SciencesPo
DEPARTMENT OF ECONOMICS

Discussion Paper

MASKS, CAMERAS, AND SOCIAL PRESSURE

Itzhak Rasooly and Roberto Rozzi

SCIENCES PO ECONOMICS DISCUSSION PAPER

No. 2022-12

Masks, cameras and social pressure*

Itzhak Rasooly[†] Roberto Rozzi[‡]

December 9, 2022

Abstract

In contrast to classical social norm experiments, we conduct experiments that *semi-continuously* randomise the share of individuals who are taking a particular action in a given environment. Using our experimental results, we are able to estimate the distributions of individual tipping points across our settings. We find that tipping points are very heterogenous, and that a substantial share choose to do the action (or not) regardless of what others are doing. We also show that, once embedded in dynamic models, our estimates predict that individuals will end up doing very different things despite engaging in copying-like behaviour.

Keywords: social norms, field experiment, dynamic models
JEL Codes: D90, C93, C73

*For useful comments and discussions, we would like to thank Adam Brzezinski, Andrea Gallice, Eduard Krkoska, Ennio Bilancini, Jasmine Theilgaard, Jean-Paul Carvalho, Johannes Abeler, Junnan He, Karl Schlag, Loukas Balafoutas, Luca Corazzini, Marco LiCalzi, Marie Claire Villeval, Michele Bernasconi, Miguel Ballester, Paolo Pellizzari, Peyton Young, Pietro Dindo, Séverine Toussaert, Valeria Maggian, Valerio Capraro, Vatsal Khandelwal and Vivek Roy-Chowdhury. We would also like to thank audiences at ASFEE Lyon, the ESA meeting in Bologna, the University of Oxford, the University of Turin, and IMT Lucca. For excellent research assistance, we are grateful to Amélie Marescaux, Antonia Ledda, Aniket Chakravorty, Declan Murray, Doaa Shabbir, Ethan Dodd, Eunice Poon, Gwen Williams, Ivan Toth-Rohonyi, Jacob Hill, Jasvin Khurana, Jordan Edwards-Zinger, Juliet Dyrud, Kirsten Fletcher, Leah Fahy, Lucy Owen, Lunchen Song, Maisie Howard, Mary Chen, Michelle Cho, Natalie Vriend, Neily Raymond, Nina Skrzypczak, Peter Brookes, Ria McDonald, Rose Poyser, Sahani de Silva, Thomas Dingwall, Tim Green, Wade Kamphuis and Wyatt Radzin. Finally, we would like to thank Exeter College, the George Webb Medley Fund, and Ca' Foscari University of Venice for generous financial support.

†Sciences Po & Paris School of Economics.
‡University of Padova & Ca' Foscari University of Venice.

1 Introduction

There is a large literature demonstrating the power of peer effects and descriptive social norms across a range of domains. For instance, studies have found that we look to others when deciding whether to evade our taxes (Bott et al., 2020), donate to charity (Agerström et al., 2016) and even whether to vote (Gerber and Rogers, 2009). Of course, these examples are somewhat arbitrary: it is hard to think of even one activity that is not somehow shaped by our expectations about the behaviour of others.

Despite the obvious importance of social norms, however, current studies only provide limited evidence regarding the exact relationship between our beliefs about the share of people who do an activity and our own inclination to do that activity. To take a fairly typical example, consider Frey and Meier (2004)'s study of the impact of informing individuals that 64% as opposed to 46% of their peers donate to charity. While their experiment reveals that higher beliefs about the prevalence of charitable giving can lead to higher donation rates, it reveals little about the exact shape of this relationship; and more generally how this function looks over the full possible range of beliefs (0% to 100%).

There are at least three reasons why one might care about how actions depend on exact beliefs about prevalence (we call this relationship the f function). First, there is a clear policy motivation: the shape of this function reveals the returns to altering perceptions about prevalence (e.g. by disclosing information). Second, estimating the shape of this function allows us to test economic theories since certain economic models, e.g. those in evolutionary game theory, make distinctive predictions about the observed functional form. Third, the shape of this function turns out to be absolutely crucial for understanding long-run equilibria in dynamic models.

In this paper, we begin by elaborating on this third point by demonstrating theoretically how long-run equilibria in a plausible dynamic model depend on the shape of this function. We show that curvature is critical. If the function is first convex and then concave, then our dynamic system converges to an extreme equilibrium (as in Kreindler and Young 2013). On the other hand, if the function is first concave and then convex, we obtain convergence to an interior equilibrium; meaning that individuals end up doing very different things even though each of them is engaging in copying-like behaviour. As a result, understanding the curvature of this function is crucial for understanding long-run outcomes. In addition, we show that the intercepts of this function play a key role in pinning down its fixed points, thereby also shaping long-run equilibria.

This motivates our first experiment which aims to estimate the shape of this function in a particular setting, namely face mask usage. The basic idea of the experiment was straightforward. Subjects entered a room (one at a time) thinking that they were there solely to answer a decision problem involving lotteries. Unbeknownst to them, the number of the four experimenters in the room wearing a face mask had been randomised (leading to treatments in which 0/4, 1/4, 2/4, 3/4, or 4/4 experimenters were wearing a mask). We then observed whether each subject themselves chose to wear a face mask.

The experiment took place in Oxford over the course of nine days in February/March 2022. In total, we conducted fourteen three-hour sessions across twelve different colleges; and repeated our experimental protocol 646 times (each time with a different subject). Importantly, the experiment took place at a time in which face masks were no longer required by law or university rules, but still remained not abnormal. As a result, this was an ideal setting for capturing the implications of social pressure.

Our first experiment yielded four main results. First, according to our point estimates, the function is strictly increasing. That is, the greater the number of experimenters who were wearing a mask, the more likely were subjects to wear a mask. Reassuringly, this increasing relationship is evident across all of the specifications we estimate, including those that include and omit demographic controls and college fixed effects.

Second, we observe that many individuals defy social pressure. For example, 20% of the subjects chose to wear a mask even when none of the experimenters were wearing one (the 0/4 treatment); and 51% of the subjects did not wear the mask when all the other experimenters were wearing it (the 4/4 treatment). Similar results can be obtained by looking at changes, i.e. whether individuals chose to put on or take off a face mask during the experiment. For instance, out of the 106 subjects in the 4/4 treatment who entered the room without wearing a mask, only 39 chose to put on a mask during the experiment — a fact that illustrates the limits of social pressure in our setting.

Third, according to our point estimates, the largest jump in mask wearing arises between the 3/4 and 4/4 treatments. For instance, while increasing the number of mask wearers in the room from 1 to 2 experimenters raises the probability that a subject will wear a mask by around 4 percentage points, increasing the number of mask wearers from 3 to 4 raises the probability that a subject will wear a mask by a full 12 percentage points. This finding is consistent with an 'everybody effect' where social pressure becomes especially acute if everybody in the relevant environment chooses to do a particular activity.

Fourth, and perhaps most importantly, our estimated function has an interior fixed point, which is close to 23%. When embedded in our dynamic models, our results therefore suggest that, in the long-run, around 23% will choose to wear a face mask. As a result, calibrating our models using our estimates predicts convergence to an interior equilibrium, despite the existence of copying-like behaviour.

In order to assess the robustness of our findings, we conducted an analogous experiment in a very different context: camera use in online calls. The idea of this experiment was also straightforward. Subjects joined a Zoom call (one at a time) knowing only that they were attending in order to participate in an economics experiment. Unbeknownst to them, the number of the four experimenters on the call with their laptop camera on had been randomised (leading again to five treatments, corresponding to 0/4, 1/4, 2/4, 3/4, and 4/4 experimenters with their camera on). We then observed whether each subject themselves chose to use their video camera. In total, we repeated this process 1,114 times, leading to a sample size that was almost twice as large as that obtained in our first experiment.

Conducting this experiment led to similar, although not identical, results. We again find evidence of an everywhere increasing f function, i.e. that the share who use their camera is everywhere increasing in the number of experimenters who use their camera. We also again find high levels of non-compliance, with many participants choosing to use their cameras (or not) regardless of how many others are doing the same. Most importantly, once we use our estimates to calibrate our dynamic models, we again obtain convergence to an interior equilibrium, now with around 37% using a camera. Despite these similarities, the estimated f function in this context is not precisely the same as that estimated in the mask setting; and appears to be substantially more linear.

Finally, we discuss which models could give rise to our experimental findings; and could explain both the commonalities and differences between them. We observe that, assuming that all individuals have tipping point preferences, our f function can be interpreted as the cumulative distribution of individual tipping points. Viewed in this way, our experiments can be interpreted as an attempt to estimate the distribution of individual tipping points using randomisation. In both experiments, we find that individual tipping points are very heterogeneous, in contrast to canonical models in evolutionary game theory (e.g. Young 1993). We also provide a simple model to explain where these tipping points come from. In our model, tipping points are the result of the interaction of intrinsic preferences to take the action along with (potentially non-linear) social pressure effects.

Our study contributes to a number of literatures across economics and related disciplines. First, our study contributes to the broader literature on the importance of peer effects and descriptive social norms. The current literature consists of a series of generally binary experiments across a variety of domains (see Cialdini 2007, Mascagni 2018 and Farrow et al. 2017 for reviews).[1] In contrast, our study is the first to semi-continuously randomise the share taking an action in subjects' immediate environment; and the first to do so in any setting (not just the settings of face masks and video calls).[2]

Second, our study contributes to the literature on tipping points and long-run dynamics. Especially relevant references include papers like Young (1993), Kandori et al. (1993), Jackson and Yariv (2007), Young (2009) and Kreindler and Young (2013) in the economics literature; as well as the sociology literature following Granovetter (1978) (see Dodds and Watts 2011 for an overview). Our study can be viewed as a first attempt to experimentally estimate the shape of the 'aggregate best response function' (or equivalently, tipping point distribution) that is crucial for driving the results of such models.

Third, and more narrowly, we contribute to the literature on the social determinants of face mask wearing. The existing papers in this literature rely either on vignette-based experiments and surveys (Bokemper et al., 2021; Barceló and Sheen, 2020; Rudert and Janke, 2021; Goldberg et al., 2020; Barile et al., 2021) or instead observational data (Freidin et al., 2022; Woodcock and Schultz, 2021). We contribute to this literature by conducting the first ever randomised field experiment on the social determinants of face mask use.[3]

Fourth, we contribute to the literature on the social determinants of video camera use. Existing papers in this literature are again based on surveys: see, for example, Castelli and Sarvary (2021), Gherheș et al. (2021), Sederevičiūtė-Pačiauskienė et al. (2022) and Bedenlier et al. (2021). Our study is the first to examine this topic through use of a randomised field experiment.

[1] Studies which provide some information about f functions in various contexts include: Cialdini et al. (1990); Cason and Mui (1998); Ichino and Maggi (2000); Borsari and Carey (2003); Heldt (2005); Fortin et al. (2007); Goldstein et al. (2008); Martin and Randal (2008); Krupka and Weber (2009); Gerber and Rogers (2009); Allcott (2011); Ferraro and Price (2013); Ayres et al. (2013); Costa and Kahn (2013); Bursztyn et al. (2014); Damm and Dustmann (2014); Smith et al. (2015); Thöni and Gächter (2015); Efferson et al. (2015); Lefebvre et al. (2015); Allcott and Kessler (2019); Novak (2020); Linek and Traxler (2021).

[2] Our study also connects with the conformity literature following Asch (1951). In contrast to this literature, our study concerns individuals' actions (e.g. whether to wear a mask) as opposed to their cognitive judgements. Perhaps more importantly, our study also uses semi-continuous randomisation, in contrast to experiments in the Asch paradigm (see Bond and Smith 1996 for an overview).

[3] Our use of a randomised field experiment allows us to side-step some of the issues that afflict previous studies of the social determinants of face mask wearing. For example, attempts to study this problem using hypothetical questions (as in Bokemper et al. 2021) suffer from the issue that individuals may not know what they would do in a hypothetical situation — an especially pressing concern since imitative behaviour may well rest on unconscious cognition. Meanwhile, attempts to study this problem using observational data (as in Woodcock and Schultz 2021) can suffer from both omitted variable bias and reverse causality issues (see Manski 1993 for an influential exposition of this latter point). Our randomised experiment avoids both of these issues.

The remainder of this article is structured as follows. Section 2 motivates our experiments with a theoretical discussion of the long-run implications of various f functions. Section 3 outlines the design of our face mask experiment and the associated results. Section 4 presents the design and results for our experiment on video cameras. Section 5 uses our results to calculate the distribution of individual tipping points across our contexts and discusses what could give rise to these distributions. Finally, Section 6 concludes with a discussion of future research suggested by our experiments.

2 Dynamics

To motivate our experiments, we begin by discussing how the relationship between beliefs about prevalence and the actual prevalence of an activity pin down long-run equilibria in a plausible dynamic model. Time is discrete, indexed by $t = 0, 1, 2, \ldots$. Let $\hat{s}_t \in [0, 1]$ denote the belief (assumed to be common) about the share doing an activity at time $t \in \mathbb{N}$. The belief $\hat{s}_t \in [0, 1]$ generates the actual share $s_t \in [0, 1]$ via the function $f \colon [0, 1] \to [0, 1]$. That is, $s_t = f(\hat{s}_t)$ for all t. If we assume that $\hat{s}_t = s_{t-1}$, then we obtain the relation $s_t = f(s_{t-1})$: a dynamic process whose outcomes we can study. We will write f^t to denote the tth iterate of f. For example, $f^3(s_0) = f(f(f(s_0)))$.

In canonical evolutionary game theory models (Young, 1993; Kandori et al., 1993; Young, 2009), there exists some 'tipping point' at which all individuals will switch from not taking the action to taking it. This gives rise to a convex and then concave shaped f function (see Figure 1(a) for a smooth version). While this is a plausible model in many settings, it is not the only possible model. For example, one might instead think that individuals are quantitatively quite insensitive, so treat shares like 40% and 45% as 'the same'. In contrast, however, one might think that there is an important qualitative difference between nobody and a minority taking an action, which leads to an f function which is steep near zero; and one might similarly assume an f function that is steep near 1. The resulting f function — which is reminiscent of the probability weighting function proposed in Kahneman and Tversky (1979) — is concave and then convex, and is plotted in Figure 1(b).

Proposition 1. Suppose that f is continuous, strictly increasing, and has three fixed points at $s = 0$, $s = \hat{s} \in (0, 1)$ and $s = 1$. Then

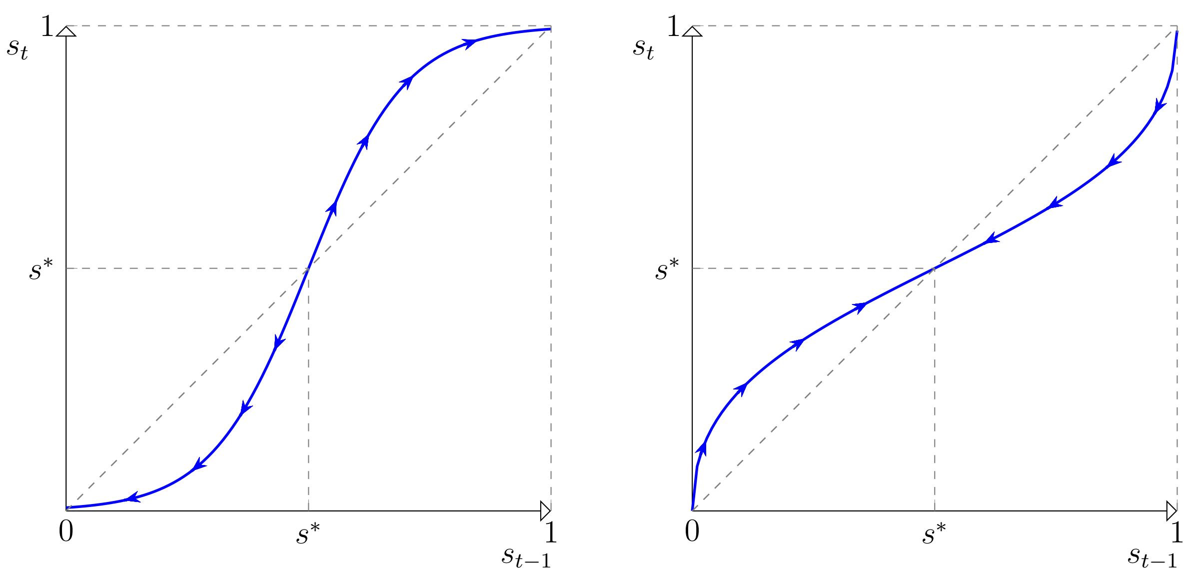
- If f is convex on $[0, \hat{s}]$ and concave on $[\hat{s}, 1]$, then $\lim_{t\to\infty} s_t \in \{0, 1\}$ provided that $s_0 \neq \hat{s}$.
- If f is concave on $[0, \hat{s}]$ and convex on $[\hat{s}, 1]$, then $\lim_{t\to\infty} s_t = \hat{s}$ provided that

$s_0 \notin \{0, 1\}$.

Proposition 1 shows how long-run equilibria crucially depend on the shape of the f function. If one assumes a convex and then concave f function, as in Kreindler and Young (2013), then one generically obtains convergence to an extreme equilibrium in which either nobody or everybody does the action.[4] We do not outline the dynamics in any detail since they are already familiar, but they are displayed graphically in Figure 1(a). On the other hand, Proposition 1 also states that if the f function is concave and then convex, then one obtains convergence to an *interior* equilibrium as displayed in Figure 1(b). This illustrates how differently shaped f functions can generate very different equilibria.

Figure 1: Two possible f functions

(a) Convex then concave (b) Concave then convex



Notes. This figure describes the evolution of s_t given two different f functions. When $f(s_{t-1}) > s_{t-1}$, the share doing the activity rises. When $f(s_{t-1}) < s_{t-1}$, the share doing the activity falls.

While Proposition 1 assumes that $f(0) = 0$ and $f(1) = 1$, this need not be the case. Instead, one might think that some individuals always do the action (leading to $f(0) > 0$) and that some others never do the action (leading to $f(1) < 1$). We now show that understanding the intercepts of the f function is also crucial for understanding long-run equilibria.

Proposition 2. Suppose that f is increasing. Then if s^* is the limit of $f^t(s_0)$ as $t \to \infty$, $s^* \in [f(0), f(1)]$.

[4] One can then use stochastic stability arguments to identify which of these equilibria is more likely to emerge: see, for example, Young (1993); Kandori et al. (1993).

Proposition 2 states that, assuming the f function is increasing, then the long-run share of individuals who do the activity is bounded by $f(0)$ and $f(1)$. Intuitively, this is because the fixed points of the f function must be bounded in this way; and any limit of the sequence $\{s_t\}_{t=0}^{\infty}$ must be a fixed point of f. As a result, estimating $f(0)$ and $f(1)$ can provide valuable information about long-run equilibria.

In our experimental settings, behaviour is largely pinned down by beliefs about the share doing the action in the individual's immediate environment. Our results apply immediately to such cases, barring some discreteness issues, if one defines this environment as the relevant population. Alternatively, one can consider a network model in which individuals interact locally in small but interconnected communities (see Appendix B). As one might expect, this model yields very similar results.

In this section, our primary goal is not to insist on a particular dynamic model. Indeed, we believe that a large number of reasonable models are possible; and that one can make substantial variations on the assumptions made above. Rather, the main goal is to emphasise how, in any reasonable model, the shape of the f function is going to be a crucial determinant of long-run outcomes. This motivates our experimental investigation of f functions in the next two sections.

3 Masks

3.1 Experimental design

We now describe our first experiment aimed at estimating the shape of the f function in a particular context. The basic idea of the experiment was straightforward. Subjects entered a room thinking that they were there solely to answer a decision problem involving lotteries. Unbeknownst to them, the number of experimenters in the room wearing a face mask had been randomized. We then observed whether each subject themselves chose to wear a face mask (and how this varied with the number of experimenters wearing a mask in their immediate environment).[5]

This first experiment took place in Oxford in late February and early March of 2022. At this

[5] The experiment received approval from the University of Oxford's Departmental Research Ethics Committee (ECONCIA21-22-50). In line with the recommendations of the committee, we told subjects in advance that taking part in the experiment might involve interacting with unmasked individuals (which was common at the University of Oxford at the time). We also took reasonable social distancing precautions, including making sure that the experimental settings were well ventilated. We should also emphasise that, although we did not reveal the main purpose of our experiment to participants (as is not unusual in social science experiments), we did not explicitly deceive participants at any stage.

time, masks were not required by either law or university rules – however, they were also not unusual. This gave us an ideal setting in which to study the effects of social pressure. In total, we conducted 14 three-hour sessions in 12 different colleges over 7 days (with the help of 16 research assistants, some of whom participated in multiple sessions). On average, around 46 participants attended each session; which led to a total sample size of 646 experimental subjects (see Table C1 for the distribution of subjects across treatment groups).[6]

The structure of the experiment was as follows:

1. Subjects were asked to arrive at a room within a particular time slot.
2. Before each subject entered the room, the number of the four experimenters in the room who were wearing a mask (and the allocation of masks to experimenters) had been randomised. Thus, there were five treatment groups, corresponding to: 0/4 masks, 1/4 masks, 2/4 masks, 3/4 masks, 4/4 masks. We denote these treatments by T0, T1, etc.
3. Once a subject entered, they were asked to sit at a table in a way that gave them a clear view of the four experimenters. On the table were a box of masks as well as a bottle of hand sanitiser (such a set-up was common within the University of Oxford at the time). As a result, any subject who wished to wear a mask was able to do so.
4. Once the subject had sat down, each of the four experimenters introduced themselves by stating their name and subject of study. The purpose of this was to further ensure that each subject fully processed the number of experimenters who were wearing a mask.
5. The subject was asked their name, age, college and subject of study; and then given a decision problem involving lotteries.
6. We then asked the subject to leave the room, and repeated the process for the next subject (see Appendix D for a more detailed description of the experimental protocol which includes the decision problem).

We recorded whether each subject was wearing a mask when they entered the room (this variable is labelled 'pre' in the tables). Naturally, we also recorded whether they chose to wear a mask after interacting with the experimenters. Finally, we recorded their choice in the lottery problem; as well as whether they asked if they ought to wear a mask (in such cases, each was told 'it's up to you' by the data recorder).

[6] The experiment was pre-registered here: https://www.socialscienceregistry.org/trials/9013.

Based on post-experimental conversations, it seemed that most subjects believed that our goal was to measure risk aversion. Importantly, none of the subjects appeared to suspect that the experiment had anything to do with face masks; and there was nothing in the experimental design that could have revealed this.[7] This is reassuring since subjects might have acted in unnatural and unrepresentative ways if they had known that they were taking part in a face mask experiment.

Once all experimental sessions had been completed, we debriefed all subjects on the underlying purpose of the experiment. During the debriefing, subjects were given the opportunity to take part in an online survey. In the survey, subjects were asked to imagine that they entered a room and saw 4 people sitting around a table. They were then asked if they would wear a mask if none of the 4 people were wearing a face mask, if 1 of the 4 people were wearing a face mask, and so forth. Finally, they were asked to give an explanation for their answers, as well as whether they had contracted COVID-19 at any point during the pandemic. The purpose of the follow-up survey was to obtain some suggestive evidence on mechanisms, as well as some data on individual level f functions (see Section 5 for discussion).

3.2 Results

We now turn to our main results, beginning with a brief description of our sample. As shown by Table C2, our average participant was around 21 years old; and approximately half of our sample was male. Participants were fairly evenly distributed across subject divisions, with social sciences students being most represented (33% of the sample). Turning to Table 1, we see that genders, subjects and ages were reasonably balanced across our five treatment groups. However, we do observe some imbalance in the share of participants who entered the room wearing a mask: for example, the share is 27% in treatment T2 but only 14% in T0. Given that this variable turns out to be highly predictive for our outcome (whether participants continued to wear a mask), we control for it in our main specification.

Our regressions take the form

$$y_i = \beta_0 + \sum_{i=1}^{4} \beta_i T_i + \gamma x_i + u_i, \quad (1)$$

where y_i denotes whether an individual chose to wear a mask, the T_i are dummy variables indicating treatment assignment, and x_i is a vector of covariates (including whether they

[7] We also required all research assistants to sign an agreement specifying that they would keep the main purpose of the experiment confidential throughout its duration.

Table 1: Balance table (experiment 1)

Variable	T0	T1	T2	T3	T4	p-value
Age	21.0	21.3	20.1	20.6	20.8	.143
	[.361]	[.539]	[.165]	[.219]	[.268]	
Pre	.142	.157	.266	.242	.203	.060
	[.031]	[.032]	[.039]	[.039]	[.035]	
Male	.535	.522	.461	.548	.421	.189
	[.044]	[.043]	[.044]	[.045]	[.043]	
Humanities	.323	.246	.250	.347	.256	.237
	[.042]	[.037]	[.038]	[.043]	[.038]	
Social	.268	.403	.336	.298	.353	.177
	[.039]	[.043]	[.042]	[.041]	[.042]	
MPLS	.213	.209	.305	.242	.233	.380
	[.036]	[.035]	[.041]	[.039]	[.037]	
Medical	.181	.104	.102	.105	.143	.235
	[.034]	[.027]	[.027]	[.028]	[.030]	

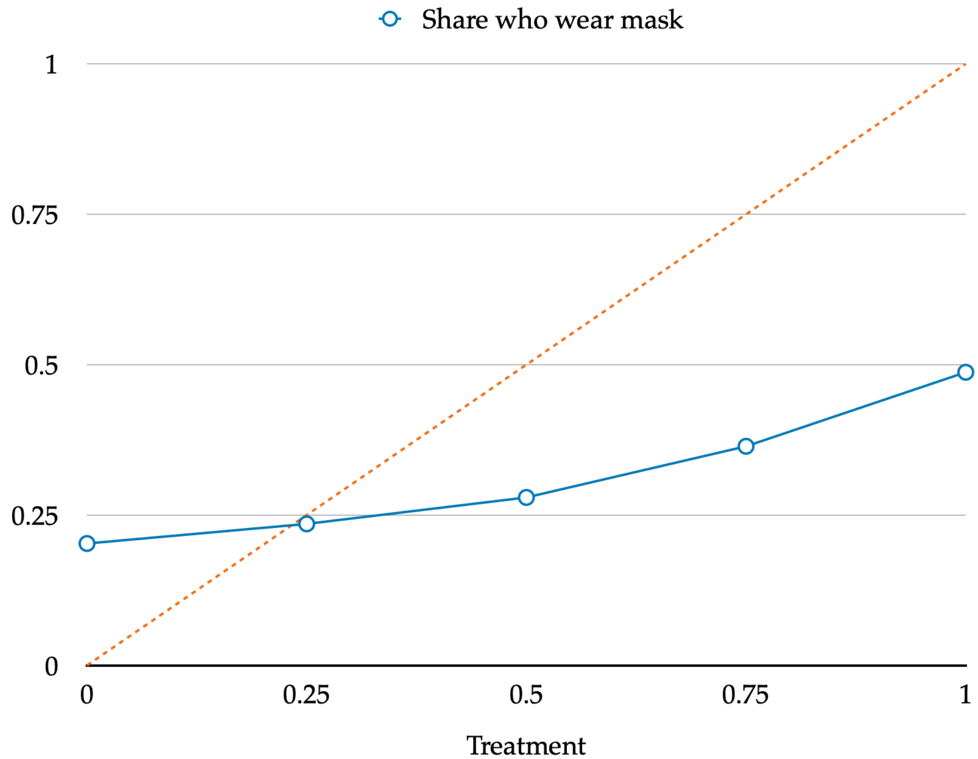
Notes. This table shows the average value of various variables across the five treatments. The variables are age, whether the subject entered wearing a mask ('pre'), gender, division of study (Humanities; Social Sciences; Mathematical, Physical & Life Sciences; Medical Sciences). The final column reports the p-value obtained from regressing the relevant variable on all treatment dummies and testing the hypothesis that the coefficients on all treatment dummies are equal to zero.

entered the room wearing a mask). In our main specification, we control for participant age, gender, and whether they entered the room wearing a mask (the 'pre' variable). However, we also report uncontrolled regressions, as well as regressions that use the full set of controls that are available (including session and college fixed effects).

Figure 2 plots the results from our main specification (see Table 2 for the corresponding estimates, and Table C3 and C4 for the near identical results obtained by estimating probit and logit regressions). The x-axis indicates the treatments, expressed as the fraction of experimenters wearing a mask (0, 0.25, 0.5, 0.75, 1). The y-axis displays the predicted share of individuals wearing a mask in each treatment. To obtain this predicted share, we set the three control variables (age, gender, and pre) equal to their mean values; so we are implicitly correcting for any imbalance in the pre variable. Thus, in the language of Section 2, Figure 2 displays our preferred estimates of $f(0)$, $f(0.25)$, $f(0.5)$, $f(0.75)$, and $f(1)$.

Several features of the data are apparent. First, we find evidence that the frequency of mask wearing is everywhere increasing in the share of experimenters who wear a mask. This pattern is evident in all the specifications that we estimate, regardless of whether they include controls, use logit or probit, etc. (again, see Tables 2, C3 and C4). From a statistical point of view, we can reject the hypothesis that lower treatments lead to the same levels of mask

Figure 2: Mask wearing by treatment group



Notes. This figure shows how mask wearing varies by treatment group, after setting all covariates in the 'main specification' to their mean value.

wearing as higher treatments for the large majority of treatment pairs (see Table C5), with the exception of the comparison of T0/T1 and the comparison of T1/T2. While we discuss mechanisms later on, we note that this is consistent with a model in which higher rates of mask wearing lead to greater social pressure to wear a mask.

Second, we see that many individuals defy social pressure. In the treatment in which no experimenters wear a mask (T0), 20.0% of the participants nonetheless choose to wear a mask, a share which is statistically different from zero ($p < 0.0001$). In the language of Angrist et al. (1996), these people can be interpreted as 'always wearers', i.e. individuals who choose to wear a mask no matter how many others are doing the same (see Section 5 for elaboration). Similarly, in the treatment in which all experimenters wear a mask (T4), only 48.7% choose to wear a mask, which is again statistically different from 1 ($p < 0.0001$). The remaining 51.3% of individuals (who do not wear a mask) can be interpreted as 'never wearers', i.e. individuals who will never choose to wear a mask, no matter how many others are doing so (again, see Section 5 for a more formal discussion of this point).

Similar results are available if we look at changes. In treatment T0, out of the participants

Table 2: Regressions (experiment 1)

Variable	No controls	Main Specification	All Controls
Treatment 1	.044	.032	.020
	[.048]	[.029]	[.033]
Treatment 2	.171***	.078**	.075**
	[.053]	[.032]	[.035]
Treatment 3	.238***	.163***	.156***
	[.055]	[.039]	[.041]
Treatment 4	.331***	.284***	.289***
	[.054]	[.043]	[.046]
Pre		.757***	.741***
		[.029]	[.035]
Age		.002	.001
		[.005]	[.005]
Male		-.007	-.007
		[.026]	[.028]
Constant	.157***	.014	.130
	[.032]	[.107]	[.144]
n	646	646	646
R^2	0.070	0.494	0.517

Notes. This table reports our main regressions. To obtain the estimates in the first column, we regress whether subjects wore a mask on the treatment dummies. In the second column, we control for subject age, gender, and whether they entered wearing a mask. The third column also includes session and college fixed effects. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

who entered the room wearing a mask, only 5.6% chose to take off their mask (see Table C6). Similarly, in treatment T4, out of the participants who entered the room not wearing a mask, only 36.8% chose to put on a mask. It is quite striking that the majority of those who entered without a mask in T4 decided to defy social pressure in this way, especially given that all four experimenters were clearly visible and that a box of masks was available.

Third, our estimated f function appears to be non-linear. Estimating a model with a quadratic term suggests some convexity ($p = 0.04$): see Table C7. Insofar as estimates appear non-linear, this is due to a large jump between the 3 and 4 treatments (the difference is 12 percentage points, as opposed to the average difference between treatments of 7 percentage points). This is indicative of a potential 'everybody effect', i.e. that a particularly large change in behaviour is induced by changing the share who are doing an action from 'most people' to 'everybody'.

Finally, we examine what our estimates imply when embedded in plausible dynamic models of the style discussed in Section 2. This delivers our fourth and perhaps most important finding: when embedded in such models, our estimates predict convergence to an interior equilibrium. In the very simple model discussed in Section 2, our model predicts global convergence to the fixed point of the estimated f function, which is about 23.3%.[8] We obtain very similar results when we use our estimates to calibrate our network model (see Appendix B), which predicts that around 23%-24% should wear a mask (with almost no dependence on the initial conditions). In these equilibria, around $0.20/0.23 \approx 87\%$ of mask wearers wear the mask because they always wear one; with the remainder wearing a mask due in part to copying behaviour. We discuss these interior equilibria in more detail in the next section (which obtains even more striking evidence of interiority).

Before moving to our second experiment, we briefly discuss the results of our online follow-up survey. As explained earlier, this survey directly asked participants how their decision to wear a face mask would vary with the number of individuals in the room who were also wearing a face mask. Given that individuals might not always know what they would do in a hypothetical situation, we do not emphasise the estimated f function obtained from this survey (although, reassuringly, it is also monotone increasing in the number of mask wearers). However, we use the follow-up survey to address two issues that our original experiment could not speak to, namely individual level f functions and mechanisms.

Our first finding from the online survey is that individual decision rules are plausibly monotone in the share of experimenters who are wearing a face mask. Indeed, over 99% of subjects report weakly increasing decision rules: if such subjects chose to wear a mask in some treatment Tk, they would also choose to wear the mask in all treatments Tk' for $k' > k$. This finding helps validate our assumption in Section 5 that individual preferences have a tipping point representation, which in turn provides an insightful decomposition of the observed aggregate behaviour. We should perhaps also stress that this finding cannot be obtained from the data from our main experiment, which is in principle consistent with the possibility that many individuals have decreasing decision rules.

Second, we obtain some suggestive evidence on why individuals are more likely to wear a mask if they see more mask wearing in their immediate environment. To do this, we consider only those individuals who reported that they would change their mask-wearing behaviour depending on the share of others wearing a mask. We then placed the explanation into various categories, including whether individuals were trying to avoid being judged, trying to put others at ease, or taking high rates of mask wearing as a sign of high COVID risk levels

[8] This fixed point is obtained by linearly interpolating between $f(0)$ and $f(0.25)$. However, given that the fixed point is close to 0.25, near identical fixed points are obtained through other methods, e.g. quadratic interpolation.

(see Appendix E for a more detailed explanation of our categories along with examples). The main message from this exercise is that the health-based mechanism (i.e. that masks are used as a signal of COVID rates) is extremely unlikely to be driving our results: see Table C8 for details. Instead, the observed changes seem to be driven by a variety of social learning and social pressure mechanisms, although exactly identifying the relative importance of these mechanisms is challenging.[9]

4 Cameras

4.1 Experimental design

In order to study the generality of our results, we conducted a second experiment which used a near-identical methodology in a very different context. The basic idea of this second experiment was the following. Subjects joined a Zoom call knowing solely that they were taking part in some kind of economics experiment. Unbeknownst to them, the number of experimenters on the call with their video camera on had been randomised. We then observed whether each subject themselves chose to use their camera. Thus, this second experiment was essentially the same as the first, except with the subject of video-camera instead of face mask usage.[10]

This second experiment took place online in late July and early August of 2022. We conducted 16 two-hour sessions over the course of 8 days (with the help of 20 research assistants, some of whom participated in multiple sessions). On average, each session was attended by around 70 participants, leading to a sample size of 1,113 participants in total (see Table C9 for the distribution of subjects across treatment groups). We recruited all participants from Prolific, and required all participants to have a working microphone and video camera.[11]

The structure of the experiment was as follows:

1. Subjects were asked to join a Zoom call at a particular time.
2. Before each subject joined the call, the number of the four experimenters in the meeting with their camera on (and which experimenters had their camera on) had been randomised. Thus, there were again five treatment groups: 0/4 cameras (denoted

[9] One particular issue is that individuals may not be entirely honest about the reasons for their behaviour. For example, they might overstate the extent to which their behaviour is driven by altruistic reasons (e.g. trying to put others at ease), as opposed to a fear of being judged.

[10] This experiment also received approval from the University of Oxford's Departmental Research Ethics Committee (ECONCIA21-22-44).

[11] The experiment was pre-registered here: https://www.socialscienceregistry.org/trials/9829.

treatment T0), 1/4 cameras (T1), 2/4 cameras (T2), 3/4 cameras (T3), 4/4 cameras (T4).

3. Once a subject joined the call, all four experimenters introduced themselves by stating their name. The purpose of this was to ensure that each subject fully processed the number of experimenters whose cameras were on.

4. The subject was asked for their age, and whether they would hypothetically want to donate half of a bonus payment to the next subject on the call.

5. We then asked the subject to leave the call, and repeated the process for the next subject (again, see Appendix D for a more detailed description of the experimental protocol).

Similarly to before, we recorded whether each subject had already turned their camera on when they joined the call; and whether they chose to turn their camera on after interacting with the experimenters. We also recorded their choice in the decision problem; as well as whether they asked if they ought to turn their camera on (in such cases, each was told that 'it's up to you'). Finally, if a subject had not turned their camera on at any point during the call, we asked them if there were any issues with their video camera.[12]

4.2 Results

We now turn to our results, beginning again with a description of our sample. In contrast with the student population studied in our first experiment, the average participant in this experiment was around 42 years old, with a standard deviation of 13.9 years (see Table C10). Around 46% of the sample was male. As shown by Table 3, ages and genders were reasonably balanced across each of our five treatments. However, we again observe some imbalance in the share who joined the call with their camera on (the 'pre' variable), and so control for this variable in our main specification.

Our regressions take the same form as Equation 1. That is, we regress whether an individual used their camera on the treatment dummies (using treatment T0 as the omitted category), and a vector of covariates. In our main specification, we control for participant age, gender, and whether they joined the call with their camera on. However, we once again also report uncontrolled regressions, as well as regressions that include the full set of possible controls (including session fixed effects).

[12] Unsurprisingly, asking this question occasionally had the effect of prompting participants to turn their video camera on. In such cases, we still recorded such participants as having chosen to not use their camera (on the basis that they had chosen not to use their camera until effectively asked to do so).

Table 3: Balance table (experiment 2)

Variable	T0	T1	T2	T3	T4	p-value
Age	42.2	43.4	42.3	41.3	42.7	.615
	[.940]	[.931]	[.903]	[.906]	[.990]	
Pre	.116	.039	.058	.074	.070	.039
	[.021]	[.014]	[.016]	[.017]	[.018]	
Male	.472	.441	.439	.455	.516	.486
	[.033]	[.035]	[.033]	[.032]	[.034]	

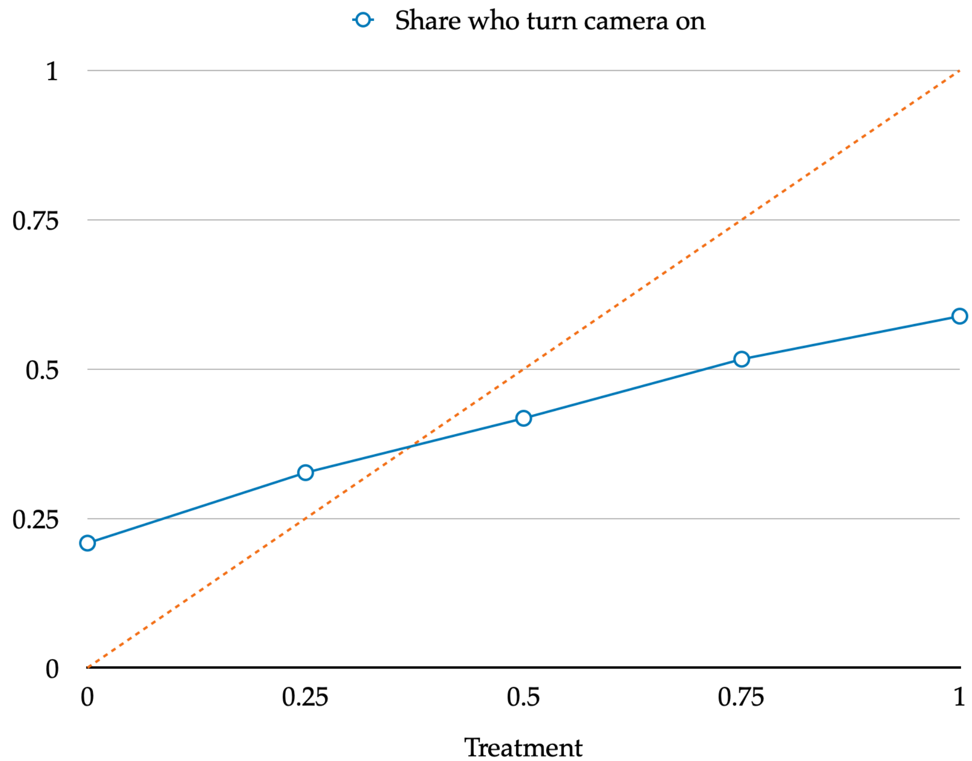
Notes. This table shows the average value of various variables across the five treatments. The variables are age, whether the subject joined the call with their camera on ('pre'), and gender. The final column reports the p-value obtained from regressing the relevant variable on all treatment dummies and testing the hypothesis that the coefficients on all treatment dummies are equal to zero.

Figure 3 plots the results from our main specification (see Table 4 for the corresponding estimates, and Tables C11 and C12 for the near identical results obtained by estimating probit and logit regressions). Several points are apparent. First, similarly to the face mask experiment, we once again observe a monotone f function: the frequency of camera use is everywhere increasing in the share of experimenters who use a camera. This pattern arises in all of the specifications we estimate (see Tables 4, C11 and C12). In our main specification, we can reject the hypothesis that treatment i and treatment $i+1$ lead to the same rates of camera usage ($p < 0.05$) for all i except $i = 3$; and we can always reject the hypothesis that treatment i and treatment $i+2$ lead to the same rates of camera usage ($p < 0.01$) — see Table C13 for details. As before, this monotonicity is consistent with a model in which higher rates of camera use lead to greater social pressure to use a camera.

Second, we once again observe that many individuals defy social pressure. In the treatment in which no experimenters use a camera (T0), 20.9% of the participants nonetheless choose to use a camera, a share which is statistically different from zero ($p < 0.0001$). As explained in Section 5, such participants can be interpreted as 'always users', i.e. individuals who use a camera no matter how many others do the same. Similarly, in the treatment in which all experimenters use a camera (T4), only 58.7% choose to use a camera, which is again statistically different from 1 ($p < 0.0001$). The remaining 41.3% of individuals (who do not use a camera) can be interpreted as 'never users', i.e. individuals who will never choose to use a camera, no matter how many others are doing so. As before, similar results can be obtained by examining changes — see Table C14.

Third, the estimated f function in this context appears to be more linear. Statistically, we cannot reject a linear model: see Table C15. However, the jump between the 0 and 1 treatments (about 12 percentage points, in the main specification) is larger than the other

Figure 3: Camera use by treatment group



Notes. This figure shows how camera use varies by treatment group, after setting all covariates in the 'main specification' to their mean value.

3 jumps (which are 9 percentage points, 10 percentage points, and 8 percentage points respectively). This provides some suggestive evidence on non-linearity, although one would need to obtain a larger sample to investigate this issue in greater detail. We should perhaps emphasise that, linear or not, our estimated f function is clearly different to that generated in standard evolutionary game theory models, which predict an S shape (see Section 2 for elaboration).

Fourth, and most importantly, our estimates once again predict convergence to an interior equilibrium when embedded in plausible dynamic models. When embedded in the model from Section 2, our estimates predict convergence to the fixed point of the estimated f function, which is about 37.0%.[13] We obtain similar results in our network model, which predicts that about 34.3% should turn the camera on (averaging across our models and initial conditions). In the equilibrium of the simple model (for example), around $0.209/0.370 \approx 56\%$ of camera users use a camera because they always use one; with the remainder using a camera

[13] This fixed point is calculated by linearly interpolating between $f(0.25)$ and $f(0.5)$. Given that the f function appears roughly linear, this seems like a sensible approach. Moreover, our network model delivers similar predictions without the need to rely on any kind of interpolation.

Table 4: Regressions (experiment 2)

Variable	No controls	Main Specification	All Controls
Treatment 1	.077*	.118***	.125***
	[.043]	[.040]	[.041]
Treatment 2	.176***	.209***	.214***
	[.043]	[.039]	[.044]
Treatment 3	.281***	.308***	.320***
	[.043]	[.039]	[.049]
Treatment 4	.355***	.380***	.386***
	[.044]	[.041]	[.057]
Pre		.579***	.581***
		[.033]	[.034]
Age		.000	.000
		[.001]	[.001]
Male		.024	.023
		[.027]	[.027]
Constant	.241***	.155***	.094
	[.028]	[.047]	[.061]
n	1,113	1,111	1,109
R^2	0.069	0.161	0.183

Notes. This table reports our main regressions. To obtain the estimates in the first column, we regress whether subjects used a camera on the treatment dummies. In the second column, we control for subject age, gender, and whether they joined the call with their camera on. The third column also includes session fixed effects. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

due in part to copying behaviour.

In canonical evolutionary game theory models (Young, 1993, 2009), the system typically converges to a situation in which either everybody or nobody does the relevant activity. In light of this, it may be worth explaining why our models predict interior equilibria (e.g. that 56% of individuals use a camera) despite the presence of copying-like behaviour. The explanation is as follows. According to our estimates, there is a substantial share of individuals who do the relevant activity no matter what others are doing. Given the behaviour of these individuals, others are induced to also do the activity, leading to a gradual increase in the share who do the activity (if the initial share is low). Eventually, the share of those doing the activity reaches the (unique) fixed point of our estimated f function, at which point the process stops. Likewise, if the share doing the activity is initially very high, then it gradually falls until it reaches the fixed point of our f function.

5 Discussion

In Section 2, we discuss what our estimates imply for long-run equilibria once embedded within dynamic models. However, we have not yet examined which models could give rise to our empirical findings; and what might explain the commonalities and differences in f functions across our two contexts. We turn to this question in the present section.

To provide a formal explanation for our results, let us assume that individual preferences have a tipping point representation. Formally, this means that, for every individual i, there exists a number $t_i \in [0, 1]$ such that the individual does the action ($a_i = 1$) if and only if $s \geq t_i$, where $s \in [0, 1]$ is the share of others who are doing it. Such an assumption seems very plausible in our two contexts; and it is given some empirical validation by the online survey discussed in Section 3.2.

Under this assumption,

$$f(s) = \frac{1}{n} \sum_i \mathbb{1}(a_i = 1|s) = \frac{1}{n} \sum_i \mathbb{1}(t_i \leq s), \tag{2}$$

where $\mathbb{1}(a_i = 1|s)$ is an indicator function that equals one if an individual takes the action (given that the share taking the action is s), and $\mathbb{1}(t_i \leq s)$ is an indicator function that equals one if an individual's tipping point exceeds the share. We thus see that, under our assumption, the f function is *exactly* the cumulative distribution of individual tipping points. Viewed in this way, our two field experiments can be seen as an experimental investigation of tipping point distributions.

Given that our experiments can be used to estimate the cumulative distribution of individual tipping points, it is straightforward to compute the probability distribution. To see how this works in practice, consider the data from the face mask experiment and define p_i as the share with a tipping point of i, for $i \in \{0, 1, 2, 3, 4\}$ (for convenience, we now use non-normalised tipping points, i.e. we do not divide by the population size). That is, for such values of i, p_i is the share of individuals who take the action if and only if they observe i or more of the four people in the room doing the same. Let us also define p_5 as the share who *never* do the action; and observe that $p_5 = 1 - \sum_{i=0}^{4} p_i$.

Table 5 reveals how the expected frequency of mask-wearing depends on the p_i parameters. In treatment T0, the only type who will do the action are the 'always doers', so the predicted share is p_0. In treatment T1, the types who do the action are the 'always doers' in combination with those who tip when they see one person doing the action. More generally, in treatment Tk, the expected share who will do the action is $\sum_{i=0}^{k} p_i$. Given this, one can estimate the p_i by matching the parameters with the sample frequencies (as suggested, e.g., by maximum likelihood). For example, we obtain the estimates $\hat{p}_0 = 0.203$; and obtain $\hat{p}_1$, $\hat{p}_2$, $\hat{p}_3$, $\hat{p}_4$ by computing the difference of mask wearing between neighbouring treatments. Finally, our estimate for the 'never do-ers' is obtained using $\hat{p}_5 = 1 - \sum_{i=0}^{4} \hat{p}_i$.[14]

Table 5: Tipping points (experiment 1)

Treatment	Frequency	Predicted frequency
0	0.203	p_0
1	0.235	$p_0 + p_1$
2	0.279	$p_0 + p_1 + p_2$
3	0.364	$p_0 + p_1 + p_2 + p_3$
4	0.487	$p_0 + p_1 + p_2 + p_3 + p_4$

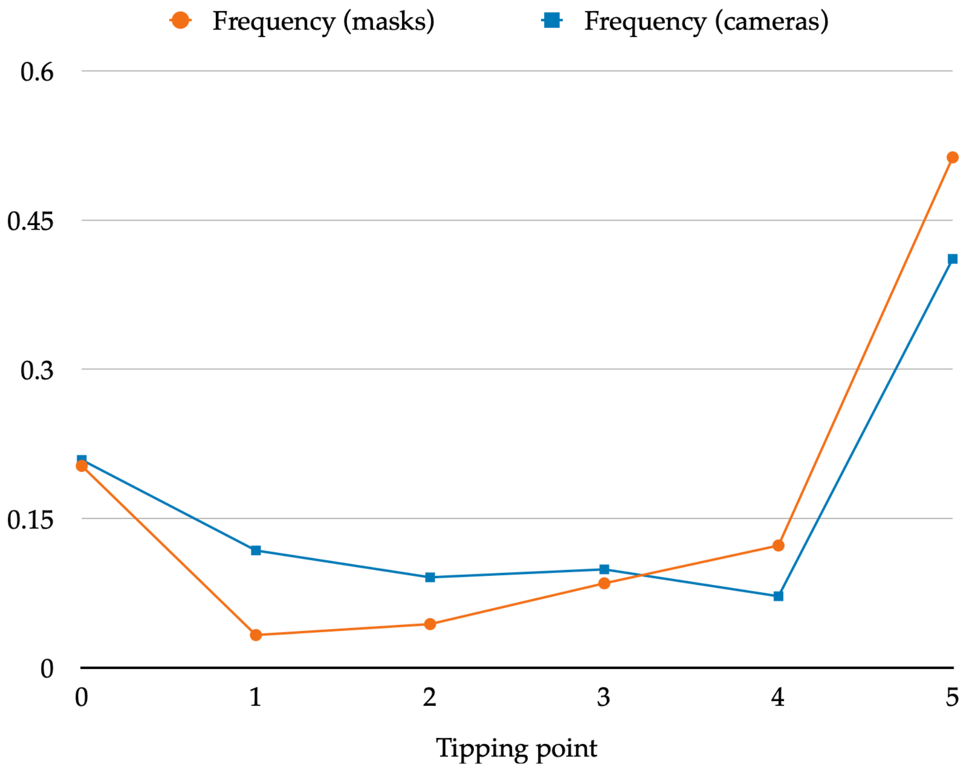
Figure 4 plots the results for both experiments. By construction, the distribution of tipping points plotted in the figure exactly generates the experimentally observed f functions. As a result, it is possible to rationalise any observed differences across contexts by postulating a difference in the distribution of tipping points. For example, in the face mask experiment, a substantially lower share are estimated to have tipping points of 1 and 2 than in the Zoom experiment. This can be used to explain why the estimated f function is flatter over the $0 - 2$ range in the face mask experiment. Similarly, the observed non-linearity in the mask experiment can be rationalised by postulating that an especially large fraction have a tipping point of 4.

Although a model of heterogeneous tipping points is able to rationalise our results — and indeed can rationalise *any* non-decreasing f function — the explanation is rather mechanical. More precisely, while our results can be viewed as the cumulative distribution of tipping points, the question remains as to *why* the distributions take the form that they do. We now address this question using the simplest possible model of individual tipping points.

To this end, consider an individual i who is deciding whether to take the action or not given that a fraction s is already doing so. If they take the action, they obtain utility $u(a_i = 1) = \alpha_i + m(s)$, where $\alpha_i \in \mathbb{R}$ describes their intrinsic preference for taking the action and $m(s)$ describes their 'coordination payoff' from doing the same thing as a fraction s of their neighbours (assume m is differentiable). If they do not take the action, they get utility

[14] This approach implicitly treats the set of possible tipping points at discrete. It is also possible, however, to view (normalised) tipping points as continuously distributed on $[0, 1]$. In that case, our estimates can be used to calculate the share of normalised tipping points that are zero, the share that are between 0 and 0.25, etc.

Figure 4: Tipping point distributions



Notes. This figure shows the distributions of individual tipping points calculated from our two sets of experimental estimates. Individuals with tipping points of 0 and 5 are 'always doers' and 'never doers' respectively.

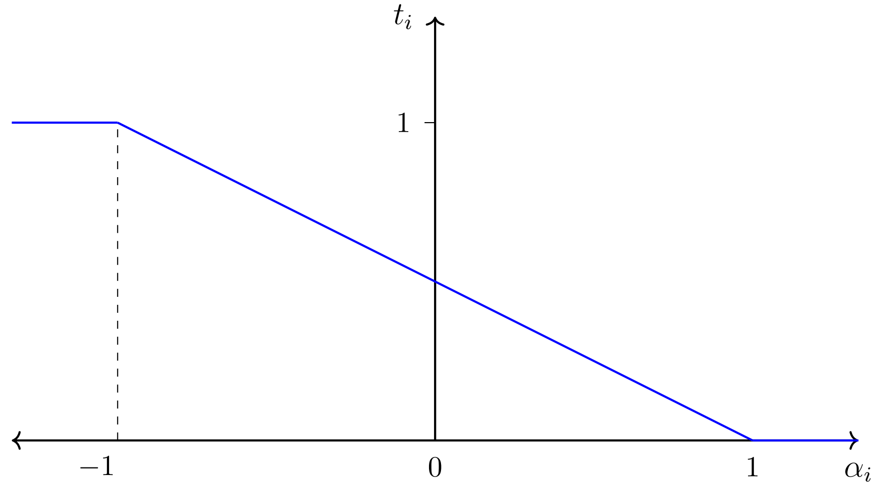
$u(a_i = 0) = m(1 - s)$. An (interior) tipping point $t_i \in (0, 1)$ is a share s that makes the individual indifferent between doing the action or not doing it, i.e. $\alpha_i + m(t_i) = m(1 - t_i)$.

To illustrate, consider the case in which of $m(s) = s$. In that case, it is easy to check that $t_i = 1$ for all individuals whose preferences satisfy $\alpha_i \leq -1$. Such individuals never take the action, no matter how many others are doing so. Similarly, we have $t_i = 0$ for all individuals for which $\alpha_i \geq 1$: such individuals always take the action. Interior tipping points satisfy the equation $\alpha_i + t_i = 1 - t_i$, or $t_i = 0.5(1 - \alpha_i)$ (see Figure 5). As a result, interior tipping points are strictly decreasing in the α_i parameter (and equal to 0.5 when the individual has no intrinsic reason to do the action, i.e. $\alpha_i = 0$). Intuitively, this is because individuals with a high intrinsic preference for taking the action will be indifferent between taking the action or not even when the share of others who are taking the action is very low.

We now verify that this result holds for *any* m function with an everywhere positive slope.

Proposition 3. *If $m'(s) > 0$ for all $s \in [0, 1]$, then each individual has a well defined tipping point $t_i \in [0, 1]$. Furthermore, if $t_i \in (0, 1)$, then $\frac{\partial t_i}{\partial \alpha_i} < 0$.*

Figure 5: Tipping points when $m(s) = s$



Notes. This figure shows how an individual's tipping point t_i varies with their intrinsic preference to take the action α_i when $m(s) = s$.

Proposition 3 says that (interior) tipping points are strictly decreasing in an individual's intrinsic preference for taking the action. As a result, it provides us with an explanation for the differences we observe across our two contexts. In the mask experiment, the estimated tipping points tend to be higher (again, see Figure 5). One simple way to rationalise this, as suggested by Proposition 3, is to postulate that individuals are generally less willing to wear masks than to use their laptop cameras.

Although intrinsic preferences for taking actions do plausibly influence tipping points in the manner just discussed, we should emphasise that they are not the *only* factor that determines the shape of the f function. In our view, the f function is driven by the complicated interplay of both intrinsic preferences (captured by α_i) *and* potentially non-linear social pressure effects (captured by m). Future work could attempt to decompose these two channels.

6 Concluding remarks

In this paper, we conduct multi-treatment social norm experiments to obtain a quantitative understanding of how individuals' behaviour varies with the share doing an action in their immediate environment. Despite some differences between the estimates across our contexts (which we rationalise using a simple theory), we obtain many commonalities across the two experiments: increasing f functions, high levels of non-compliance, etc. Perhaps most importantly, when embedded in dynamic models, our estimates can explain how copying can plausibly lead to heterogenous behaviour. In our view, this is an important insight with applications to many settings (e.g. religious adoption, fashion trends, etc.): despite the ubiquity of social pressure and copying behaviour, different individuals nonetheless often end up doing very different things.

It may be worth briefly emphasising why our models generate convergence to interior equilibria, as opposed to the extreme equilibria predicted by canonical game theory models (e.g. Young, 1993, 2009; Kandori et al., 1993). In these models, each individual effectively has the *same* tipping point, which in turn generates the S shaped f function discussed in Section 2. This in turn leads to three fixed points, of which only two (the extreme equilibria) are stable. In contrast, our empirical estimates suggest: i) reasonably high levels of non-compliance ii) substantial heterogeneity in individual tipping points. These two factors generate an f function with a unique fixed point in the interior, which is the globally stable equilibrium.

Despite the large number of social norm experiments, we believe that our findings open up several avenues for future research. First, it may be worthwhile to conduct more experiments with semi-continuous randomisation in additional contexts. In particular, this could provide further evidence on whether our key finding of interior equilibria is robust. Second, it may be worthwhile to conduct such experiments with an even larger number of treatment groups, thus allowing for a more fine grained estimate of the f function. Given the very large sample sizes required to do this, however, such experiments are likely to be even more logistically challenging to implement than the two field experiments whose results we report here.

References

Agerström, J., Carlsson, R., Nicklasson, L., and Guntell, L. (2016). Using descriptive social norms to increase charitable giving: The power of local norms. *Journal of Economic Psychology*, 52:147–153.

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

Allcott, H. and Kessler, J. B. (2019). The welfare effects of nudges: A case study of energy use social comparisons. *American Economic Journal: Applied Economics*, 11(1):236–76.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables. *Journal of the American Statistical Association*, 91(434):444–455.

Asch, S. E. (1951). Effects of group pressure upon the modification and distortion of judgments. *Organizational Influence Processes*, 58:295–303.

Ayres, I., Raseman, S., and Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization*, 29(5):992–1022.

Barceló, J. and Sheen, G. C.-H. (2020). Voluntary adoption of social welfare-enhancing behavior: Mask-wearing in spain during the COVID-19 outbreak. *PLOS ONE*, 15(12):e0242764.

Barile, J. P., Guerin, R. J., Fisher, K. A., Tian, L. H., Okun, A. H., Vanden Esschert, K. L., Jeffers, A., Gurbaxani, B. M., Thompson, W. W., and Prue, C. E. (2021). Theory-based behavioral predictors of self-reported use of face coverings in public settings during the COVID-19 pandemic in the united states. *Annals of Behavioral Medicine*, 55(1):82–88.

Bedenlier, S., Wunder, I., Gläser-Zikuda, M., Kammerl, R., Kopp, B., Ziegler, A., and Händel, M. (2021). Generation invisible? Higher education students' (non) use of webcams in synchronous online learning. *International Journal of Educational Research Open*, 2:100068.

Bokemper, S. E., Cucciniello, M., Rotesi, T., Pin, P., Malik, A. A., Willebrand, K., Paintsil, E. E., Omer, S. B., Huber, G. A., and Melegaro, A. (2021). Experimental evidence that changing beliefs about mask efficacy and social norms increase mask wearing for COVID-19 risk reduction: Results from the united states and italy. *PlOS ONE*, 16(10):e0258282.

Bond, R. and Smith, P. B. (1996). Culture and conformity: A meta-analysis of studies using Asch's (1952b, 1956) line judgment task. *Psychological Bulletin*, 119(1):111.

Borsari, B. and Carey, K. B. (2003). Descriptive and injunctive norms in college drinking: a meta-analytic integration. *Journal of Studies on Alcohol*, 64(3):331–341.

Bott, K. M., Cappelen, A. W., Sørensen, E. Ø., and Tungodden, B. (2020). You've got mail: A randomized field experiment on tax evasion. *Management Science*, 66(7):2801–2819.

Bursztyn, L., Ederer, F., Ferman, B., and Yuchtman, N. (2014). Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions. *Econometrica*, 82(4):1273–1301.

Cason, T. N. and Mui, V.-L. (1998). Social influence in the sequential dictator game. *Journal of Mathematical Psychology*, 42(2-3):248–265.

Castelli, F. R. and Sarvary, M. A. (2021). Why students do not turn on their video cameras during online classes and an equitable and inclusive plan to encourage them to do so. *Ecology and Evolution*, 11(8):3565–3576.

Cialdini, R. B. (2007). Descriptive social norms as underappreciated sources of social control. *Psychometrika*, 72(2):263–268.

Cialdini, R. B., Reno, R. R., and Kallgren, C. A. (1990). A focus theory of normative conduct: Recycling the concept of norms to reduce littering in public places. *Journal of Personality and Social Psychology*, 58(6):1015.

Costa, D. L. and Kahn, M. E. (2013). Energy conservation "nudges" and environmentalist ideology: Evidence from a randomized residential electricity field experiment. *Journal of the European Economic Association*, 11(3):680–702.

Damm, A. P. and Dustmann, C. (2014). Does growing up in a high crime neighborhood affect youth criminal behavior? *American Economic Review*, 104(6):1806–32.

Dodds, P. and Watts, D. J. (2011). Threshold models of social influence.

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

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

Farrow, K., Grolleau, G., and Ibanez, L. (2017). Social norms and pro-environmental behavior: A review of the evidence. *Ecological Economics*, 140:1–13.

Ferraro, P. J. and Price, M. K. (2013). Using nonpecuniary strategies to influence behavior: evidence from a large-scale field experiment. *Review of Economics and Statistics*, 95(1):64–73.

Fortin, B., Lacroix, G., and Villeval, M.-C. (2007). Tax evasion and social interactions. *Journal of Public Economics*, 91(11-12):2089–2112.

Freidin, E., Acera Martini, L., Senci, C. M., Duarte, C., and Carballo, F. (2022). Field observations and survey evidence to assess predictors of mask wearing across different outdoor activities in an Argentine city during the COVID-19 pandemic. *Applied Psychology: Health and Well-Being*, 14(1):81–100.

Frey, B. S. and Meier, S. (2004). Social comparisons and pro-social behavior: Testing "conditional cooperation" in a field experiment. *American Economic Review*, 94(5):1717–1722.

Gerber, A. S. and Rogers, T. (2009). Descriptive social norms and motivation to vote: Everybody's voting and so should you. *The Journal of Politics*, 71(1):178–191.

Gherheș, V., Șimon, S., and Para, I. (2021). Analysing students' reasons for keeping their webcams on or off during online classes. *Sustainability*, 13(6):3203.

Goldberg, M., Gustafson, A., Maibach, E., van der Linden, S., Ballew, M. T., Bergquist, P., Kotcher, J., Marlon, J. R., Rosenthal, S., and Leiserowitz, A. (2020). Social norms motivate COVID-19 preventive behaviors. *PsyArXiv*.

Goldstein, N. J., Cialdini, R. B., and Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3):472–482.

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

Heldt, T. (2005). Conditional cooperation in the field: cross-country skiers' behavior in sweden. In *Economic Science Association, European Meeting, Alessandria, Italien, September 15-18, 2005*.

Ichino, A. and Maggi, G. (2000). Work environment and individual background: Explaining regional shirking differentials in a large Italian firm. *The Quarterly Journal of Economics*, 115(3):1057–1090.

Jackson, M. O. and Yariv, L. (2007). Diffusion of behavior and equilibrium properties in network games. *American Economic Review*, 97(2):92–98.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2):263–292.

Kandori, M., Mailath, G. J., and Rob, R. (1993). Learning, mutation, and long run equilibria in games. *Econometrica*, pages 29–56.

Kreindler, G. E. and Young, H. P. (2013). Fast convergence in evolutionary equilibrium selection. *Games and Economic Behavior*, 80:39–67.

Krupka, E. and Weber, R. A. (2009). The focusing and informational effects of norms on pro-social behavior. *Journal of Economic Psychology*, 30(3):307–320.

Lefebvre, M., Pestieau, P., Riedl, A., and Villeval, M. C. (2015). Tax evasion and social information: an experiment in belgium, france, and the netherlands. *International Tax and Public Finance*, 22(3):401–425.

Linek, M. and Traxler, C. (2021). Framing and social information nudges at Wikipedia. *Journal of Economic Behavior & Organization*, 188:1269–1279.

Manski, C. F. (1993). Identification of endogenous social effects: The reflection problem. *The Review of Economic Studies*, 60(3):531–542.

Martin, R. and Randal, J. (2008). How is donation behaviour affected by the donations of others? *Journal of Economic Behavior & Organization*, 67(1):228–238.

Mascagni, G. (2018). From the lab to the field: A review of tax experiments. *Journal of Economic Surveys*, 32(2):273–301.

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

Rabin, M. (1993). Incorporating fairness into game theory and economics. *The American Economic Review*, pages 1281–1302.

Rudert, S. C. and Janke, S. (2021). Following the crowd in times of crisis: Descriptive norms predict physical distancing, stockpiling, and prosocial behavior during the COVID-19 pandemic. *Group Processes & Intergroup Relations*, page 13684302211023562.

Sederevičiūtė-Pačiauskienė, Ž., Valantinaitė, I., and Asakavičiūtė, V. (2022). 'Should I turn on my video camera?' The students' perceptions of the use of video cameras in synchronous distant learning. *Electronics*, 11(5):813.

Smith, S., Windmeijer, F., and Wright, E. (2015). Peer effects in charitable giving: Evidence from the (running) field. *The Economic Journal*, 125(585):1053–1071.

Thöni, C. and Gächter, S. (2015). Peer effects and social preferences in voluntary cooperation: A theoretical and experimental analysis. *Journal of Economic Psychology*, 48:72–88.

Woodcock, A. and Schultz, P. W. (2021). The role of conformity in mask-wearing during COVID-19. *PLOS ONE*, 16(12):e0261321.

Young, H. P. (1993). The evolution of conventions. *Econometrica*, pages 57–84.

Young, H. P. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. *American Economic Review*, 99(5):1899–1924.

Appendices

A Proofs

Proof of Proposition 1. To prove the first statement, assume that f is first convex and then concave; and also that $s_0 \neq \hat{s}$. (The proof of the second statement is entirely analogous and therefore omitted). As a preliminary, observe that, if $s_0 = 0$, then $s_t = f^t(0) = 0$ for all $t \in \mathbb{N}$ (recall that 0 is a fixed point). Hence, if $s_0 = 0$, then $\lim_{t\to\infty} s_t = 0$. Similarly, $\lim_{t\to\infty} s_t = 1$ if $s_t = 1$. Therefore, the statement holds trivially in the cases of $s_0 = 0$ and $s_1 = 1$. It remains to consider the cases of $s_0 \in (0, \hat{s})$ and $s_0 \in (\hat{s}, 1)$.

Suppose then that $s_0 \in (0, \hat{s})$. (The argument when $s_0 \in (\hat{s}, 1)$ follows similar lines and is therefore omitted.) Given that f is convex on $[0, \hat{s}]$, and furthermore that $f(0) = 0$ and $f(\hat{s}) = \hat{s}$, one may check that $f(s) < s$ for all $s \in (0, \hat{s})$. In addition, given that f is increasing, $f(s) > f(0) = 0$ for all such s. So for all $s \in (0, s^*)$, $f(s) \in (0, s)$.

Given that this fact, and also that $s_0 \in (0, s^*)$, one can show by induction that $s_t \in (0, s_{t-1})$ for all $t \in \mathbb{N}$. Hence, the sequence $\{s_t\}_{t=0}^{\infty}$ is strictly decreasing and bounded from below by zero. By the monotone convergence theorem, it therefore has a limit s^*. Furthermore, since f is continuous, every limit s^* must be a fixed point:

$$s^* = \lim_{t\to\infty} s_t = \lim_{t\to\infty} s_{t+1} = \lim_{t\to\infty} f(s_t) = f(\lim_{t\to\infty} s_t) = f(s^*), \tag{3}$$

where the penultimate equality uses the continuity of f. By assumption, f only has three fixed points: 0, $\hat{s}$, and 1. Since $s^* < s_0 < \hat{s}$, we see that the only possible limit is 0.

The argument given above establishes that $\lim_{t\to\infty} s_t = 0$ if $s_0 \in (0, \hat{s})$. By an analogous argument, one may show that $\lim_{t\to\infty} s_t = 1$ if $s_0 \in (\hat{s}, 1)$. Furthermore, we have already observed that $\lim_{t\to\infty} s_t \in \{0, 1\}$ if either $s_0 = 0$ or $s_0 = 1$. This establishes that $\lim_{t\to\infty} s_t \in \{0, 1\}$ for any initial value $s_0 \neq \hat{s}$. □

Proof of Proposition 2. As noted earlier, given that f is continuous, we know that every limit s^* must be a fixed point. Moreover, since f is increasing, $f(s) \geq f(0)$ for all $s \in [0, 1]$. In particular, then, $f(s^*) \geq f(0)$. However, since s^* is a fixed point (established earlier), $s^* = f(s^*)$. From this, we conclude that $s^* \geq f(0)$; and a symmetric argument establishes that $s^* \leq f(1)$. □

Proof of Proposition 3. Let us define the difference in utilities by

$$\Delta(\alpha_i, s) \equiv U(a_i = 1) - U(a_i = 0) = \alpha_i + m(s) - m(1 - s) \tag{4}$$

To show that the individual has tipping point preferences, we consider three cases:

Case 1. For all $s \in [0, 1]$, $\Delta(\alpha_i, s) \geq 0$. In that case, $a_i^* = 1$ for all s (where a_i^* denotes the optimal action). Equivalently, $a_i^* = 1$ if and only if $s \geq t_i$, for any tipping point that satisfies $t_i \leq 0$. Thus, the individual has tipping point preferences (for example, we may set $t_i = 0$).

Case 2. For all $s \in [0, 1]$, $\Delta(\alpha_i, s) < 0$. Similarly to before, this means that $a_i^* = 0$ for all s. Equivalently, $a_i^* = 1$ if and only if $s \geq t_i$ for any tipping point that satisfies $t_i > 1$: for example, we can set $t_i = 2$. Thus, the individual again has tipping point preferences.

Case 3. $\Delta(\alpha_i, s) \geq 0$ for some $s \in [0, 1]$; but also $\Delta(\alpha_i, s) < 0$ for some $s' \in [0, 1]$. Differentiating with respect to s, we see that

$$\Delta'(s) = m'(s) + m'(1 - s) > 0. \tag{5}$$

Thus, $\Delta(s)$ is strictly increasing (and continuous) in s. This means that there is a unique $s^* \in (0, 1]$ such that $\Delta(s) > 0$ for $s > s^*$, $\Delta(s) = 0$ when $s = s^*$, and $\Delta(s) < 0$ for $s < s^*$. Thus, the individual has tipping point preferences for $t_i = s^*$.

To prove the second statement, observe that, if $t_i \in (0, 1)$, then

$$\alpha_i + m(t_i) - m(1 - t_i) = 0 \tag{6}$$

We can totally differentiate to obtain

$$\frac{\partial \alpha_i}{\partial \alpha_i} + \frac{\partial m(t_i)}{\partial t_i}\frac{\partial t_i}{\partial \alpha_i} + \frac{\partial m(1 - t_i)}{\partial t_i}\frac{\partial t_i}{\partial \alpha_i} = 0 \tag{7}$$

which implies that

$$\frac{\partial t_i}{\partial \alpha_i} = -\frac{1}{m'(t_i) + m'(1 - t_i)} < 0 \tag{8}$$

where the inequality holds since both derivatives are strictly positive. Thus, the derivative is negative (as claimed). □

B Local interaction

In this section, we extend the model described in Section 2 to allow for local interaction in overlapping networks. The model presented here shares some similarities to the model studied by Efferson et al. (2020). An important difference is that, while Efferson et al. (2020) assume that decision makers choose randomly, we instead assume that they choose deterministically but with heterogenous decision rules. In addition, our model assumes that individuals respond to the decisions of their 'neighbours' (in line with our experimental settings); whereas Efferson et al. (2020) assume that they best respond to the entire population.

In our baseline model, we assume the following:[15]

- There are l^2 agents, each located on a node of a grid with side length $l \in \mathbb{N}^+$. Let (r, c) denote the agent located at row r and column c; so the set of agents is the set $N = \{(r, c)\colon r \in \{1, ..., l\}, c \in \{1, ..., l\}\}$.
- As in our experiments, agents are faced with a binary choice: they must either take an action (denoted $a_{r,c} = 1$) or not take the action (denoted $a_{r,c} = 0$).
- Each agent (r, c) has a set of 'neighbours' $N_{r,c}$ whose actions they can see. For each agent, we assume that $N_{r,c} = \{(i, j)\colon (i, j) \in N, |i - r| \leq 1, |j - c| \leq 1, (i, j) \neq (r, c)\}$. Observe that agents in the interior have 4 neighbours, agents on the edge have 3 neighbours, and agents in the corners have 2 neighbours.
- We define $m^1_{r,c}$ as the share of individual (r, c)'s neighbours who have chosen to do the action. Formally, $m^1_{r,c} = \frac{1}{|N_{r,c}|} \sum_{(r,c) \in N_{r,c}} a_{r,c}$ where $|N_{r,c}|$ is the cardinality of $N_{r,c}$.
- Each agent is endowed with a (fixed) tipping point $t_{r,c} \in [0, 1]$. As in the main text, we assume that they choose $a_{r,c} = 1$ if and only if $t_{r,c} \geq m^1_{r,c}$.
- Agents interact over multiple periods. In each period, one agent is chosen to move at random; and updates their action (if necessary) by comparing their tipping point $t_{r,c}$ with the share of their neighbours who are taking the action $m^1_{r,c}$.

To assess the robustness of our results, we also study an alternative model that departs from the model sketched above in various ways.[16] In this model — which we label the *edgeless model* — each agent is linked with the same number of neighbours. In addition, each agent has a probability $\epsilon \in [0, 1]$ of making a 'mistake', i.e. choosing the opposite action as that required by their tipping point. Finally, a share $p \in [0, 1]$ of agents are selected in period to

[15] The corresponding code can be viewed here: https://github.com/Itzhak95/tipping_points
[16] The corresponding code can be viewed here: https://github.com/rrozzi/tipping_point-netlogo-

revise their action; so in principle multiple agents can update their action simultaneously.

To simulate the results of our models, we use the following procedure:

- We specify a distribution of tipping points in the population, and randomly scatter these tipping points across the agents.
- We also specify the share of agents who initially take the action; and we randomly scatter the agents who are taking the action on grid.
- We then allow the model to run for 1000 periods (or until it is 'stable' so no further changes can occur).

As discussed above, the results of the model could in principle depend on the way in which tipping points and initial actions are scattered. As a result, we conduct all simulations 1000 times and report the distribution of results across simulations.

Before turning to our main results, we provide a simple example to illustrate the mechanics of the model. To generate this example, we suppose that, initially, 40% of agents are taking the action; and we set $s = 5$. In addition, we assume (for expositional simplicity) that all agents have a tipping point $t_{r,c} = 0.5$, so choose $a_{r,c} = 1$ if and only if half or more of their neighbours are doing the action. While one would normally repeat the simulation many times, here we just report the outcome of one simulation.

After randomly scattering the initial actions, we obtain the initial state

0	1	1	1	0
0	1	0	0	1
1	0	1	1	0
0	0	0	0	0
1	0	0	1	0

As can be seen, 10 of the $s^2 = 25$ agents initially take the action (indicated by a 1); the rest do not. Several rounds now progress in which the player chosen to move does not wish to update their action. Eventually, however, the player at row 3 and column 1 is chosen to move (they are coloured in red). Since none of their neighbours (coloured in blue) were taking the action, they choose to switch to action 0. This yields the new state

$$\begin{matrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \end{matrix}$$

As the process continues, additional players are given the opportunity to also revise their action. After 13 such revisions, we finally obtain the state

$$\begin{matrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{matrix}$$

This state is *stable* in the sense that no agent has an incentive to change their behaviour. The agent in the top left is surrounded by neighbours who choose $a_{r,c} = 1$, so would also want to choose $a_{r,c} = 1$ if allowed to update their action. The agent at $(2, 2)$ is surrounded by 4 neighbours, half of whom are taking the action; so also chooses $a_{r,c} = 1$ (recall that all tipping points are set at $t_{r,c} = 0.5$). Meanwhile, the agents at $(2, 1)$ and $(1, 2)$ are each surrounded by 3 neighbours, 2 of whom are choosing the action; so they also wish to choose the action. Finally, one can verify that the agents choosing $a_{r,c} = 0$ are choosing optimally given their tipping point and the share of their neighbours who are taking the action.

We now calibrate our model using the tipping point distributions calculated in Section 5. We assume a population size of 100; and the edgeless model further assumes an error probability $\epsilon = 0.01$ and a probability of revision $p = 0.07$. As stated above, each simulation is run for 1000 periods (or until the obtained state is stable); and all simulations are conducted $1,000$ times. Tables B1 and B2 display the results for experiment 1 (face masks) and experiment 2 (Zoom calls) respectively. The first row specifies the initial share who are assumed to do the action. The rows 'mean (main)' and 'mean (edgeless)' display the average share who end up doing the activity in the main specification and edgeless model respectively. The rows 'Var (main)' and 'Var (edgeless)' specify the variance of outcomes across simulations.

Table B1: Simulation results (experiment 1)

Initial share	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
Mean (main)	.228	.230	.230	.231	.231	.233	.234	.235	.235	.236
Var (main)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000
Mean (edgeless)	.237	.242	.242	.242	.240	.242	.240	.239	.240	.239
Var (edgeless)	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000

Notes. This table shows the results of simulating our models using the distribution of tipping points obtained by experiment 1 (see Table 2). That is, we set $p_0 = .203$, $p_1 = .033$, $p_2 = .044$, $p_3 = .085$, $p_4 = .123$, $p_5 = .513$.

Table B2: Simulation results (experiment 2)

Initial share	0.05	0.15	0.25	0.35	0.45	0.55	0.65	0.75	0.85	0.95
Mean (main)	.311	.316	.321	.329	.333	.338	.341	.345	.349	.353
Var (main)	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001
Mean (edgeless)	.353	.353	.347	.356	.355	.362	.354	.360	.356	.355
Var (edgeless)	.001	.001	.001	.001	.001	.001	.001	.001	.001	.001

Notes. This table shows the results of simulating our models using the distribution of tipping points obtained by experiment 2 (see Table 4). That is, we set $p_0 = .209$, $p_1 = .118$, $p_2 = .091$, $p_3 = .099$, $p_4 = .072$, $p_5 = .411$.

Three results are apparent. First, we see that the results of the simulations are relatively insensitive to the initial share who are assumed to do the activity. This is especially true in the edgeless model since this assumes that agents occasionally make errors, which weakens dependence on initial conditions in the usual way (Young, 1993). Second, the variance in outcomes across simulations is very low, which is again points to the lack of importance of initial conditions (since different simulations generate different outcomes only due to variation in initial conditions). Finally, and most importantly, we see that our models generate convergence to interior equilibria that resemble those obtained from the simple model of Section 2. This should not come as a surprise given that some agents always do the action, that other agents never do the action, and that a final group of agents engage in copying behaviour.

C Tables and figures

Table C1: Sample allocation (experiment 1)

Treatment	Frequency	Percentage
0	127	19.7
1	134	20.7
2	128	19.8
3	124	19.2
4	133	20.6
Total	646	100.0

Notes. This table shows how many subjects were allocated into each of the five treatments in the first experiment.

Table C2: Descriptive statistics (experiment 1)

Variable	Mean	Std. Dev.
Age	20.8	3.90
Male	.497	.500
Humanities	.283	.451
MPLS	.240	.427
Medical Sciences	.127	.333
Social Sciences	.333	.471
Pre	.201	.401
n	646	

Notes. This table shows the descriptive statistics for experiment 1 (see Table 1 for a description of the variables).

Table C3: Logit regressions (experiment 1)

Variable	No controls	Main Specification	All Controls
Treatment 1	.044	.033	.029
	[.047]	[.030]	[.034]
Treatment 2	.171***	.073**	.079**
	[.053]	[.032]	[.035]
Treatment 3	.238***	.162***	.168***
	[.055]	[.040]	[.043]
Treatment 4	.331***	.283***	.304***
	[.054]	[.042]	[.046]
Pre		.504***	.498***
		[.030]	[.031]
Age		.003	.002
		[.005]	[.004]
Male		-.006	-.002
		[.026]	[.028]
n	646	646	620

Notes. This table reports the exact same specifications reported on in Table 2, except using logistic instead of linear regressions. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table C4: Probit regressions (experiment 1)

Variable	No controls	Main Specification	All Controls
Treatment 1	.044	.036	.029
	[.047]	[.031]	[.034]
Treatment 2	.171***	.078**	.078**
	[.053]	[.033]	[.035]
Treatment 3	.238***	.163***	.162***
	[.055]	[.040]	[.043]
Treatment 4	.331***	.284***	.298***
	[.054]	[.043]	[.046]
Pre		.518***	.512***
		[.024]	[.027]
Age		.002	.001
		[.004]	[.004]
Male		-.007	-.004
		[.026]	[.028]
n	646	646	620

Notes. This table reports the exact same specifications reported on in Table 2, except using probit instead of linear regressions. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table C5: Comparisons (experiment 1)

Comparison	No controls	Main specification	All controls
T0 vs T1	.355	.278	.536
T1 vs T2	.019	.205	.163
T2 vs T3	.269	.051	.068
T3 vs T4	.131	.019	.017
T0 vs T2	.001	.014	.032
T1 vs T3	.001	.002	.002
T2 vs T4	.008	.000	.000

Notes. This table reports p-values corresponding to hypothesis that the effect of treatment k is the same as the effect of treatment k', for all possible $k \neq k'$. We do this for the three specifications considered in Table 2.

Table C6: Changes (experiment 1)

	T0	T1	T2	T3	T4
Putting mask on	.028	.080	.106	.223	.368
Taking mask off	.056	.143	.059	.067	.037

Notes. The first row shows the share who put a mask on given that they entered the room without wearing a mask. The second row shows the share who took their mask off given that they entered the room wearing a mask.

Table C7: Polynomial regressions (experiment 1)

Variable	Linear	Quadratic	Cubic
Masks	.070***	0.008	0.024
	[-.010]	[-.028]	[-.062]
Masksˆ2		.016**	.004
		[-.008]	[-.045]
Masksˆ3			.002
			[-.008]
Pre	.752***	.757***	.757***
	[-.029]	[-.029]	[-.029]
Age	.002	.002	.002
	[-.005]	[-.005]	[-.005]
Male	-.008	-.007	-.007
	[-.026]	[-.026]	[-.026]
Constant	-.022	.016	.014
	[-.102]	[-.107]	[-.107]
Joint test	.000	.000	.000
R^2	.491	.494	.494

Notes. In this table, we regress whether subjects chose to wear a mask on the number of experimenters wearing a mask, as well higher order terms to capture potential non-linearity (we also control for 'pre', age, and gender). The penultimate row reports p-values corresponding to the hypothesis that the coefficients on all mask variables are zero.

Table C8: Explanations from online survey

Explanation	Frequency
Trying to avoid judgement	.148
Trying to cater to others' preferences	.511
Trying to follow rules	.148
Reciprocity	.023
COVID risks	.011
Not answering question	.159
n	88

Notes. This table shows the frequencies of the explanations given by subjects (see Appendix E for a detailed description of the categories).

Table C9: Sample allocation (experiment 2)

Treatment	Frequency	Percentage
0	232	20.8
1	204	18.3
2	223	20.0
3	241	21.7
4	213	19.1
Total	1113	100.0

Notes. This table shows how many subjects were allocated into each of the five treatments in the second experiment.

Table C10: Descriptive statistics (experiment 2)

Variable	Mean	Std. Dev.
Age	42.4	13.9
Male	.465	.499
n	1113	

Notes. This table shows the descriptive statistics for experiment 2.

Table C11: Logit regressions (experiment 2)

Variable	No controls	Main specification	All controls
Treatment 1	.077*	.127***	.133***
	[.043]	[.039]	[.039]
Treatment 2	.176***	.215***	.218***
	[.043]	[.039]	[.040]
Treatment 3	.281***	.314***	.323***
	[.043]	[.039]	[.045]
Treatment 4	.355***	.385***	.389***
	[.044]	[.041]	[.051]
Pre		.741***	.743***
		[.092]	[.092]
Age		.000	.000
		[.001]	[.001]
Male		.023	.023
		[.027]	[.027]
n	1,113	1,111	1,109

Notes. This table reports the exact same specifications reported on in Table 4, except using logistic instead of linear regressions. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table C12: Probit regressions (experiment 2)

Variable	No controls	Main specification	All controls
Treatment 1	.077*	.125***	.130***
	[.043]	[.039]	[.039]
Treatment 2	.176***	.216***	.218***
	[.043]	[.039]	[.040]
Treatment 3	.281***	.312***	.321***
	[.043]	[.039]	[.046]
Treatment 4	.355***	.385***	.389***
	[.044]	[.040]	[.052]
Pre		.701***	.699***
		[.075]	[.076]
Age		.000	.000
		[.001]	[.001]
Male		.024	.025
		[.027]	[.027]
n	1,113	1,111	1,109

Notes. This table reports the exact same specifications reported on in Table 4, except using probit instead of linear regressions. Robust standard errors in parentheses (*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$).

Table C13: Comparisons (experiment 2)

Comparison	No controls	Main Specification	All Controls
T0 vs T1	.074	.003	.002
T1 vs T2	.035	.043	.051
T2 vs T3	.022	.028	.020
T3 vs T4	.116	.116	.152
T0 vs T2	.000	.000	.000
T1 vs T3	.000	.000	.000
T2 vs T4	.000	.000	.001

Notes. This table reports p-values corresponding to hypothesis that the effect of treatment k is the same as the effect of treatment k', for all possible $k \neq k'$. We do this for the three specifications considered in Table 4.

Table C14: Changes (experiment 2)

	T0	T1	T2	T3	T4
Turning camera on	0.156	0.296	0.381	0.491	0.566
Turning camera off	0.111	0.125	0.000	0.059	0.000

Notes. The first row shows the share who turned their camera on given that they joined the call without video. The second row shows the share who turned their camera off given that they joined the call with video.

Table C15: Polynomial regressions (experiment 2)

Variable	Linear	Quadratic	Cubic
Cameras	.095***	.119***	.119
	-.009	-.032	-.074
Cameras^2		-.006	-.006
		-.008	-.049
Cameras^3			.000
			-.008
Pre	.576***	.578***	.578***
	-.033	-.033	-.033
Age	.000	.000	.000
	-.001	-.001	-.001
Male	.023	.024	.024
	-.027	-.027	-.027
Constant	.169***	.156***	.156***
	-.046	-.047	-.047
Joint test	.000	.000	.000
R^2	.161	.161	.161

Notes. In this table, we regress whether subjects chose to use their camera on the number of experimenters using a camera, as well higher order terms to capture potential non-linearity (we also control for ‘pre’, age, and gender). The penultimate row reports p-values corresponding to the hypothesis that the coefficients on all camera variables are zero.

D Experimental protocols (online appendix)

In this section, we provide a more detailed outline of the experimental protocols followed in both experiments.

Experiment 1 (face masks). There were four experimenter roles, labelled 1 through 4. Experimenter 1's role was to greet the subject and (at the end) bid them goodbye. Experimenter 2's role was to record the data and ask some demographic questions. Experimenter 3's role was to ask the question about the lotteries. Experimenter 4's only role was to was to introduce themselves when asked to do so and wear a face mask when this was required by the randomisation.

Subjects were asked to arrive at a room within a particular time slot. Importantly, it was not possible to view inside the room without entering it; and once a subject had entered, the only people they could see were the experimenters inside the room. Before each subject entered the room, the number of the four experimenters in the room who were wearing a mask (and the allocation of masks to experimenters) had been randomised. Thus, there were five treatment groups, corresponding to: 0/4 masks, 1/4 masks, 2/4 masks, 3/4 masks, 4/4 masks. All four experimenters were seated in front of a table on which a box of face masks, hand sanitiser, and bag of checkers had been placed.

Once a subject arrived, the experiment proceeded in the following manner:

1. Experimenter 1 welcomed the participant in and asked all other experimenters to introduce themselves. The other three experimenters then did this by stating their name and subject of study.

2. Experimenter 2 asked the subject for their name, age, and academic division. They recorded these on a spreadsheet, along with their apparent gender and whether they had entered the room wearing a mask.

3. Experimenter 3 asked the subject the following question. '*As you may know, we have issued a fixed number of lottery tickets for an Amazon voucher. I am now going to give you two options to choose from. The first option is simply to get one lottery ticket for the voucher. The second option is a gamble between 2 and 0 lottery tickets. Specifically, if you take the second option, then you will take a checker from the bag in front of you. If you get a black checker — and there are six of these — then you will get two lottery tickets. However, if you get a white checker — and there are five of these — then you will not get any lottery tickets. So what do you choose — getting one lottery ticket for sure, or taking the gamble between 2 and 0 lottery tickets?*'

4. Person 1 thanked the participant for coming and told them that the experiment had concluded.

Occasionally, a subject asked if they should wear a face mask. In response to such questions, Experimenter 2 always replied: 'it's up to you'.

Experiment 2 (Zoom calls). As before, there were four experimenter roles, labelled 1 through 4. Experimenter 1's role was to admit subjects into the Zoom room and guide them through the experiment. Experimenter 2 was the data recorder, and Experimenter 3 double checked all data. Experimenter 4 pasted a link to the survey in the Zoom chat just before the subject was asked to leave the room.

Subjects were asked to join the Zoom call at a particular time slot. Before the subject joined the call, the number of experimenters with their camera on, and which experimenters had their camera on, had been randomised. Once a subject arrived in the Zoom waiting room, the experiment proceeded in the following manner:

1. Experimenter 1 thanked the subject for joining and asked if whether they could hear the audio. They stated their name, and said that the other experimenters would now introduce themselves.

2. The other three experimenters on the call now introduced themselves by stating their name.

3. Experimenter 1 asked the subject for their age. Experimenter 2 recorded this on a spreadsheet along with their apparent gender, and whether they had joined the call with their video camera on.[17]

4. Experimenter 1 then asked the subject the following question: '*If we were to hypothetically give you a £10 bonus payment, would you choose to share half of it with the next person on the call?*'

5. If the subject's camera had remained off throughout the call, Experimenter 1 asked them if there were any issues with their camera.

6. Experimenter 1 then thanked the subject for participating, asked them to click the survey link, and removed them from the room.

Some notes:

[17] In contrast to the first experiment, we did not require Experimenter 2 to ask the demographic questions due to the more rapid pace of data collection.

- All experimenters ensured that they did not have a Zoom profile photo; so when they turned their camera off, only the text of their name was visible.
- If a subject asked if they should turn their camera on, then Experimenter 1 told them that ‘it’s up to you’.
- If a subject turned their camera on when the host asked if they had issues with their camera, then we ignored this from a data recording point of view (see previous discussion).

E Explanations (online appendix)

In this section, we elaborate on the way in which we categorise subject explanations for 'switching behaviour' (recorded in the online survey). The categories are as follows:

1. 'Trying to avoid judgement'
   Elaboration: if you see many others wearing a mask, you might infer that these others want you to wear a mask. This in turn might induce you to wear a mask if you do not want to be negatively judged by the others.
   Example from dataset: 'Don't see the point in wearing a mask now, but if everyone else was then social conformity and not wanting to be the odd one out would mean I probably would.'

2. 'Trying to cater to others' preferences'
   Elaboration: If you see many others wearing a mask, you might infer that these others want you to wear a mask. This in turn might induce you to wear a mask if you want to altruistically cater to their preferences (e.g. to make them feel more comfortable).
   Comment: Observe that, like the explanation before, this explanation is based on learning about the preferences of others through their actions.
   Example from dataset: 'if i see someone wearing a mask it makes me think that they might be uncomfortable about the virus so if i had one on me i would wear it.'

3. 'Trying to follow rules'
   Elaboration: If you see many others wearing a mask, you might conclude that a (formal or informal) rule requires wearing a mask — and you might generally try to follow rules.
   Comment: In practice, it can be hard to distinguish this from the first explanation: individuals may follow informal rules to avoid judgement. However, we used this category since some participants mentioned rules without mentioning a fear of being judged.
   Example from dataset: 'If majority of people wearing mask, I assume there is written/unwritten rules regarding this, in that room that I am not aware of'

4. 'Reciprocity'
   Elaboration: if you see many others wearing a mask, you might infer that they are trying to protect you. As a result, you might want to protect them (as in Rabin (1993)).
   Example from dataset: 'I want to protect others, but if they aren't willing to protect me then I'm not willing to protect them'

5. ‘COVID risks’

   Elaboration: if you see many others wearing a mask, you might conclude that the COVID risk around you is high: for example, these people might be wearing a mask because they have COVID. Assuming that you want to avoid COVID, you might therefore choose to wear a mask.

   (Only) example from dataset: ‘I’d think if anyone were wearing a mask they probably have a good reason to, like being a close contact of a positive tester. Or if someone is just being particularly careful I would also think they have a good reason to and try to respect that.’

6. ‘Not answering question’

   Elaboration: Some subjects explained the various factors which determine whether they choose to wear a mask, but did not explain why their decision to wear a mask would vary with the number of others wearing a mask (the question we were interested in).

   Example from dataset: ‘It depends on the setting, and if I were carrying a mask at the time. If asked in advance I would always wear a mask, and would never want to make someone feel uncomfortable. However if the situation was relatively safe, I would not feel a need to wear a mask.’