

Policy to activate cultural change to amplify policy

Charles Efferson^{a,1}

This manuscript was compiled on April 22, 2021

Social tipping | Behavior change | Social norms | Coordination | Cultural evolution

1 If you happened to be driving down the road in Sweden at 04:50
2 on 3 September 1967, the Swedish government required you to
3 stop. You then had to move slowly from the left to the right side of
4 the road, and at 05:00 you could continue on your way. Although
5 Sweden invested heavily in preparing for this pivotal ten minutes,
6 the transition from left to right created some inevitable confusion (1).
7 Nonetheless, the transition to a new equilibrium was fast. Traffic
8 accidents and insurance claims actually declined immediately after
9 the change, presumably because of extra caution behind the wheel,
10 but they soon returned to normal (2). With a one-time government
11 initiative, Swedes tipped from driving on the left to driving on the
12 right, where they have remained ever since. The rest of us gained
13 a compelling metaphor, arguably too compelling, for how social
14 tipping can support society-wide changes in culture consistent with
15 policy goals.

16 I say “arguably too compelling” because choosing a side of the
17 road is a special kind of coordination problem maximally suited
18 to rapid change. The question is, when does the potential for
19 rapid social tipping extend to other coordination problems that are
20 similar in some ways but different in others? More broadly, can
21 we predict and even control tipping in settings that are typical pre-
22 cisely because they are more complex than choosing the left or
23 right side of the road? In a companion article, Andreoni et al. (3)
24 examine exactly these questions with a theoretical and experimen-
25 tal approach. Apart from basic scientific interest, the questions
26 are relevant across an impressive array of policy domains where
27 social norms, applied cultural evolution, and tipping appear as
28 related mechanisms for behavior change (4, 5). Example domains
29 range from equality, social justice, and health (6, 7) to resource
30 conservation (8, 9) and climate change (10, 11).

31 Choosing a side of the road is a special problem for at least
32 three reasons. Simple preferences to coordinate with people
33 nearby do not mix with other motives. Moreover, these prefer-
34 ences are the same for everyone, and they are stable through
35 time. Intuitively, from an ex ante perspective before a society has
36 chosen left or right, everyone agrees that either side is and will
37 remain just as good as the other. The one and only concern is that
38 everyone makes the same choice. Language is similar. “Der Hund”
39 and “le chien” both work fine and will continue to do so; we just

40 need to agree (12, 13). Step outside these two domains, however,
41 and many coordination problems involve a number of additional
42 complexities.

43 Andreoni et al. (3) add important and realistic complexity by
44 abandoning exactly the three characteristics that make driving
45 and language special problems. They examine a setting in which
46 individuals are randomly paired to play a game. Each player must
47 choose blue or green, and everyone faces incentives to coordinate
48 their choices with their partners. Players play, receive a payoff,
49 update their beliefs about how others play, and then pair off and
50 play again. So far, this sounds like driving, but the similarities end
51 there. Specifically, each player has a ranking over the equilibria of
52 the game, which means the player either prefers coordinating on
53 blue over coordinating on green or vice versa. Players also differ
54 from each other in terms of their rankings, and player rankings
55 change through time.

56 Andreoni et al. (3) emphasize the evolution of social norms as
57 an organizing principle. A norm is a common behavior together
58 with the widespread belief that the behavior is and should remain
59 common. A norm helps people pick a specific behavior when
60 everyone values choosing the same behavior, which is a problem
61 with multiple solutions. This pressure to behave like others is
62 also why norm evolution can exhibit tipping. If a status quo norm
63 becomes unstable, the pressure to conform can lead the population
64 to coalesce quickly around a new norm.

65 To develop a modeling framework for how norms evolve, An-
66 dreoni et al. (3) decompose preferences into three parts. First,
67 each player faces a basic material incentive that favors either coor-
68 dinating on blue over coordinating on green or vice versa. Second,
69 each player faces material incentives that are relevant when two
70 players choose different options. Specifically, in addition to the
71 opportunity costs of miscoordination, each player in a miscoordina-
72 ting pair pays a cost that increases as the player’s choice becomes
73 more unusual. We can interpret this cost as punishment. These
74 first two components of the incentive structure are material in the
75 sense that they were monetized in Andreoni et al.’s experiment.
76 More broadly, they represent the public features of decision mak-
77 ing that would be readily available for policy intervention. A policy
78 maker, for example, can subsidize the behavior she prefers, tax

^aFaculty of Business and Economics, University of Lausanne, 1015 Lausanne, Switzerland.
CE wrote the paper.

¹To whom correspondence should be addressed. E-mail: charles.efferson@unil.ch.
The author declares no competing interests.

79 the behaviors she does not prefer, and punish deviants. The third
80 component of preferences is an idiosyncratic psychological quantity
81 that appears in the predictive model of Andreoni et al. but was
82 not monetized in their experiment. Among other interpretations,
83 variation in this quantity represents the fact that some people are
84 more open to new experiences than others, a form of ordinary heterogeneity
85 that can affect the spread of innovations in a population
86 (14, 15).

87 With all three parts of the theoretical incentive structure in place,
88 each individual has an indifference point. If the proportion of individuals
89 recently choosing green is at least as large as this indifference point,
90 the individual in question chooses green by assumption. The population
91 consists of a distribution of indifference points. This distribution changes
92 through time and in turn influences how behavior and associated norms evolve.
93

94 In Andreoni et al.'s (3) experimental sessions, material incentives
95 initially favored coordinating on blue over coordinating on green, and
96 groups immediately adopted a blue norm as a result. With a blue norm
97 in place, material incentives began to change. At a given point in time,
98 for any individual whose material incentives favored blue over green,
99 these incentives would switch the ranking with probability 0.1. As these
100 new incentives trickled into the population, the distribution of indifference
101 points should have become increasingly favorable for green.
102

103 Fig. 1 shows a stylized simulation in which this steady trickle
104 leads to tipping. In $t = 1$, no one faces material incentives that
105 favor coordinating on green. All three parts of the incentive structure
106 combine to create a distribution of indifference points that is not
107 favorable for green, and no one chooses green. Material incentives
108 then begin to change, and the distribution of indifference points
109 drifts downward. For a while, behavior change lags behind as everyone
110 continues to conform to the status quo blue norm. At $t = 6$,
111 changes in behavior start to race ahead of the changes in material
112 incentives, and by $t = 9$ the entire population has switched to
113 choosing green. This is social tipping. Coordination and conformity
114 oppose the behavioral effects of changing incentives at first, but
115 then suddenly a new regime appears in which they amplify these
116 effects.

117 This kind of tipping, however, may not occur, and altogether
118 Andreoni et al. (3) implemented nine experimental treatments to
119 examine a variety of behavioral mechanisms. Four treatments operated
120 directly via material incentives. Andreoni et al. manipulated the
121 material incentives related to coordinating, and they manipulated the
122 material punishment associated with miscoordinating. Their model
123 does an outstanding job of predicting observed tipping (3, Fig. 4).
124 In one especially revealing treatment, Andreoni et al. allowed the
125 participants themselves to set the punishment costs of miscoordinating.
126 This is like a situation in which a policy maker uses a combination
127 of taxes and subsidies to promote a specific behavior, but the
128 punishment of norm violations is an informal affair that citizens
129 handle themselves. In this treatment, participants consistently set
130 punishment costs too high. Doing so saved them the short-run costs
131 of miscoordinating while transitioning to a new norm, but using
132 punishment to block transitions brought substantial opportunity costs
133 in the long-run.

134 Four additional treatments manipulated the information and
135 expectations participants had about the changes occurring in their
136 groups. In one treatment, participants received immediate feedback
137 about what others were choosing, an approach designed to mimic
138 the speed of modern communications. One can imagine that readily
139 available information would have facilitated tipping, but

140 it did not. Instead, it seems to have made the early prevalence
141 of blue salient, and this treatment had no effect on tipping. In
142 another treatment, Andreoni et al. (3) cut the size of experimental
143 groups from 20 to 10, which increased the relative influence of
144 each decision maker. This significantly increased tipping. Surprisingly,
145 however, when transitions to a green norm occurred, they were long
146 drawn-out affairs with a lot of miscoordination along the way.
147 Average earnings were especially low as a result. This result shows
148 that transitions to a new, socially beneficial equilibrium can actually
149 be socially harmful depending on how long the transition takes.
150

151 In the "Public awareness" and "Preference poll" treatments, Andreoni
152 et al. (3) introduced two mechanisms designed to make private
153 information public (6). Under public awareness, the experimenters
154 gave participants a running log of the kinds of changes in material
155 incentives taking place within the group. The preference poll
156 polled group members about their preferred norm after several
157 periods of play and immediately made the poll results public. Both
158 of these treatments revealed information about participants that
159 would have otherwise remained private, and even trivial revelations
160 of this sort can strongly affect cultural evolution (16). The result
161 in both treatments was a significant increase in tipping to the socially
162 beneficial norm.

163 Finally, Andreoni et al. implemented a treatment that rewarded
164 those who first attempted to instigate norm change, but only when
165 these attempts were successful. This extra reward for agents of
166 change seems to have motivated individuals predisposed to change
167 anyway, but it also ignored people with a status quo bias. As the
168 authors point out, tipping requires behavior change among both
169 types, both those who are ready to lead the way to a new norm
170 and those who are not. The results across groups in this treatment
171 were highly unpredictable, with half of the groups tipping to
172 green and half sticking with blue. Altogether, Andreoni et al.
173 used a convincing policy-inspired mix of treatments to detail
174 several behavioral subtleties related to tipping. At the same time,
175 their study highlights how much we still need to learn about the
176 various scenarios in which a policy maker might want to activate
177 endogenous cultural change.

178 One important scenario is when the population is sub-divided
179 into groups that have distinct social identities tied to the norms
180 and behaviors in question. For example, imagine a situation in
181 which some people have tied their social identities to their shared
182 decision to wear face masks in a pandemic, while others have
183 based their social identities on rejecting masks (17). In cases like
184 this, the distribution of indifference points will look quite different
185 from those assumed in Andreoni et al. (see also Fig. 1). The
186 distribution will tend to be strongly bimodal, with one mode for
187 the group that likes one behavior and another mode for the group
188 that likes the other behavior. Tipping points may not exist in
189 situations like this, and the most challenging situation of all is
190 when the groups have social identities that are not only distinct,
191 but oppositional (18). Oppositional identities would mean, for
192 example, that the group rejecting masks values this stance precisely
193 because of the difference it creates with respect to the group
194 wearing masks (19). If preferences take this form, the policy
195 maker who sparks a commitment to her preferred norm in one
196 group likely entrenches and adds value to a different norm in
197 the other group (18). The increasingly sectarian nature of U.S.
198 politics (20) suggests that dynamics of this sort could be
199 common in the future.

200 A second issue involves the options available to the policy
maker. Andreoni et al. implement several treatments that reflect the

201 kinds of choices a policy maker might consider to try and provoke
202 tipping. Their treatments represent policy initiatives that subsidize
203 the desired behavior, punish the undesired behavior, influence the
204 information people have, and reward those who instigate change.
205 These are all important possibilities. In addition, a policy maker
206 might also want to constrain an intervention to a specific segment of
207 the population. Indeed, much of the policy appeal of tipping follows
208 from the idea that an intervention touches only some people. When
209 these people change their behavior, however, the effect spills over
210 to generate additional change among those never exposed to the
211 intervention. If a policy maker wants a constrained approach of this
212 sort, she must decide whom to target. Some strategies prioritize
213 the effects among those directly exposed to the intervention while
214 minimizing the changes that occur among those not exposed.
215 Other strategies do the opposite, with a range of trade-offs in
216 between the extremes (18).

217 Tipping has a theatrical quality, with rapid changes that some-
218 how seem both surprising and obvious after they have occurred.
219 Tipping is also tempting as a policy tool because it implies the
220 policy maker can recruit social interactions within a population to
221 point cultural evolution in a specific direction. Empirically, however,
222 people are strikingly heterogeneous in terms of how they learn
223 from and react to the choices of others (21, 22). This suggests that
224 tipping and other cultural evolutionary processes can easily involve
225 a daunting level of complexity. Andreoni et al. (3) have provided
226 an important study of ways to examine and manage some of this
227 complexity.

- 228 1 E Flock, Dagen H: The day Sweden switched sides of the road (photo). *Wash. Post* **N/A**
229 (2012).
- 230 2 D Bierend, Throwback Thursday: Hilarity ensues as Sweden starts driving on the right side of
231 the road. *Wired* **N/A** (2014).
- 232 3 J Andreoni, N Nikiforakis, S Siegenthaler, Predicting social tipping and norm change in con-
233 trolled experiments. *Proc. Natl. Acad. Sci.* ?, ? (2021).
- 234 4 HP Young, The evolution of social norms. *Annu. Rev. Econ.* **7**, 359–387 (2015).
- 235 5 M Muthukrishna, Cultural evolutionary public policy. *Nat. Hum. Behav.* **4**, 12–13 (2020).
- 236 6 C Bicchieri, *Norms in the wild: How to diagnose, measure, and change social norms.* (Oxford
237 University Press), (2016).
- 238 7 JP Platteau, G Camilotti, E Auriol, Eradicating women-hurting customs. *Towards gender equity*
239 *development* **319** (2018).
- 240 8 JC Castilla-Rho, R Rojas, MS Andersen, C Holley, G Mariethoz, Social tipping points in global
241 groundwater management. *Nat. Hum. Behav.* **1**, 640–649 (2017).
- 242 9 H Travers, J Walsh, S Vogt, T Clements, E Milner-Gulland, Delivering behavioural change at
243 scale: What conservation can learn from other fields. *Biol. Conserv.* **257**, 109092 (2021).
- 244 10 K Nyborg, et al., Social norms as solutions. *Science* **354**, 42–43 (2016).
- 245 11 MJ Dávila-Fernández, S Sordi, Attitudes towards climate policies in a macrodynamic model of
246 the economy. *Ecol. Econ.* **169**, 106319 (2020).
- 247 12 D Centola, A Baronchelli, The spontaneous emergence of conventions: An experimental study
248 of cultural evolution. *Proc. Natl. Acad. Sci.* **112**, 1989–1994 (2015).
- 249 13 D Centola, J Becker, D Brackbill, A Baronchelli, Experimental evidence for tipping points in
250 social convention. *Science* **360**, 1116–1119 (2018).
- 251 14 J Henrich, Cultural transmission and the diffusion of innovations: Adoption dynamics indicate
252 that biased cultural transmission is the predominate force in behavioral change. *Am. Anthropol.*
253 **103**, 992–1013 (2001).
- 254 15 EM Rogers, *Diffusion of Innovations.* (Simon and Schuster), (2010).
- 255 16 JK Goeree, TR Palfrey, BW Rogers, RD McKelvey, Self-correcting information cascades. *The*
256 *Rev. Econ. Stud.* **74**, 733–762 (2007).
- 257 17 C Moya, , et al., Dynamics of behavior change in the COVID world. *Am. J. Hum. Biol.* **e23485**
258 (2020).
- 259 18 C Efferson, S Vogt, E Fehr, The promise and the peril of using social influence to reverse
260 harmful traditions. *Nat. Hum. Behav.* **4**, 55–68 (2020).
- 261 19 SM Utych, Messaging mask wearing during the COVID-19 crisis: Ideological differences. *J.*
262 *Exp. Polit. Sci.* **N/A**, 1–11 (2020).
- 263 20 EJ Finkel, et al., Political sectarianism in America. *Science* **370**, 533–536 (2020).
- 264 21 A Mesoudi, L Chang, SR Dall, A Thornton, The evolution of individual and cultural variation in
265 social learning. *Trends Ecol. & Evol.* **31**, 215–225 (2016).
- 266 22 RL Kendal, et al., Social learning strategies: Bridge-building between fields. *Trends Cogn. Sci.*
267 **22**, 651–665 (2018).

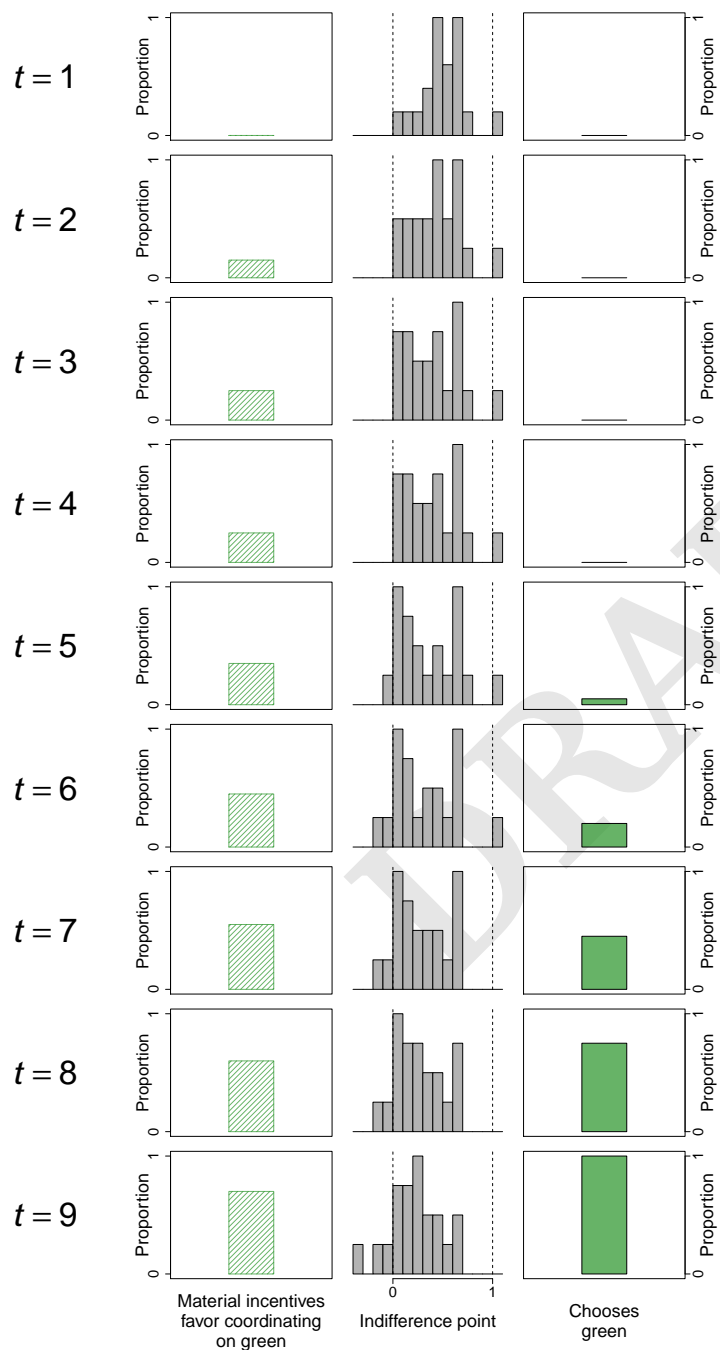


Fig. 1. An example of tipping based on the framework used in Andreoni et al. (3). In $t = 1$, everyone faces material incentives that favor coordinating on blue over coordinating on green (Column 1). The distribution of indifference points is relatively unfavorable for green as a result (Column 2), and everyone chooses blue (Column 3). As time passes, individuals experience changing material incentives. The distribution of thresholds drifts steadily downward, in favor of green, but for a while (e.g. $t \leq 5$) this generates little change in behavior. At some point (e.g. $t \geq 6$), behavior change suddenly accelerates, and the population transitions rapidly to a new norm. Broadly speaking, Andreoni et al. (3) examine when the rapid change in behavior does or does not follow the change in incentives.