kind of aggregate-level signature for settings with tipping potential. Third, when coordination incentives dominate, social tipping occurs reliably so long as the shock to the system

is sufficiently large (Centola et al. 2018, Andreoni et al. 2021) and other mechanisms like group identities (Ehret et al. 2022) do not interfere. Finally, we have several historical examples of punctuated behaviour change that seem to reflect rapid cultural transitions from one equilibrium to another (Young 2015).

At the same time, social planners have explicitly embraced tipping in some domains, only for empirical work to provide little reason to believe that tipping is feasible (Efferson et al. 2015, Bellemare et al. 2015) and little evidence that interventions are having the anticipated effects (Camilotti 2016, Platteau et al. 2018). Completely ordinary forms of heterogeneity seem to be playing some role (Hayford 2005, Novak 2020). Consistent with these examples, experimental studies show a sometimes bizarre degree of heterogeneity in social learning strategies. For example, when following the majority is a transparently beneficial use of social information, many people follow the majority, and many do not (Efferson, Lalive, Richerson, McElreath & Lubell 2008). Moreover, when following the majority is transparently detrimental, many people do not follow the majority, and many people do (Goeree & Yariv 2015). This kind of variation (Efferson, Vogt & Fehr 2020, Fig. 1) is just one example of how heterogeneity can create an apparent disconnect between individual-level social learning and aggregate-level cultural evolution.

With respect to policy, a key challenge is thus to develop practical methods for evaluating diverse situations in terms of the scope for tipping. One approach would be to measure the value of conforming as perceived by individuals in the population. This is most likely to be valuable if coupled with a reliable estimate of the preference distribution in the population. An alternative approach would be to jump straight to the aggregate level and look for evidence of multiple equilibria, specifically local homogeneity and global heterogeneity in attitudes, beliefs, norms, and behaviour (Young 2015). Looking for this aggregate-level signature is intuitively appealing because social planners are mainly interested in society, not specific individuals. At least two issues are crucial. First, what does local mean? Is it a geographic concept for the population in question, an ethnic concept, or something else?

Second, is global heterogeneity actually cultural variation rather than, say, institutional or environmental variation? Identifying cultural variation empirically can be difficult precisely because cultural boundaries often coincide with other kinds of boundaries, but we now have compelling examples of how to do so (Eugster et al. 2017, Lowes et al. 2017). In any case, if the aggregate cultural signature is present, the social planner has a relatively secure basis for relying on tipping as a way to activate cultural evolution for good or bad, whatever the case may be.

8 Acknowledgements

We would like to thank the Swiss National Science Foundation for support (Grant Nr. 100018_185417).

References

Acerbi, A. (2019), 'Cognitive attraction and online misinformation', *Palgrave Communications* **5**(1), 1–7.

Allcott, H. (2011), 'Social norms and energy conservation', *Journal of Public Economics* **95**(9-10), 1082–1095.

Alvergne, A. & Stevens, R. (2021), 'Cultural change beyond adoption dynamics: Evolutionary approaches to the discontinuation of contraception', *Evolutionary Human Sciences* pp. 1–45.

Andreoni, J., Nikiforakis, N. & Siegenthaler, S. (2021), 'Predicting social tipping and norm change in controlled experiments', *Proceedings of the National Academy of Sciences* **118**(16).

Arnot, M., Brandl, E., Campbell, O. L. K., Chen, Y., Du, J., Dyble, M., Emmott, E. H., Ge, E., Kretschmer, L. D. & Mace, R. (2020), 'How evolutionary behavioural sciences

can help us understand behaviour in a pandemic', *Evolution, Medicine, and Public Health* **2020**(1), 264–278.

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2013), 'The diffusion of microfinance', *Science* **341**(6144).

Banerjee, A., La Ferrara, E. & Orozco-Olvera, V. H. (2019), The entertaining way to behavioral change: Fighting hiv with mtv, Technical report, National Bureau of Economic Research.

Barkoczi, D. & Galesic, M. (2016), 'Social learning strategies modify the effect of network structure on group performance', *Nature Communications* **7**, 13109.

Bellemare, M. F., Novak, L. & Steinmetz, T. L. (2015), 'All in the family: Explaining the persistence of female genital cutting in West Africa', *Journal of Development Economics* **116**, 252–265.

Berl, R. E., Samarasinghe, A. N., Roberts, S. G., Jordan, F. M. & Gavin, M. C. (2021), 'Prestige and content biases together shape the cultural transmission of narratives', *Evolutionary Human Sciences* **3**.

Boddy, J. (2007), *Civilizing Women: British Crusades in Colonial Sudan*, Princeton University Press.

Bowles, S. (2009), *Microeconomics: Behavior, Institutions, and Evolution*, Princeton University Press.

Boyd, R. & Richerson, P. J. (1985), *Culture and the Evolutionary Process*, University of Chicago press.

Boyd, R. & Richerson, P. J. (1987), 'The evolution of ethnic markers', *Cultural Anthropology* **2**(1), 65–79.

Boyd, R., Richerson, P. J. & Henrich, J. (2011), 'Rapid cultural adaptation can facilitate the evolution of large-scale cooperation', *Behavioral Ecology and Sociobiology* **65**(3), 431–444.

Camilotti, G. (2016), 'Interventions to stop female genital cutting and the evolution of the custom: Evidence on age at cutting in Senegal', *Journal of African Economies* **25**(1), 133–158.

Castilla-Rho, J. C., Rojas, R., Andersen, M. S., Holley, C. & Mariethoz, G. (2017), 'Social tipping points in global groundwater management', *Nature Human Behaviour* **1**(9), 640–649.

Centola, D. (2011), 'An experimental study of homophily in the adoption of health behavior', *Science* **334**(6060), 1269–1272.

Centola, D., Becker, J., Brackbill, D. & Baronchelli, A. (2018), 'Experimental evidence for tipping points in social convention', *Science* **360**(6393), 1116–1119.

Christakis, N. A. & Fowler, J. H. (2008), 'The collective dynamics of smoking in a large social network', *New England Journal of Medicine* **358**(21), 2249–2258.

Christakis, N. A. & Fowler, J. H. (2013), 'Social contagion theory: Examining dynamic social networks and human behavior', *Statistics in Medicine* **32**(4), 556–577.

Cloward, K. (2016), *When Norms Collide: Local Responses to Activism Against Female Genital Mutilation and Early Marriage*, Oxford University Press.

Davidai, S., Gilovich, T. & Ross, L. D. (2012), 'The meaning of default options for potential organ donors', *Proceedings of the National Academy of Sciences* **109**(38), 15201–15205.

DellaVigna, S., Enikolopov, R., Mironova, V., Petrova, M. & Zhuravskaya, E. (2014), 'Cross-border media and nationalism: Evidence from serbian radio in croatia', *American Economic Journal: Applied Economics* **6**(3), 103–32.

DellaVigna, S. & Kaplan, E. (2007), 'The fox news effect: Media bias and voting', *The Quarterly Journal of Economics* **122**(3), 1187–1234.

DellaVigna, S. & La Ferrara, E. (2015), Economic and social impacts of the media, *in* 'Handbook of media economics', Vol. 1, Elsevier, pp. 723–768.

Derex, M. & Boyd, R. (2016), 'Partial connectivity increases cultural accumulation within groups', *Proceedings of the National Academy of Sciences* **113**(11), 2982–2987.

Devetag, G. & Ortmann, A. (2007), 'When and why? a critical survey on coordination failure in the laboratory', *Experimental Economics* **10**(3), 331–344.

Efferson, C. (2021), 'Policy to activate cultural change to amplify policy', *Proceedings of the National Academy of Sciences* **118**(23).

Efferson, C., Lalive, R. & Fehr, E. (2008), 'The coevolution of cultural groups and ingroup favoritism', *Science* **321**(5897), 1844–1849.

Efferson, C., Lalive, R., Richerson, P. J., McElreath, R. & Lubell, M. (2008), 'Conformists and mavericks: The empirics of frequency-dependent cultural transmission', *Evolution and Human Behavior* **29**(1), 56–64.

Efferson, C., McKay, R. & Fehr, E. (2020), 'The evolution of distorted beliefs vs. mistaken choices under asymmetric error costs', *Evolutionary Human Sciences* **2**.

Efferson, C. & Vogt, S. (2018), 'Behavioural homogenization with spillovers in a normative domain', *Proceedings of the Royal Society B: Biological Sciences* **285**(1879), 20180492.

Efferson, C., Vogt, S., Elhadi, A., Ahmed, H. E. F. & Fehr, E. (2015), 'Female genital cutting is not a social coordination norm', *Science* **349**(6255), 1446–1447.

Efferson, C., Vogt, S. & Fehr, E. (2020), 'The promise and the peril of using social influence to reverse harmful traditions', *Nature Human Behaviour* **4**(1), 55–68.

Ehret, S., Constantino, S. M., Weber, E. U., Efferson, C. & Vogt, S. (2022), 'Group identities can undermine social tipping after intervention', *Nature Human Behaviour* pp. 1–11.

Eugster, B., Lalive, R., Steinhauer, A. & Zweimüller, J. (2017), 'Culture, work attitudes, and job search: Evidence from the Swiss language border', *Journal of the European Economic Association* **15**(5), 1056–1100.

Farmer, J. D., Hepburn, C., Ives, M. C., Hale, T., Wetzer, T., Mealy, P., Rafaty, R., Srivastav, S. & Way, R. (2019), 'Sensitive intervention points in the post-carbon transition', *Science* **364**(6436), 132–134.

Gavrilets, S. (2020), 'The dynamics of injunctive social norms', *Evolutionary Human Sciences* **2**.

Goeree, J. K. & Yariv, L. (2015), 'Conformity in the lab', *Journal of the Economic Science Association* **1**(1), 15–28.

Golub, B. & Jackson, M. O. (2010), 'Naive learning in social networks and the wisdom of crowds', *American Economic Journal: Microeconomics* **2**(1), 112–49.

Granovetter, M. (1978), 'Threshold models of collective behavior', *American Journal of Sociology* **83**(6), 1420–1443.

Gruenbaum, E. (2001), *The Female Circumcision Controversy: An Anthropological Perspective*, University of Pennsylvania Press.

Harsanyi, J. C. & Selten, R. (1988), *A General Theory of Equilibrium Selection in Games*, The MIT Press.

Hayford, S. R. (2005), 'Conformity and change: Community effects on female genital cutting in Kenya', *Journal of Health and Social Behavior* **46**(2), 121–140.

Henrich, J. (2001), 'Cultural transmission and the diffusion of innovations: Adoption dynamics indicate that biased cultural transmission is the predominate force in behavioral change', *American Anthropologist* **103**(4), 992–1013.

Holtsmark, B. & Skonhoft, A. (2014), 'The norwegian support and subsidy policy of electric cars. should it be adopted by other countries?', *Environmental Science & Policy* **42**, 160–168.

Howard, J. A. & Gibson, M. A. (2017), 'Frequency-dependent female genital cutting behaviour confers evolutionary fitness benefits', *Nature Ecology & Evolution* **1**(3), 1–6.

Jensen, R. & Oster, E. (2009), 'The power of tv: Cable television and women's status in india', *The Quarterly Journal of Economics* **124**(3), 1057–1094.

Johnson, E. J. & Goldstein, D. (2003), 'Do defaults save lives?', *Science* **302**.

Johnson, J. G., Cohen, P., Smailes, E. M., Kasen, S. & Brook, J. S. (2002), 'Television viewing and aggressive behavior during adolescence and adulthood', *Science* **295**(5564), 2468–2471.

Karlsen, S., Mogilnicka, M., Carver, N. & Pantazis, C. (2019), 'Female genital mutilation: empirical evidence supports concerns about statistics and safeguarding', *BMJ* **364**.

Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M. & Jones, P. L. (2018), 'Social learning strategies: Bridge-building between fields', *Trends in Cognitive Sciences* **22**(7), 651–665.

La Ferrara, E. (2016), 'Mass media and social change: Can we use television to fight poverty?', *Journal of the European Economic Association* **14**(4), 791–827.

La Ferrara, E., Chong, A. & Duryea, S. (2012), 'Soap operas and fertility: Evidence from brazil', *American Economic Journal: Applied Economics* **4**(4), 1–31.

Lawson, D. & Gibson, M. (2022), Evolutionary approaches to population health: Insights on polygynous marriage, 'child marriage' and female genital mutiliation/cutting, *in* O. Burger, R. Lee & R. Sear, eds, 'Human Evolutionary Demography'.

Liebe, U., Gewinner, J. & Diekmann, A. (2021), 'Large and persistent effects of green energy defaults in the household and business sectors', *Nature Human Behaviour* **5**(5), 576–585.

Lowes, S., Nunn, N., Robinson, J. A. & Weigel, J. L. (2017), 'The evolution of culture and institutions: Evidence from the Kuba Kingdom', *Econometrica* **85**(4), 1065–1091.

Mackie, G. (1996), 'Ending footbinding and infibulation: A convention account', *American Sociological Review* pp. 999–1017.

Mackie, G. & LeJeune, J. (2009), 'Social dynamics of abandonment of harmful practices: A new look at the theory'.

Mäs, M. & Nax, H. H. (2016), 'A behavioral study of "noise" in coordination games', *Journal of Economic Theory* **162**, 195–208.

Mesoudi, A., Chang, L., Dall, S. R. & Thornton, A. (2016), 'The evolution of individual and cultural variation in social learning', *Trends in Ecology & Evolution* **31**(3), 215–225.

Morgan, T. J. H. & Laland, K. N. (2012), 'The biological bases of conformity', *Frontiers in Neuroscience* **6**, 87.

Morin, O. (2016), *How Traditions Live and Die*, Oxford University Press.

Moya, C., Cruz Y Celis Peniche, P. & Kline, Michelle A, S. P. E. (2020), 'Dynamics of behavior change in the COVID world', *American Journal of Human Biology* **32**.

Muthukrishna, M., Morgan, T. J. & Henrich, J. (2016), 'The when and who of social learning and conformist transmission', *Evolution and Human Behavior* **37**(1), 10–20.

Novak, L. (2020), 'Persistent norms and tipping points: The case of female genital cutting', *Journal of Economic Behavior & Organization* **177**, 433–474.

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., Adger, W. N., Arrow, K. J., Barrett, S. & Carpenter, S. (2016), 'Social norms as solutions', *Science* **354**(6308), 42–43.

Otto, I. M., Donges, J. F., Cremades, R., Bhowmik, A., Hewitt, R. J., Lucht, W., Rockström, J., Allerberger, F., McCaffrey, M. & Doe, S. S. (2020), 'Social tipping dynamics for stabilizing Earth's climate by 2050', *Proceedings of the National Academy of Sciences* **117**(5), 2354–2365.

Paluck, E. L. (2012), 'Media as an instrument for reconstructing communities following conflict', *Restoring Civil Societies: The Psychology of Intervention and Engagement Following Crisis* pp. 284–298.

Platteau, J.-P., Camilotti, G. & Auriol, E. (2018), 'Eradicating women-hurting customs', *Towards Gender Equity in Development* **319**.

Pritchett, L. & Sandefur, J. (2014), 'Context matters for size: why external validity claims and development practice do not mix', *Journal of Globalization and Development* **4**(2), 161–197.

Riedl, A., Rohde, I. M. & Strobel, M. (2016), 'Efficient coordination in weakest-link games', *The Review of Economic Studies* **83**(2), 737–767.

Salali, G. D., Uysal, M. S., Bozyel, G., Akpinar, E. & Aksu, A. (2022), 'Does social influence affect covid-19 vaccination intention among the unvaccinated?', *Evolutionary Human Sciences* **4**, E32.

Schief, M., Vogt, S. & Efferson, C. (2021), 'Investigating the structure of son bias in armenia with novel measures of individual preferences', *Demography* .

Schimmelpfennig, R., Vogt, S., Ehret, S. & Efferson, C. (2021), 'Promotion of behavioural change for health in a heterogeneous population', *Bulletin of the World Health Organization* **tbd**(tbd), tbd.

Shell-Duncan, B. (2008), 'From health to human rights: Female genital cutting and the politics of intervention', *American Anthropologist* **110**(2), 225–236.

Shell-Duncan, B. & Hernland, Y. (2000), Female" Circumcision" in Africa: Dimensions of the Practice and Debates, *in* 'Female" Circumcision" in Africa: Culture, Controversy, and Change', Lynne Rienner Publishers, p. 1.

Singh, M., Acerbi, A., Caldwell, C. A., Danchin, É., Isabel, G., Molleman, L., Scott-Phillips, T., Tamariz, M., Van den Berg, P., Van Leeuwen, E. J. et al. (2021), 'Beyond social learning', *Philosophical Transactions of the Royal Society B* **376**(1828), 20200050.

Singhal, A. & Rogers, E. M. (2003), The status of entertainment-education worldwide, *in* 'Entertainment-education and social change', Routledge, pp. 25–42.

Smaldino, P. E., Janssen, M. A., Hillis, V. & Bednar, J. (2017), 'Adoption as a social marker: Innovation diffusion with outgroup aversion', *The Journal of Mathematical Sociology* **41**(1), 26–45.

Smaldino, P. E. & Jones, J. H. (2021), 'Coupled dynamics of behaviour and disease contagion among antagonistic groups', *Evolutionary Human Sciences* **3**.

Soraperra, I., Köbis, N., Efferson, C., Shalvi, S., Vogt, S. & Offerman, T. (2019), 'A market for integrity-an experiment on corruption in the education sector', *Available at SSRN 3409962* .

Sulik, J., Efferson, C. & McKay, R. (2021), 'Collectively jumping to conclusions: Social information amplifies the tendency to gather insufficient data.', *Journal of Experimental Psychology: General* .

Thaler, R. H. & Sunstein, C. R. (2021), *Nudge: The Final Edition*, Penguin.

Thomas, L. M. (2000), ''Ngaitana (I Will Circumcise Myself)': Lessons from Colonial Campaigns to Ban Excision in Meru, Kenya'.

Vivalt, E. (2015), 'Heterogeneous treatment effects in impact evaluation', *American Economic Review* **105**(5), 467–70.

Vogt, S., Efferson, C. & Fehr, E. (2017), 'The risk of female genital cutting in europe: comparing immigrant attitudes toward uncut girls with attitudes in a practicing country', *SSM-Population Health* **3**, 283–293.

Vogt, S., Zaid, N. A. M., Ahmed, H. E. F., Fehr, E. & Efferson, C. (2016), 'Changing cultural attitudes towards female genital cutting', *Nature* **538**(7626), 506–509.

Weber, R. A. (2006), 'Managing growth to achieve efficient coordination in large groups', *American Economic Review* **96**(1), 114–126.

White, C. J., Muthukrishna, M. & Norenzayan, A. (2021), 'Cultural similarity among coreligionists within and between countries', *Proceedings of the National Academy of Sciences* **118**(37).

Wilson, D. S., Hayes, S. C., Biglan, A. & Embry, D. D. (2014), 'Evolving the future: Toward a science of intentional change', *Behavioral and Brain Sciences* **37**(4), 395.

Young, H. P. (2015), 'The evolution of social norms', *Annual Review of Economics* **7**(1), 359–387.

Young, H. P. & Burke, M. A. (2001), 'Competition and custom in economic contracts: a case study of Illinois agriculture', *American Economic Review* **91**(3), 559–573.