Memory Fault Simulator for Static-Linked Faults

Overcoming Inaction: An Agent-Based Modelling Study of Social Interventions that Promote Systematic Pro-Environmental Change

Tabea Hoffmann[1,6], Mengbin Ye[2], Lorenzo Zino[3,4], Ming Cao[4], Ward Rauws[6], and Jan Willem Bolderdijk[1,5]

[1]Faculty of Economics and Business, University of Groningen, Groningen, The Netherlands

[2]Centre for Optimisation and Decision Science, Curtin University, Perth, Australia

[3]Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy

[4]Engineering and Technology Institute Groningen, University of Groningen, Groningen, The Netherlands

[5]Faculty of Economics and Business, University of Amsterdam, Amsterdam, The Netherlands

[6]Faculty of Spatial Sciences, University of Groningen, Groningen, The Netherlands

Author Note

The data and all materials are available on the OSF repository at `https://osf.io/w8pjk/`. Correspondence should be addressed to Tabea Hoffmann (t.hoffmann@rug.nl; Nettelbosje 2, 9747AE Groningen, NL).

Abstract

Even though many people have pro-environmental convictions, oftentimes they do not actually engage in pro-environmental behaviour. We hypothesise that behavioural change is hampered by a social feedback loop that reinforces the status quo: People routinely underestimate others' pro-environmental convictions, and when they expect that others care less, they are unlikely to show more pro-environmental behaviour themselves, which reinforces the general impression that people do not care. This leads to the question of how to effectively elicit a push from the current state to a state in which pro-environmental behaviour becomes more widespread. We examine this question with an agent-based model (ABM) which was parameterised using individual-level survey data collected in several Dutch neighbourhoods. We explore whether interventions that make people talk more about their convictions versus interventions that enhance the visibility of pro-environmental behaviour can trigger individuals to update their expectations and consequently tip the system into a more environmentally-friendly state. Our simulations suggest that enhancing the visibility of pro-environmental behaviour with an intervention may be most effective to motivate durable pro-environmental change. Motivating more talk on the topic only generates temporary effects in our simulations. These results can provide valuable guidance for empirical research on norm-based interventions and it may eventually inform the development of evidence-based policies that effectively encourage pro-environmental change.

Keywords: Pro-environmental Behaviour, Feedback Loop, Pluralistic Ignorance, Agent-Based Modelling

Declaration of Competing Interest

The authors declare that there are no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgements

We extend our gratitude to the dedicated students of the "Revitalising Neighbourhoods" course at the University of Groningen during the winter semester of 2020/2021 who contributed significantly to this study by collecting the data.

Funding

T. Hoffmann was supported by the University of Groningen's interdisciplinary Ph.D. fund. M. Ye was supported by the Western Australian Government via the Premier's Science Fellowship Program. The funders had no role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the article for publication.

1. Introduction

Mitigating climate change and accomplishing global environmental sustainability requires collective behaviour change (Lenton et al., 2022; Otto et al., 2020). However, while many people indicate pro-environmental convictions (i.e., they consider it important to protect the environment), the corresponding pro-environmental behaviour often falls short (Bolderdijk & Jans, 2021). The present paper investigates the role of social influence in encouraging pro-environmental change. We propose that pro-environmental behaviour is lagging because a feedback loop maintains the current status quo: People have misguided normative expectations, which means that they expect that others have lower pro-environmental convictions than they do themselves (Bergquist, 2020; T. Bouman, Steg, & Zawadzki, 2020; Hanel et al., 2018; Leviston & Uren, 2020; Mildenberger & Tingley, 2019; Sparkman, Geiger, & Weber, 2022). On an individual level, such misguided expectations lead to less pro-environmental behaviour as people tend to adjust their behaviour to match their expectations about what others do (i.e., other's behaviour) or what others think should be done (i.e., other's convictions; Bolderdijk & Cornelissen, 2022; Griskevicius, Tybur, & Van den Bergh, 2010; Lorenzoni, Nicholson-Cole, & Whitmarsh, 2007; Noppers, Keizer, Bolderdijk, & Steg, 2014; Salmivaara, Lombardini, & Lankoski, 2021). On a collective level, misguided expectations and the ensuing inactivity reinforce the normative expectation that others do not care about pro-environmental issues, which sustains the non-sustainable status quo (T. Bouman & Steg, 2022).

The question thus becomes how to effectively elicit a push from the current state to a state in which pro-environmental behaviour becomes widely adopted. Social complexity suggests that seemingly small effects, when considered over time, can accumulate into major qualitative changes. However, such changes are hard to study using survey or cross-sectional data because they cannot capture temporal processes and account for feedback loops, where changes create a ripple effect of further changes. Herein, agent-based modelling becomes a valuable tool, given its ability to capture how long-term societal-level change emerges from a sequence of small individual-level change. The present paper uses an agent-based model (ABM) calibrated with data from several Dutch neighbourhoods. We first demonstrate how the proposed feedback loop entrenches existing low levels of pro-environmental behaviour despite high levels of pro-environmental convictions. Second, we compare the effectiveness of two types of interventions that aim to correct misguided expectations to 'tip' the system to a state where pro-environmental behaviour and normative expectations reinforce each other in driving pro-environmental change. We expect that correcting normative expectations (i.e., expectations about whether others care about pro-environmental issues) will change the outcomes of the feedback loop, accelerating the spread of pro-environmental behaviour. Moreover, once pro-environmental behaviour has been broadly adopted, the feedback loop may now play a supportive role in maintaining this adoption. Specifically, we simulate two interventions to test whether they can correct misguided expectations about other's pro-environmental convictions; the first allows individuals to more easily share their convictions, and the second makes pro-environmental behaviour more visible. We then further investigate whether these interventions can lead to widespread adoption of pro-environmental behaviour, and whether such behaviour change remains permanent after the interventions are removed. It is crucial to acknowledge that our ABM is an abstraction of the real world. As such, the present paper represents an initial step in researching these interventions, establishing the groundwork for future empirical studies that delve into the underlying mechanisms contributing to the interventions' success and validate their effectiveness in real-world scenarios.

1.1 Social Misperceptions

Prior research has shown that people expect that others care less about environmental protection than they do themselves and also expect less engagement in pro-environmental behaviour by others as compared to themselves (see, for example, Bergquist, 2020; Sparkman et al., 2022). However, large-scale surveys in the US, China, the EU, and Australia show that most people in these countries accept the anthropogenic causes for climate change, are worried about it, and willing to take measures (Drews, Savin, Van Den Bergh, & Villamayor-Tomás, 2022; Leviston, Walker, & Morwinski, 2013; Mildenberger & Tingley, 2019). It thus seems that there is so-called pluralistic ignorance — a systematic and shared misperception — surrounding the norms on action towards environmental sustainability (Prentice & Miller, 1996).

The misguided expectation that others are not really pro-environmental might partially stem from a lack of direct exchange on the topic. Prior research shows that environmental issues are rarely discussed, even though people claim they care about them (Geiger & Swim, 2016; Maibach, Leiserowitz, Rosenthal, Roser-Renouf, & Cutler, 2016). People are reluctant to steer a conversation towards the topic, let alone advise others to engage in pro-environmental behaviour or confront others' environmentally harmful behaviour to avoid conflict (Jones & Niemiec, 2023; Steentjes, Kurz, Barreto, & Morton, 2017). Given that environmental issues are rarely discussed in conversations, people have to rely on indirect sources, such as encountering information in public discourse, to form an aggregate expectation about others' convictions, meaning they are unlikely to have direct insights into the actual convictions of others.

In other words, people's normative expectations — their expectation of what others approve of (Bicchieri, 2005) — are often based on more indirect clues such as observing others (cf. Kashima, Wilson, Lusher, Pearson, & Pearson, 2013).[1] However, people rarely observe pro-environmental behaviour because it is not the common behaviour in many situations (e.g., commuting by bike versus by car) and not all pro-environmental behaviours are visible (e.g., retrofitting one's home) or unambiguously pro-environmental (e.g., maintaining a green garden for aesthetic reasons or to foster biodiversity). Lacking examples of others' pro-environmental behaviours, people are likely to expect that others do not care much, as a lack of behaviour is often mistaken as a lack of interest (cf. Epley & Caruso, 2008; Ross, 1977). Thus, while not talking about environmental issues leads to individuals having little

[1] Normative and empirical expectations can be seen as equivalent to 'perceived injunctive norms' and 'perceived descriptive norms' (Constantino et al., 2022; Farrow, Grolleau, & Ibanez, 2017). They are sometimes also called 'second-order normative beliefs´ (Drews et al., 2022).

insight into others' convictions, not observing pro-environmental behaviour may lead to the drawing of wrong conclusions about why others do what they do.

1.2 Feedback Loop

Misguided expectations about what others approve of can hinder behaviour change, as expectations that others do not care about pro-environmental behaviour can discourage individuals from engaging in it (e.g., Bolderdijk & Cornelissen, 2022; Lorenzoni et al., 2007; Noppers et al., 2014; Sparkman et al., 2022). Also, expectations based on misconceptions (i.e., underestimating others' pro-environmental convictions) can lead to further self-silencing (i.e., avoidance of the topic in conversations; Geiger & Swim, 2016), reinforcing misperceptions.

Taken together, there appears to be a feedback loop in operation where the social expectations that partly determine behaviour are formed by observing what others are doing (T. Bouman & Steg, 2022). Currently, this feedback loop works against pro-environmental change because the lack of observable pro-environmental behaviour stabilises the expectation that others do not care about environmental protection which in turn hampers pro-environmental behaviour. This feedback loop thus keeps the population locked into a steady state where pro-environmental behaviour is uncommon. Spontaneously breaking out of this state is unlikely since misguided expectations are unlikely to be corrected as people do not talk about the topic.

However, in a like manner, the proposed feedback loop has the potential to drive pro-environmental change. Changed expectations of what others approve of can produce abrupt behavioural changes, which makes interventions that effectively change people's expectations a potentially powerful policy tool (Lenton et al., 2022; Nolan, 2021; Nyborg et al., 2016; Sharpe & Lenton, 2021). We identify two ways in which an intervention may correct misguided expectations: encouraging the communication of pro-environmental convictions, for example, by stimulating talk on the topic (cf. Geiger & Swim, 2016), and making pro-environmental behaviour more visible with, for example, a visual indicator in the environment such as a sticker or a sign (cf. Hamann, Reese, Seewald, & Loeschinger, 2015). We propose that both interventions have the potential to elicit a push from the current state to a state in which pro-environmental behaviour becomes widespread (cf. Nolan, 2021; Wolske, Gillingham, & Schultz, 2020). Encouraging more conversations on the topic can clear up misperceptions and helps individuals to rely less on assumptions made from indirect cues, such as observing others' behaviour. Making already occurring pro-environmental behaviour more visible can correct misperceptions by making invisible (e.g., saving resources) or ambiguous (e.g., maintaining a biodiverse garden) pro-environmental behaviours visible. Enhancing the visibility of pro-environmental behaviour can also selectively highlight already visual pro-environmental behaviour by making it more salient than other behaviour (Cialdini, Reno, & Kallgren, 1990).

1.3 Modelling the Feedback Loop

This paper uses an agent-based model (ABM) as the main tool to study the effectiveness of the proposed interventions. ABMs are computational models that allow exploring how repeated individual-level interactions over time can lead to downstream societal-level phenomena, such as collective change (Bonabeau, 2002). ABMs have three components: the agents, the environment, and interactions. Agents represent individuals with changing states, such as behaviour and convictions, with all agents together making up the system's overall state. The environment is an artificial representation of the setting in which the simulated individuals interact. The interactions represent the connections between the agents that allow them to share and learn of each other's states; these interactions underpin the social influence that can cause agent states to evolve. The interactions also represent the interaction of an agent with their environment. The agents' states, interactions with each other, and environment are defined in the model through several parameters. For instance, some parameters govern the agent's attributes while others define the rules for the agent's behaviour (e.g., parameters that define when agents change their behaviour) or their interaction. Standard practice with ABMs involves fixing some parameters across all simulations (e.g., those estimated from empirical data), while other parameters are varied across simulations. Varying parameters and exploring their combinations is a fundamental aspect of agent-based modelling, as it allows one to understand the interplay between different parameter values in affecting the simulation outcomes, including collective change and emergent phenomena. In essence, an ABM is like a box of bricks that allows experimenting with and evaluating ideas using a computational approach. This helps to gain insight into complex systems and make informed decisions or predictions. Carrying out the explorations within the ABM is thus a preparatory step for designing better-informed real-world interventions.

The present paper contributes to the existing literature by substantiating theoretical deliberations on the role of social influence in effecting pro-environmental change with a computational investigation (cf. T. Bouman & Steg, 2022; Nolan, 2021; Nyborg et al., 2016). Survey or cross-sectional data fall short in assessing the proposed dynamics and extensive experimentation would be costly and time-consuming if carried out in the real world. Based on the theoretical background above and the collected data, we therefore propose an ABM that considers the co-evolution of normative expectations and behaviour. The ABM is based on established paradigms, including activity-driven networks to model time-varying agent interactions (Perra, Gonçalves, Pastor-Satorras, & Vespignani, 2012), threshold models to describe how agents choose their behaviour (Granovetter, 1978), and weighted averaging belief formation to capture the evolution of normative expectations (Proskurnikov & Tempo, 2017). The data inform which constructs we include in the ABM and how we can determine a subset of the ABM's parameters. By adjusting other specific model parameters, we simulate different interventions, tailoring parameter changes according to the nature of each intervention. We are specifically interested in finding which intervention method (stimulating talk or making behaviours visible) may be most effective at producing a tipping point — a point at which changes on the individual level cause the system of agents to undergo a significant and abrupt transition from the current system state to a different system state (ideally a state where many more agents adopt sustainable behaviour). We pay special attention to whether such interventions can be expected to produce a lasting change in behaviour rather than a temporary one. By exploring opportunities for a tipping point, we seek to investigate the potential of behavioural interventions to shape environmentally sustainable futures.

2. Method

In the following, we describe the data that we collected to calibrate the ABM — providing information on the sample as well as the materials that we used. Detailed information on the data collection procedure are presented in Appendix A. We conclude the description of the survey data by reporting the results. We were particularly interested in whether our data concords with previous research on the lack of talk on environmental issues, the presence of pluralistic ignorance, and whether expectations of others' convictions are associated with people's own behaviour. Afterwards, we describe the agent-based model definition, providing information on the different states of the agents, the social network formation process and the dynamics for updating the different states of agents.

2.1 Calibration Data

We surveyed eight different neighbourhoods in the city of Groningen, The Netherlands. We opted for data on a neighbourhood scale as prior research indicates that the size and quality of a neighbourhood's social network may provide enabling conditions for triggering tipping points in the adoption of pro-environmental behaviours (and technologies) (e.g., Bollinger & Gillingham, 2012). In each of the eight neighbourhoods one of four pro-environmental behaviours — avoiding car use, saving water and energy, installing solar panels, or urban greening — was surveyed. This means that in each neighbourhood the measures of convictions, social expectations and behaviour were all formulated in regards to one of the four behaviours. As noted in Appendix A, we excluded the data on resource conservation ($n = 73$) due to data quality concerns. This means that the final data set stems from six neighbourhoods.

2.1.1 Sample. We conducted a convenience sampling approach by going door-to-door, resulting in 507 participants for our study. However, we excluded several participants due to missing data and concerns regarding data quality. All exclusion criteria are listed in detail in Appendix A. Our final sample consisted of 258 residents (108 women, 112 men, 2 other, 3 prefer not to say, 33 NA) with a mean age of 45.67 years (SD = 16.09, 33 NA). Participants indicated an average of 0.51 family members ($Min = 0, Max = 10$), 2.69 friends ($Min = 0, Max = 25$) and 10.39 acquaintances ($Min = 0, Max = 75$) in their neighbourhood. On average, participants lived 121.55 months in their neighbourhood (SD = 117.33, 34 NA).

The number of private households in the neighbourhoods under analysis varies between 400 and 3,600 (CBS, 2022). The neighbourhoods have a range of 30% to 100% of single-family homes, with over 90% of the houses built after the year 2000 in four of the neighbourhoods (AlleCijfers, 2023). On average, 61.3% of homes are owned by their occupants, and four of the neighbourhoods have a higher socioeconomic status than the Dutch average (AlleCijfers, 2023; CBS, 2022). Compared to other regions in the Netherlands, Groningen has a comparatively high population density, with an average of 3,373 addresses per square kilometre (AlleCijfers, 2022).

2.1.2 Materials. To survey residents, we built an online questionnaire with the survey platform Maptionnaire.[2] A pen-and-paper questionnaire was provided for respondents with insufficient computer literacy. All questionnaires can be found in the online repository associated with this paper (`https://osf.io/w8pjk/`).

Convictions. People's own pro-environmental convictions were measured with items such as "I think that creating space for plants and greenery by reducing spaces that are covered with concrete is a good thing to do." or "I think solar panels are a good thing to have." Respondents were asked to rate how much they agreed with these statements on a scale from 1 (strongly disagree) to 5 (strongly agree).

Normative expectations. Normative expectations were measured with items such as "I expect that people in this neighbourhood approve of the installation and use

[2] https://maptionnaire.com/

of solar panels." or "I expect that people in this neighbourhood think that more people should minimize their car use." (1 = "strongly disagree" to 5 = "strongly agree").

Empirical expectations. Empirical expectations were measured with items such as "How many neighbours use public transportation or ride a bike to travel to nearby areas (around 5 km) instead of using the car?" or "How many neighbours have plants on the balcony or on the window railing?" We asked respondents to indicate their empirical expectations on a scale from 1 (none) to 5 (all).

Pro-environmental behaviour. People's own pro-environmental behaviour was measured with items such as "I have requested an estimate on having solar panels installed." or "I try to minimise my car use as much as possible." We asked respondents to rate how much these statements were true of them on a scale from 1 (very untrue of me) to 7 (very true of me).

Conversations. Finally, we were interested in how often people talk about environmental issues. Conversations between neighbours were measured with one item: "Do you talk with the neighbours about problems related to the environment?" (1 = "very untrue of me" to 7 = "very true of me"). Note that while the other questions were behaviour-specific, this question was the same across all questionnaire versions.

2.1.3 Survey Results. Pre-processing of the data was performed in R (version 4.2.2 R Core Team, 2022). Analyses were conducted using JASP (version 0.16 JASP Team, 2023). Missing data was imputed using the MICE package in R (van Buuren & Groothuis-Oudshoorn, 2011). We only imputed data for the convictions, expectations and behaviour measures provided that respondents had answered at least one item for the respective measures. All data and analysis scripts are available at the online repository associated with this paper (`https://osf.io/w8pjk/`).

Do neighbours talk about environmental issues? The overall data shows that neighbours hardly talk with each other about problems related to the environment ($M = 2.25, SD = 0.87, NA = 34$). Table 1 shows that this pattern was similar across all neighbourhoods. This pattern is unlikely caused by a lack of connections between residents (i.e., that respondents do not talk to their neighbours in general) given that many respondents indicated to have acquaintances, friends and family in their neighbourhood.

Table 1

Descriptive statistics for the question "Do you talk with the neighbours about problems related to the environment?" for each neighbourhood.

	Talk about Environmental Issues					
	B	DL	O	P	R	VS
M	1.95	2.5	2.13	2.59	2.63	2.13
SD	1	0.53	0.91	0.62	0.79	0.81
Minimum	1	2	1	2	1	1
Maximum	4	3	4	4	4	3

Note. Respondents answered the question on a scale from 1 (very untrue of me) to 7 (very true of me). Each letter denotes one neighbourhood: B = Bloemenbuurt; DL = De Linie; O = Oosterpoort; P = Piccardthof; R = Reitdiep; VS = Van Starkenborgh.

Is there pluralistic ignorance? The distributions of convictions ratings and normative expectations ratings — plotted as a one-sided violin plot in the most right panel of figure 1 — indicate that respondents generally underestimated their neighbours' pro-environmental convictions: While most respondents indicated high pro-environmental convictions ($M = 4.015$), they indicated lower normative expectations ($M = 3.206$). Their expectations did not coincide with the convictions of their neighbours which are reflected in the distribution of pro-environmental convictions. Using a Bayesian Wilcoxon signed-rank test, we found a Bayes factor of $BF_{-0} = 5.515e + 8$ which means that the observed data are five hundred million times more likely under the hypothesis H_{-} that respondents underestimated their neighbours' convictions than under H_0.[3] A Bayes factor of this size is considered extreme evidence (van Doorn et al., 2021). For this analysis, we collapsed the data across the different behaviours; the results of the paired samples t-test for each of the behaviours

[3] We specified a folded Cauchy prior with $r = 0.707$ under the alternative hypothesis H_{-}

separately can be found in Appendix A.

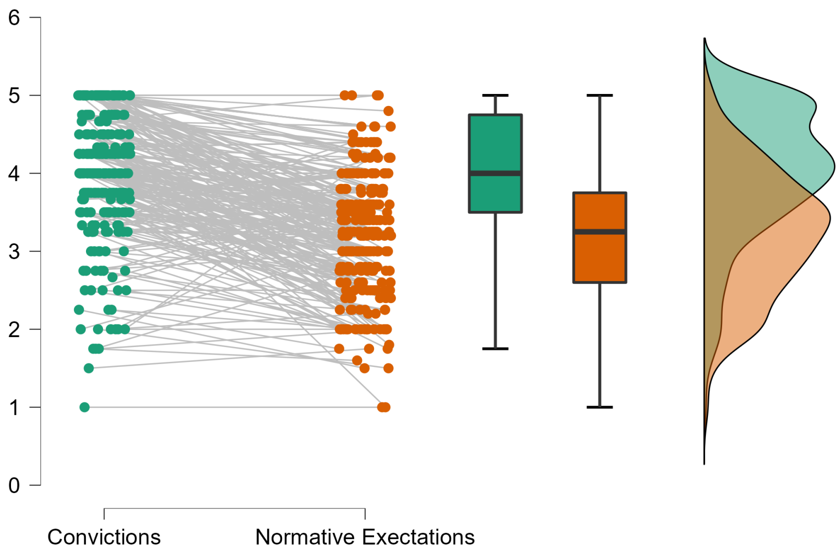


Figure 1. The plot visualises respondent's answers to the convictions and normative expectations measures with a cloud of points, a box plot and a one-sided violin plot. The whiskers of the box plot illustrate the lowest and highest data points while the box shows the first quartile, the median and the third quartile of the data. Convictions are plotted in green and normative expectations are plotted in orange.

Is there a linear relationship between social expectations and pro-environmental behaviour? We calculated correlations between normative as well as empirical expectations and pro-environmental behaviour for each of the behaviours separately. Further, we calculated the correlation between respondent's (behaviour-specific) convictions and their pro-environmental behaviour. All Kendall's τ coefficients are reported in Table 2. Across the three data sets, the correlations are variable, but convictions emerge as a consistently correlated with pro-environmental behaviour. Normative expectations seem to be associated with pro-environmental behaviour for both the data on solar panels and for the data on travel behaviour. Empirical expectations only show a noteworthy correlation with pro-environmental behaviour for the data on solar panels. In the context of greening the correlations between social expectation and behaviour are generally very weak. For all three behaviours, there seems to be a consistent, moderately strong relationship between individuals' normative and empirical expectations. We report the results of a regression analysis in Appendix A.

Table 2

Kendall's τ correlations between the mean scores of all key variables for each behaviour separately.

		M	SD	1	2	3
Solar Panels ($N = 44$)						
1.	Behaviour	3.455	1.150			
2.	Normative Expectations	4.136	0.458	0.264		
3.	Empirical Expectations	2.972	0.532	0.298	0.255	
4.	Convictions	4.205	0.660	0.430	0.443	0.343
Greening ($N = 106$)						
1.	Behaviour	4.097	0.987			
2.	Normative Expectations	3.183	0.623	0.036		
3.	Empirical Expectations	2.245	0.572	−0.026	0.320	
4.	Convictions	4.245	0.769	0.371	0.309	0.064
Travel Behaviour ($N = 108$)						
1.	Behaviour	4.507	1.154			
2.	Normative Expectations	2.850	0.726	0.154		
3.	Empirical Expectations	2.574	0.494	0.053	0.379	
4.	Convictions	3.711	0.847	0.334	0.310	0.235

Note. M = mean; SD = standard deviation.

In sum, in line with prior research, our data suggests that people are biased to underestimate how much others care about the environment, and this underestimation may hamper more pro-environmental behaviour (cf. Sparkman et al., 2022). In other words, the necessary ingredients for a self-sustaining feedback loop that entrenches a low level of pro-environmental behaviour are in place (cf. T. Bouman & Steg, 2022). Our data set, however, provides only a static picture of all variables, which impedes testing how the proposed dynamic plays out over time. To conduct a computational analysis of the proposed feedback loop, we used the data to calibrate an ABM which will be described below.

2.2 Model Definition

A schematic representation of the ABM is depicted in Figure 2. In this section, we begin with describing the different states (displayed in the yellow box in Figure 2) that are represented in the model. We proceed with explaining the network structure (displayed in the grey box in Figure 2). Lastly, we detail the dynamics for updating the different states. A detailed description of the simulations and their results follows in the next section.

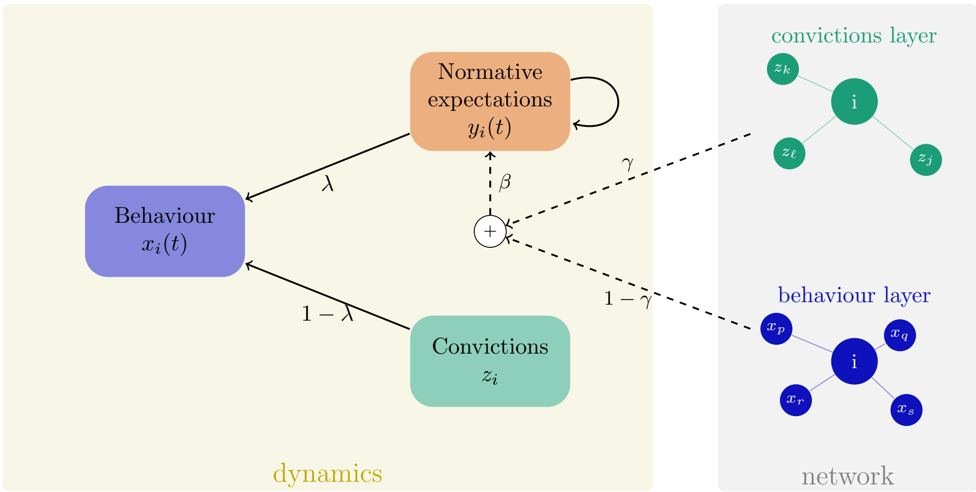


Figure 2. Schematic representation of the ABM. Solid arrows represent direct influences between the variable states of the same individual (e.g., the current convictions of an individual influences her own future behaviour); dashed arrows represent social influences from the two-layer network (e.g., the behaviours that an individual observes in the network influences the individual's normative expectation).

Our model considers a population of $n \geq 2$ individuals (agents), indexed by the set $\mathcal{V} = \{1, 2, 3, \ldots, n\}$. Individuals interact with one another, and may update their states, at discrete time-steps $t = 0, 1, 2, \ldots$, where a time-step represents a fixed interval of time after which individuals may revise their normative expectations and behaviour (e.g., a day or a week). Associated with each individual i are three values:

- The variable state $x_i(t) \in \{0, 1\}$ is the *behaviour* of individual i at time t. We assume the state takes on a binary value, with 0 and 1 representing an individual not adopting and adopting the pro-environmental behaviour, respectively.
- The variable state $y_i(t) \in [0, 1]$ is a real-valued scalar that represents individual i's *normative expectations* with respect to the pro-environmental behaviour considered. That is, $y_i = 0$ and $y_i = 1$ represents individual i expecting that there is complete social approval for not adopting and adopting the pro-environmental behaviour, respectively.
- The scalar value $z_i \in [0, 1]$ is the *conviction* of individual i with respect to the considered behaviour. Thus, $z_i = 0$ and $z_i = 1$ represent individual i having complete conviction for not adopting and adopting the pro-environmental behaviour, respectively. Based on prior research showing that convictions evolve at a relatively slow pace, we assume here that they remain constant for the entire duration of the simulations (cf. Rozin, 1999).

Hence, each individual $i \in \mathcal{V}$ is characterised by the triple $(x_i(t), y_i(t), z_i)$, representing their behaviour, normative expectations, and convictions, respectively. Note that we did not include empirical expectations in the model. We did so because the simulations focus on the question of how pluralistic ignorance (i.e., the discrepancy between a individual's convictions and the expected convictions of others, which we call normative expectations) may hamper pro-environmental change and how interventions may resolve this dynamic. Beyond these theoretical considerations, our survey data suggests that the association between empirical expectations and behaviour may be negligible. The precise updating rules for $y_i(t)$ and $x_i(t)$ will be presented in the sequel, but we first define the social network through which individuals interact and obtain information about each other.

2.2.1 Social Network. The interactions between agents are represented by means of a two-layer time-varying network. One layer mirrors the observations people may make, while the other layer mirrors the conversations that they may have with neighbours. Thus, links in the *behaviour layer* could represent an individual walking in the neighbourhood and observing newly installed solar panels or seeing a new garden, or it may represent driving to work and seeing someone else commuting by bike. Similarly, links in the *convictions layer* could represent having a conversation with a neighbour on, for example, composting and recycling, or seeing a social media post that deals with environmental sustainability by the other person. As such, the two layers are *abstract* representations of potentially many different real-world interactions.

Mathematically, the two-layer time-varying network is denoted by $\mathcal{G}(t) = (\mathcal{V}, \mathcal{E}^b(t), \mathcal{E}^c(t))$ which describes that the individuals $\mathcal{V} = \{1, \ldots, n\}$ interact on two distinct layers: they observe the behaviour of others on the *behaviour layer*, captured by the set of time-varying, directed links $\mathcal{E}^b(t)$, and they talk and share information about their convictions on the *convictions layer*, which is represented by the ordered set of links $\mathcal{E}^c(t)$. We then use an activity-driven network (ADN) to simulate the instantaneous, time-varying, and occasional interactions between individuals (Perra et al., 2012). At each time-step, new links are being created according to a random, stochastic procedure. The exact network formation process is described in more detail in Appendix B. Roughly speaking, each individual i establishes m_b links with m_b other individuals on the behaviour layer and m_c links with m_c other individuals on the convictions layer. The set of individuals that individual i created links to at time t on the behaviour and convictions layers are denoted by $\mathcal{N}_i^b(t)$ and $\mathcal{N}_i^c(t)$, respectively. Figure 3 illustrates a realisation of two time-steps of the temporal network formation process.

2.2.2 Dynamics for Updating Normative Expectations and Behaviour. Based on prior research and our survey data, we set up the ABM in a way such that agents update their normative expectations based on what they see others doing and based on what they hear others are saying; and they change their behaviour depending

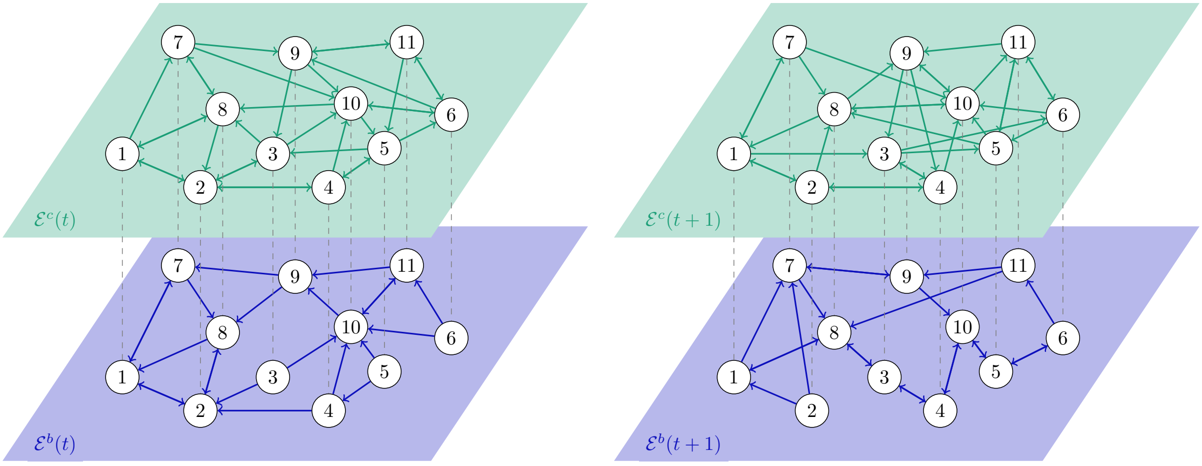


Figure 3. Illustration of two consecutive time-steps of the temporal network formation with 11 agents, $m_b = 2$ (links with others on the behaviour layer, plotted in blue) and $m_c = 3$ (links with others on the convictions layer, plotted in green). For example, at time t, agent 3 links to agents 8, 2 and 10 on the convictions layer. At time $t + 1$, the same agent links to agents 4, 5, and 6 on the convictions layer.

on their own convictions and normative expectations. For example, seeing a neighbour installing solar panels likely leads to the expectation that this neighbour cares about using clean, renewable energy sources. Such a change in normative expectations, combined with a person's own conviction that using renewable energy sources is the right thing to do, may bring the individual to also install solar panels. Formally, we state that at each time-step t, and after links have been created in the two network layers, each individual i will revise their normative expectations $y_i(t)$ and behaviour $x_i(t)$. Both updating processes will be described in more detail below.

To update their normative expectations, agents consider their current expectations and the information they gather from the two network layers, and then average them together. In addition, there is a parameter that takes into account that expectations may be more influenced by what they see others doing rather than what they hear others saying, or vice versa. For instance, seeing a neighbour setting up a compost heap may lead to the expectation that the neighbour strongly approves of composting while only hearing them talk about it may have a weaker influence on normative expectations. Formally, we write

$$y_i(t+1) = (1-\beta_i)y_i(t) + \beta_i \left[\gamma_i \frac{1}{m_c} \sum_{j \in \mathcal{N}_i^c(t)} z_j + (1-\gamma_i) \frac{1}{m_b} \sum_{j \in \mathcal{N}_i^b(t)} x_j(t) \right]. \tag{1}$$

The first part of Eq. (1) shows the individual i's current normative expectations $y_i(t)$. The second part describes the information that individual i receives from the two-layer network. This part of the equation has two components. The first component $\frac{1}{m_c} \sum_{j \in \mathcal{N}_i^c(t)} z_j$ denotes the average convictions of other agents, which individual i learns from the convictions layer. The second component $\frac{1}{m_b} \sum_{j \in \mathcal{N}_i^b(t)} x_j(t)$ denotes the average behaviour displayed by other agents, which individual i observes through the behaviour layer. The parameter $\gamma_i \in [0,1]$ weights these two components. Smaller values for γ_i indicate that individual i gives more relative importance to what they see other agents doing. The parameter $\beta_i \in [0,1]$ determines how quickly individual i changes their expectations based on the information they receive from the two-layer network. A higher β means a faster rate of change, while a lower β means a slower rate of change.[4]

Agents change their behaviour depending on their own convictions and normative expectations. To reflect the idea that people have their own pro-environmental convictions, but also want to fit in with everyone else in the neighbourhood, a parameter is included in the model that governs how people balance their own convictions with their normative expectations. We also include a parameter that allows us to differentiate behaviours depending on the difficulty of adoption. For instance, installing solar panels requires a significant financial investment that not everybody can afford. Hence, the adoption of solar panels is, on average, characterised by a larger threshold than, for example, greening the garden. Mathematically, we formalise such a

[4] Eq. (1) arises from classical weight-averaging models for opinion formation, such as the French–DeGroot model (French Jr, 1956; Proskurnikov & Tempo, 2017). Weighted averaging has been commonly employed to model social influence through networked interactions, and describe how individuals try to resolve discrepancies between their opinions or attitudes (Proskurnikov & Tempo, 2017).

rational decision by means of the following threshold rule:

$$\mathcal{D}_i(y_i(t), z_i) = \begin{cases} 1, & \text{if } (1-\lambda_i)z_i + \lambda_i y_i(t) \geq \alpha_i\,, \\ 0, & \text{otherwise}\,. \end{cases} \tag{2}$$

where $\lambda_i \in [0, 1]$ is the scalar *weighting parameter* and $\alpha_i \in [0, 1]$ is the *threshold parameter*. The quantity $(1 - \lambda_i)z_i + \lambda_i y_i(t)$ represents a weighted average of individual i's convictions to adopt the pro-environmental behaviour, z_i, and individual i's normative expectation $y_i(t)$. As noted above, λ_i captures individual i's level of conformity to their normative expectations. When the weighted average of convictions and normative expectations exceeds $\alpha_i \in [0, 1]$, individual i will adopt the pro-environmental behaviour (cf. Goldstone & Janssen, 2005; Granovetter, 1978). A value of α_i greater or less than 0.5 implies the pro-environmental behaviour has a higher or lower threshold before being adopted.

Further, we incorporate two additional mechanisms in the ABM to capture the bounded rationality of human decisions and the fact that behaviour change may occur only occasionally, in contrast to the continuous updating of normative expectations (which is caused by a constant exposure to information about others; for example, by seeing them or by talking to them). Specifically, the dynamics for updating behaviour are defined as follows. At each time-step t, each individual i updates their behaviour according to the following probabilistic rule with four outcomes:

$$x_i(t+1) = \begin{cases} x_i(t) & \text{with probability } 1-\rho, \\ \mathcal{D}_i(y_i(t), z_i & \text{with probability } \rho(1-\varepsilon_i), \\ 0 & \text{with probability } \rho\varepsilon_i/2, \\ 1 & \text{with probability } \rho\varepsilon_i/2, \end{cases} \tag{3}$$

where $\varepsilon_i \in [0, 1]$ is a *randomness parameter* and $\rho \in (0, 1]$ is a *frequency parameter*. The *frequency parameter* ρ thus captures the fact individuals do not always revise their behaviour at every possible opportunity, but rather with a certain frequency, as opposed to the continuously evolving formation process of normative expectations. The *randomness parameter* ε_i allows for including some level of bounded rationality in the behaviour update process, capturing the fact that individuals can occasionally be irrational, or show behaviour that deviates from their convictions (cf. Peyton Young, 1993). Briefly, at each time-step t, individual i decides to consider revising their behaviour with probability ρ; otherwise, with probability $1-\rho$, they maintain their current behaviour (so $x_i(t+1) = x_i(t)$)). Thus, the higher is ρ, the more often the individual will consider revising their behaviour. If i considers to revise, then, with probability $1-\varepsilon_i$, the individual follows the rational decision defined in Eq. (2), so $x_i(t+1) = \mathcal{D}_i(y_i(t), z_i)$. Otherwise, with probability ε_i, they choose their next behaviour uniformly at random among the two options 0 or 1. Here, we assume that the adoption of pro-environmental behaviours is reversible, i.e., an individual is allowed to change behaviour $x_i(t)$ multiple times during a simulation of the ABM. Note that it is also possible to adjust the model to capture one-shot behaviours that cannot be reversed, similar to a linear threshold model presented by Granovetter (1978). Table 3 provides a summary of all the states and parameters of the ABM along with their meanings.

3. Results

We now use the ABM to study collective pro-environmental change. Specifically, we perform Monte Carlo simulations of the ABM to test different scenarios and intervention policies, using the survey data to set the initial values for agents' behaviour, normative expectations, and convictions. The ABM code is written using MATLAB R2023a and the MATLAB's Statistics and Machine Learning toolbox (v12.5 or newer), and it is openly available at the online repository associated with this paper (`https://osf.io/w8pjk/`).

3.1 Simulation setup

We consider a population of $n = 1,000$ agents, which is close to the average number of private households of the surveyed neighbourhoods. For each individual $i \in \mathcal{V}$, we sampled $x_i(0)$, $y_i(0)$, and z_i from the joint density function, estimated from

Table 3

Overview of the symbols and their meaning for all states and parameters that are used in the ABM. The third column provides information on how the parameter (or the initial value for the variables) are set in the simulations: C = calibrated from survey data; G = based on reasonable guess (with additional simulations to test robustness); V = varied across the simulation scenarios to investigate their impact.

symbol	meaning	value
n	number of individuals	C
$x_i(t)$	behaviour of individual i at time t	C
$y_i(t)$	normative expectations of individual i at time t	C
z_i	conviction of individual i	C
m_b	number of links established on the behaviour layer per time-step	G
m_c	number of links established on the conviction layer per time-step	G
β_i	rate of change of individual i's normative expectations	G
γ_i	relative weight of the information gathered on the conviction layer in the normative expectations update for individual i	V
λ_i	individual i's level of conformity to their normative expectations	V
α_i	individual i's perceived difficulty of adoption	V
ε_i	individual i's randomness in behaviour update	G
ρ	frequency of re-consideration of the behaviour	G

the survey data.[5] Due to the stochastic nature of the model, i.e., the randomness in Eq. (3) and network generation process, simulation outcomes under the same parameter settings can differ. We thus perform Monte Carlo simulations over 100 independent runs for each scenario, re-sampling the initial conditions at each run, and each result of our study will be illustrated in terms of the average outcome of the 100 independent runs and their dispersion.

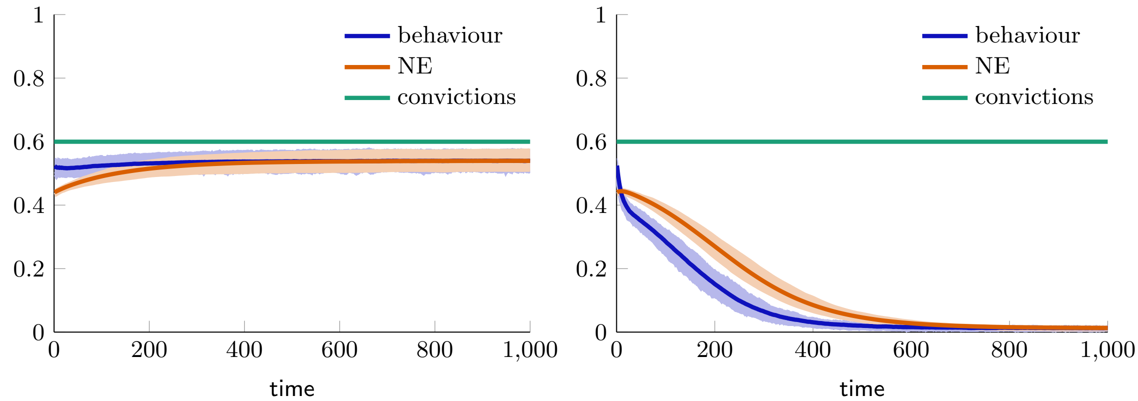
[5] The empirically reported measures were normalised to be between 0 and 1, to match the scaling of the states in the ABM.

The other parameters of the ABM, which cannot be directly inferred or informed by the data, are set within a known or reasonable range. In particular, an agent parameter that is common to all simulations and across all agents is $\beta_i = 0.01$ (the rate of change of agent's normative expectations). When the convictions layer and behaviour layer are not suppressed, we fix $m_c = 10$ (the number of links on the convictions layer) and $m_b = 10$ (the number of links on the behaviour layer), respectively. Moreover, we set the two parameters regulating the behavioural update to $\varepsilon_i = 0.01$ for all agents, and $\rho = 0.1$. Additional simulations reported in the Supplementary Material demonstrate that our findings are robust with respect to the values of these parameters and heterogeneity across the population. The remaining parameters (λ_i, α_i, and γ_i) will be varied across the different simulation scenarios in order to investigate their impact on long term, population-level, phenomena.

We fixed an initial simulation window of $T = 1000$ time-steps when there are no interventions, so that the agents are given sufficient time to reach a steady state. This allowed us to test whether the system, given the current lack of talk and the fact that people underestimate others' pro-environmental convictions, indeed is likely to stay 'locked' in an environmentally-unfriendly state. Subsequently, we simulate the introduction of an intervention (encouraging people to talk or making pro-environmental behaviour more visible) at $t = 1000$ time-steps, and its removal at $t = 2000$ time-steps. Repeated simulations indicate that 1000 time-steps are sufficient for all agents to reach a stable state and changing the time-steps does not change the results.

3.2 Development of Pro-Environmental Behaviour Without an Intervention

In our first set of simulations, we use the ABM to examine what would happen if we take the current state of the population (as obtained from the survey data) as the initial conditions for the model, with no interventions being introduced. Since the survey data suggested that individuals rarely talk about environmental issues, we initially set $\gamma_i = 0$ for all i to suppress the convictions layer (i.e., to simulate that agents do not learn about others' view on environmental sustainability). We fix the difficulty of the behaviour $\alpha_i = 0.56$ for all individuals. This means that adopting the pro-environmental behaviour is more difficult than not doing so, although it is not all too challenging using these settings. We examine the impact of the parameter λ_i — the weight given by individual i to their normative expectations $y_i(t)$ when deciding how to behave. Specifically, we tested two scenarios by setting $\lambda_i = 0.1$ and $\lambda_i = 0.4$ for all i to compare the differences between normative expectations having a weak or strong influence on pro-environmental behaviour, respectively.



(a) $\lambda = 0.1$ (b) $\lambda = 0.4$

Figure 4. Results of the simulations in the scenario with no interventions. Solid lines represent the Monte Carlo estimation of the average behaviour $\hat{x}(t) = \frac{1}{n}\sum_{i=1}^{n} x_i(t)$ (blue), the average normative expectations $\hat{y}(t) = \frac{1}{n}\sum_{i=1}^{n} y_i(t)$ (orange) and the average convictions $\hat{z} = \frac{1}{n}\sum_{i=1}^{n} z_i$ (green) of the population over time. Note that convictions were set to be time-invariant. The shaded areas represent the envelope in which 90% of the simulations lie.

The results are presented in Fig. 4 which shows that when individuals only observe others' behaviour without talking about their convictions, the best-case scenario is a maintenance of the status quo, while the worst-case scenario is the abandonment of previous pro-environmental behaviour. First, for $\lambda_i = 0.1$, we see in Fig. 4a that the average normative expectation $\hat{y}(t)$ converges to the average behaviour $\hat{x}(t)$. However, pluralistic ignorance is still present, since normative expectations differ from convictions. Collective pro-environmental change can also not be observed as the average behaviour

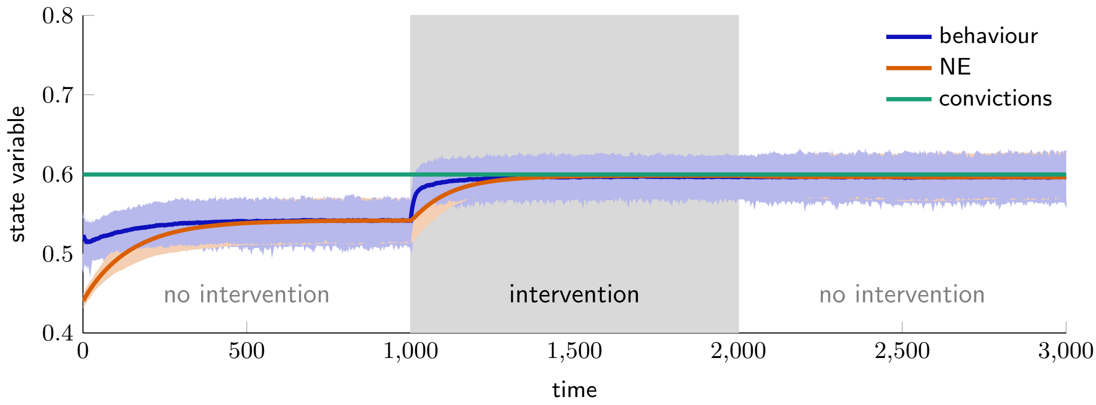
$\hat{x}(t)$ does not change over time. If we increase the influence of normative expectations on decision-making to the extremely high value of $\lambda_i = 0.4$, we see in Fig. 4b that most people have pro-environmental convictions but no one realises this as the level of pro-environmental behaviour drops (cf. Centola, Willer, & Macy, 2005). More specifically, although individuals in the population on average have convictions that support pro-environmental behaviour ($\hat{z} = 0.6$), the average behaviour $\hat{x}(t)$ drops close to 0, which indicates that the vast majority of individuals are not adopting pro-environmental behaviour. Both scenarios arise because people base their normative expectations solely on the observation that others are not adopting pro-environmental behaviour, and accordingly, perceive that others do not care much about environmental protection (i.e., underestimate others' pro-environmental convictions). As alluded to in the introduction, such far-reaching misperceptions may be dispelled by stimulating more talk on the topic. We investigate this proposition in the next simulation.

3.3 First Intervention for Stimulating Pro-Environmental Behaviour: Encouraging Talking

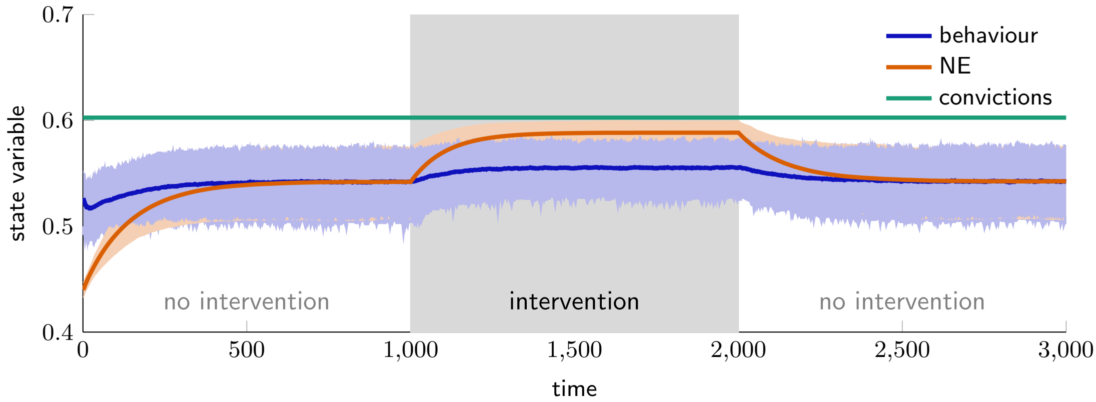
One way to correct misperceptions about others' pro-environmental convictions is by encouraging more conversations on environmental issues. This allows individuals to directly gain insights into other people's actual convictions, rather than relying on proxies like observation of behaviour. Such conversations can be facilitated in various modes, for example, by encouraging one-on-one conversations between direct neighbours, organising local meet-ups and workshops or providing an online space for conversations. To simulate the impact of making people talk more, we change the ABM such that the convictions layer is no longer suppressed. Instead, we set $\gamma_i = 0.75$, which implies that individuals can now learn the convictions of others. This impacts the formation of normative expectations (see Eq. (1)). We implement the intervention in our model for a 1,000 time-steps window, from $t = 1,000$ to $t = 2,000$.[6] Assuming that

[6] As demonstrated in our Supplementary Material (`https://osf.io/w8pjk/`), the intervention has virtually identical effects if introduced from the start (at t = 0).

the influence of normative expectations on behaviour is comparatively small, we use the low value of $\lambda_i = 0.1$ from above.



(a) $\alpha = 0.56$



(b) $\alpha = 0.58$

Figure 5. Results of the simulations in the three-phase scenario with the first intervention, which is implemented in the gray time-window. Solid lines represent the Monte Carlo estimation of the average behaviour (blue), normative expectations (orange) and conviction (green) of the population. The shaded areas represent the envelope in which 90% of the simulations lie.

The effect of the intervention is demonstrated in the grey time-window in Figure 5 which shows that more easy pro-environmental behaviours ($\alpha_i = 0.56$) benefit from increased conversations about environmental issues, while more difficult behaviours ($\alpha_i = 0.58$) require further intervention to become mainstream. In the upper panel of Figure 5, we observe a reduction in pluralistic ignorance over time (i.e., the discrepancy between normative expectations and convictions). This occurs because normative expectations are now informed by individuals' true convictions, rather than inferred from their observable behaviour, as a result of the intervention. However, the discrepancy cannot be resolved for relatively difficult behaviours, as shown in the lower panel of Figure 5. Changing the difficulty of the behaviour — denoted with $\alpha_i \in [0, 1]$ in the ABM — encodes the idea that, for instance, replacing some concrete tiles in the garden with a few bushes is easier than installing solar panels from a monetary perspective, while making it a routine to cycle to work is more challenging for most people than refraining from turning up the heating to save energy (cf. Wolske et al., 2020). Thus, we increased α_i from 0.56 to 0.58 to examine the same intervention's effect on a more difficult behaviour. As shown in Figure 5b, pluralistic ignorance reduces over time, but pro-environmental behaviour lags behind convictions. Our simulations stabilise with a fraction of about 55% of individuals adopting pro-environmental behaviour due to the increased threshold value α_i, being visibly lower than the 60% of adoption reached with the lower threshold value as depicted in Figure 5a.

To explore the persistence of the effect of the intervention, we run the simulations over a time period of $T = 3,000$ and lift the interventions at $t = 2,000$ (by setting $\gamma = 0$). The results are illustrated in the third phase of both panels of Fig. 5. Once the intervention is removed, normative expectations are once again exclusively shaped by observed behaviours, and in the second scenario ($\alpha = 0.58$), these expectations regress to match the fact that most individuals did not change their behaviour. As a result, the fraction of individuals adopting pro-environmental behaviour decreases to 54%, returning to the value observed before the intervention was implemented. Such decrease is not observed for $\alpha = 0.56$, where such fraction remains stable at about 60%.

3.4 Second Intervention for Stimulating Pro-Environmental Behaviour: Increasing Visibility of Pro-Environmental Behaviour

Empirical studies show that signs and stickers that communicate pro-environmental convictions or signal pro-environmental behaviour may have a cascading effect in promoting pro-environmental behaviour (Merkelbach, Dewies, &

Denktas, 2021; Reese, Loeschinger, Hamann, & Neubert, 2013). We assume that such visual indicators make people's pro-environmental efforts more visible to others in the neighbourhood which may cause people to update their normative expectations and even change their own behaviour. Inspired by this idea, we therefore test an intervention with our ABM that increases the visibility of pro-environmental behaviours.[7]

To increase the visibility of pro-environmental behaviours in our ABM, we modify the social network formation process, taking inspiration from Alessandretti, Sun, Baronchelli, and Perra (2017) and Moinet, Barrat, and Pastor-Satorras (2018), such that agents who adopt pro-environmental behaviour have a higher chance of connecting with others (i.e., forming a link on the behaviour layer). A more detailed description of the mathematical implementation of this intervention is provided in Appendix B. Roughly speaking, we include a parameter σ that governs the increased probability of connecting with agents that currently adopt the pro-environmental behaviour. We start with the same simulation setup as in Fig. 4a (i.e., there is no intervention), and implement the intervention at $t = 1,000$ by setting $\sigma = 0.5$ (i.e., increasing the average visibility of pro-environmental behaviour by 50%), and observe the system state at $t = 2,000$.

The effects of increasing the visibility of pro-environmental behaviour are shown in the time-window shaded in grey in Fig. 6. We see that both the average pro-environmental behaviour $\hat{x}(t)$ and average normative expectations $\hat{y}(t)$ increase over time, eventually exceeding average convictions. Hence, we see systematic and collective change in the population, who take up pro-environmental behaviour. In contrast to the results of the first intervention, the findings still hold when the behaviour is perceived as more difficult (i.e., when we increase the threshold value to $\alpha_i = 0.58$): the population still undergoes change, whereby $\hat{x}(2000) \approx 0.64$ (approximately 64% of the population adopts pro-environmental behaviour; see Fig. 6b).

Finally, we consider the scenario in which the intervention is removed at time

[7] While we assume that an increase in visibility represents an intervention (such as disseminating yard signs or stickers), an increase in visibility could also be seen as representing pro-environmental behaviour that is per se more visible (e.g., solar panels).

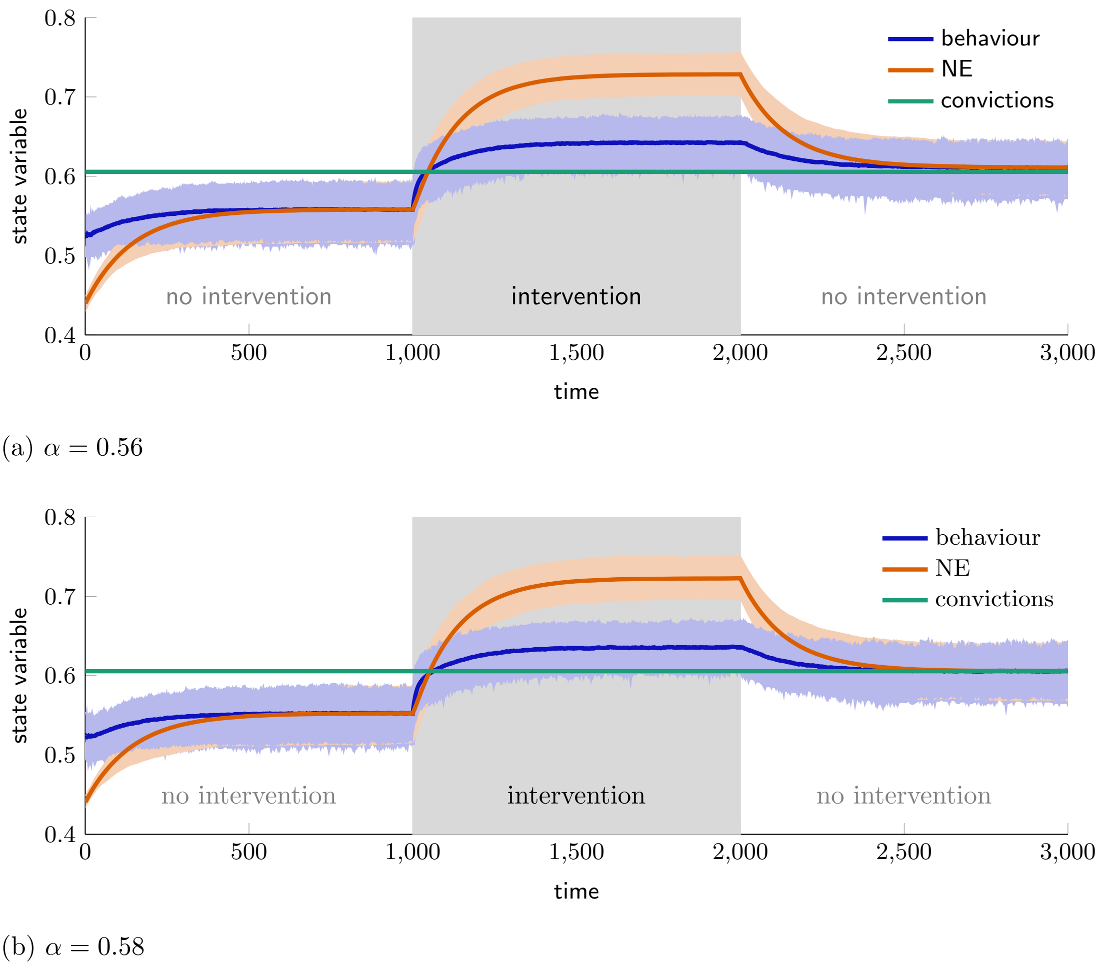


Figure 6. Results of the simulations in the three-phase scenario with the second intervention, which is implemented in the gray time-window. Solid lines represent the Monte Carlo estimation of the average behaviour (blue), normative expectations (orange) and conviction (green) of the population. The shaded areas represent the envelope in which 90% of the simulations lie.

$t = 2,000$, by setting $\sigma = 0$, and evaluating the output of the system over $T = 3,000$ time units. In a real-world setting this could, for example, mean that neighbours remove their stickers. The simulations in Fig. 6 show that, while the overestimation of the normative expectations slowly reduces, the effects of the intervention on individuals' behaviour persist even after its uplift. Eventually, the gaps between pro-environmental behaviour, normative expectations, and convictions are effectively eliminated, yielding $\hat{x}(3000) \approx 0.6$ which means that the fraction of adopters of pro-environmental behaviour is about 60% of the population. Additional simulations reported in the Supplementary

Material demonstrate that our findings are robust with respect to the values of σ.

4. Discussion

Although there is growing concern about climate change worldwide, pro-environmental behaviour is lagging. Mitigating climate change, however, requires rapid collective change (Lenton et al., 2022), involving technological, economic, and social change. In the present paper, we focused on how social influence may shape pro-environmental behaviour. We argue that there is a feedback loop where expectations about others' behaviour impact one's own behaviour and vice versa (cf. T. Bouman & Steg, 2022). Currently, the feedback loop furthers behavioural stagnation because people systematically underestimate how much others care about environmental protection, as they do not hear them talk about it and do not see many people engage in pro-environmental behaviours (Drews et al., 2022; Sparkman et al., 2022). This creates misguided expectations which can discourage those interested in pro-environmental behaviour to act — especially because addressing climate change requires coordinated action (T. Bouman, van der Werff, Perlaviciute, & Steg, 2021).

Using an agent-based model (ABM), we investigated the potential impact of the proposed feedback loop on a network of individuals. Specifically, we examined how this feedback loop might either maintain a state where pro-environmental behaviour is uncommon or propel collective pro-environmental change. We used survey data from 258 respondents to parameterise the ABM. This data from several Dutch neighbourhoods showed that residents underestimated their neighbours' pro-environmental convictions, which may inhibit their own pro-environmental behaviour. They also indicated not talking about environmental issues with neighbours, suggesting that misperceptions might not be corrected spontaneously.

Our first set of ABM simulations show that underestimating others' convictions perpetuates the gap between an individual's own convictions and their pro-environmental behaviour. Without intervention, development towards a more pro-environmental system state is therefore unlikely and the best-case scenario is maintaining the status quo (i.e., current levels of pro-environmental behaviour).

In our remaining simulations, we tested two interventions that aim to disrupt the negative feedback loop: encouraging people talk more about environmental issues (e.g., with a neighbourhood meet-up on the issue) or making pro-environmental behaviour more visible (e.g., with a sticker or a sign). Our simulations suggest that making pro-environmental behaviour more visible may be more effective than encouraging more talk, especially for difficult or costly behavioural changes. Encouraging conversations on the topic corrects normative expectations, but this only leads to change in behaviour which is comparatively easy to carry out. Removing the intervention leads to lowered normative expectations when widespread behaviour change did not occur. Conversely, making pro-environmental behaviour more visible creates lasting change, as it leads people to overestimate how important environmental protection is to others in their social network. This intervention is also effective at producing long-lasting change, with pro-environmental behaviour and normative expectations remaining at new, higher levels once the intervention is removed.

In sum, our results show that wielding social influence processes can create a tipping point for collective pro-environmental change. Notably, both interventions that we simulate focus on the individual level, but their aim is to create collective change by resolving pluralistic ignorance. In our simulations, the ABM-system state (e.g., whether pro-environmental behaviour is widespread or not) is an aggregate of the agents' states (e.g., whether they adopt pro-environmental behaviour or not), which is akin to how collective change in real life arises from changes in individuals' opinions and choices (see, L. Bouman et al., 2021; Feddes, de Lange, & te Brömmelstroet, 2020). Neighbourhoods may be a specifically relevant context for kick-starting such processes: Besides facilitating social influence processes, the connections between residents of different neighbourhoods may also scale up change to larger parts of society (Centola & Macy, 2007; Granovetter, 1978). Nevertheless, we want to stress that the proposed interventions should be used as a complement to other system-level interventions such as changing economic incentives, creating technological innovations, and reforming governance systems (Chater & Loewenstein, 2022). As our simulations show, other system-level interventions can impact the widespread adoption of a specific behaviour by making it easier or more difficult to adopt, for example, by reducing the cost of a technology or creating the required infrastructure.

4.1 Limitations and Suggestions for Future Research

While ABMs are tremendously useful in investigating how macro phenomena emerge from individual behaviour, they remain an abstract, simplified representation of the real world. As such, one must be mindful that conclusions drawn from the simulations must be interpreted while keeping in mind the abstractions and simplifications. In line with this consideration, our ABM employs a limited number of variables and rules to simulate pro-environmental change, making simplified assumptions about the agents, their interactions with each other and the environment.

For instance, the model assumes stable convictions over time, which is a reasonable assumption given that prior research shows that convictions are slow to change (Rozin, 1999). At the same time, the assumption oversimplifies reality as convictions can change as behaviour changes (Poortinga, Whitmarsh, & Suffolk, 2013). Similarly, we assume that the interventions maintain a consistent effect throughout their implementation for our simulations. Hearing about environmental issues too often (especially if the messages lack variety) may, however, lead to a decreased sensitivity or responsiveness to such messages. This, in turn, may diminish the adoption of pro-environmental behaviours (Gifford, 2011). In the ABM, this would mean the parameters associated with the interventions — γ and σ — are time-varying. Yet there is hope that the effects of the interventions that we simulate do not shade over time by establishing new norms (Frey & Rogers, 2014). Research on the longevity of social-norm interventions yields mixed results; while some show lasting effects after two years and minor decay after discontinuation, other studies show no lasting effects (Allcott & Rogers, 2014; Ito, Ida, & Tanaka, 2018). Such mixed findings emphasise the need for further research on the long-term effectiveness of the simulated interventions.

The interactions between agents are modelled based on simple assumptions, too. While we model the influence of normative expectations to be the same across agents, individuals may in reality vary in their susceptibility to normative expectations. Future work could integrate this by collecting data on the distribution of the values for the parameter λ — which captures the weight of normative expectations on an agent's behaviour — in real-world populations. Likewise, the unilateral network links in our model neglect the two-sided exchange of information during a conversation (e.g., neighbours who talk to one another about installing solar panels likely influence each other's thoughts and beliefs about installing them) and the way in which we define interactions between agents ignores the tendency for people to interact with like-minded individuals (McPherson, Smith-Lovin, & Cook, 2001). Incorporating these more sophisticated features into the ABM will be important future work, especially if interventions for specific situations have been identified.

Further, in the current model, talking to others and engaging in a pro-environmental behaviour are treated as separate factors, but they might be most potent when combined. Following the theorising on credibility-enhancing displays (Henrich, 2009), which posits that beliefs are spread more effectively by actions than by words alone, our simulated interventions may have a cumulative effect when implemented together. Empirical research on credibility-enhancing displays supports this, revealing that individuals who themselves engaged in a pro-environmental behaviour were more effective advocates than those who promoted its virtues but did not engage in the behaviour themselves (Kraft-Todd, Bollinger, Gillingham, Lamp, & Rand, 2018). Future work building on our ABM may thus consider exploring possible synergistic effects of the simulated interventions.

Likewise, the simulations do not account for real-life obstacles that may arise when implementing the proposed interventions. For example, people may be willing to do their part for the environment by adopting pro-environmental behaviours, but they might not be keen to talk to others about it (Jones & Niemiec, 2023). Encouraging conversations about environmental topics within a neighbourhood setting may prove difficult as people may anticipate that such conversations can escalate into a heated debate (cf. Bolderdijk & Cornelissen, 2022; Greenebaum, 2012; Minson & Monin, 2012). Encouraging pro-environmental behaviours with visual indicators may also be more complex than we modelled in the simulation. People may not want to make non-mainstream behaviours visible to avoid ridicule or social exclusion (cf. Brick, Sherman, & Kim, 2017). For example, a person who composts may not want to display a sign saying "I compost" if they are the only one in their community doing so.

The data that we use to calibrate the model is also not without limitations. First, while the measures of all the key variables of the ABM were formulated as behaviour-specific items, the item that we used to ask if respondents usually talk with their neighbours about environment-related problems was not behaviour-specific. The item may thus miss certain conversations such as technical questions to a neighbour who owns an electric car. Moreover, besides actively talking, people can also indirectly encounter discourse on the topic (e.g., via the radio or in print media), and that can also change their normative expectations. It is worth noting, however, that such indirect cues likely have a stronger influence on people's general perception of societal attitudes toward the topic rather than how they perceive their specific neighbours' convictions and pro-environmental efforts. Second, the difference we detect between convictions and normative expectations might be attributed, at least in part, to the unidentical nature of the items used to measure each construct. Our results are, however, consistent with prior research demonstrating pluralistic ignorance regarding environmental issues (e.g., Geiger & Swim, 2016; Sparkman et al., 2022). Third, our measure of behaviour may be less predictive of environmental impact than we would want it to be. As demonstrated by Nielsen et al. (2022), pro-environmental behaviour scales can be weakly linked to environmental impact. Therefore, future research employing this ABM should exercise caution in selecting measurement metrics that more accurately capture environmental impact to yield more informative results, enhancing the ability to assess the effectiveness and usefulness of any simulated interventions.

Conducting further research and piloting the simulated interventions before implementing them on a larger scale is thus necessary to test their effectiveness in real-world settings (cf. IJzerman et al., 2020). When designing and testing the interventions in a real-world setting, it is important to carefully consider neighbourhood-specific factors as those may impact their success. For instance, interventions promoting the visibility of pro-environmental behaviour must account for the spatial layout of the neighbourhood, which impacts the visibility of visual cues due to factors like windows, lighting, walls, and distances (Gehl, 1987). Another critical aspect of real-world implementations is collecting data on the type and quality of connections between people. Not all neighbourhoods are characterised by a tight social fabric (Henriksen & Tjora, 2014), which can affect the effectiveness of social influence processes, as they rely on local social ties (cf. Wolske et al., 2020). Future work might also want to tailor the model to a specific pro-environmental behaviour, taking into account specific characteristics of that behaviour. Importantly, the effectiveness of the simulated interventions is likely to depend on the existing prevalence of a behaviour and how widespread particular convictions are (cf. Zino, Ye, & Cao, 2022). For example, interventions that facilitate the identification of pro-environmental behaviour (e.g., visible cues) might only become notably effective once a specific threshold of behaviour adaptation is reached (Rogers, 1962). In situations where pro-environmental behaviour adoption is generally low, targeting pro-environmental convictions could be more effective, considering their potentially broader prevalence compared to actual behaviour. Examining how prevailing behaviour and prevailing convictions affect the interventions' effectiveness provides valuable insights for tailoring the interventions to specific contexts.

5. Conclusion

Our simulations demonstrate that the suggested feedback loop can prompt a network of individuals to adopt varying levels of pro-environmental behaviour and maintain that level of behaviour over time. Hooking into this feedback loop with an intervention can create a tipping point for pro-environmental change. Especially making already occurring pro-environmental behaviour more visible can clear up misperceptions about the adoption of pro-environmental convictions in a community. Given that solving the climate crisis is a problem that requires collective and coordinated action, learning that others are just as concerned about climate change as oneself or learning that others are already acting to mitigate climate change can have a powerful and lasting effect.

References

Alessandretti, L., Sun, K., Baronchelli, A., & Perra, N. (2017). Random walks on activity-driven networks with attractiveness. *Physical Review E*, *95*.

Allcott, H., & Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, *104*(10), 3003–3037.

AlleCijfers. (2022, January). *Dichtheid per km2 van bevolking, adressen en auto's.* https://allecijfers.nl/gemeente/groningen/#dichtheid.

AlleCijfers. (2023, March). *Woningkenmerken per buurt in de gemeente Groningen.* https://allecijfers.nl/ranglijst/woningkenmerken-per-buurt-in-de-gemeente-groningen/#tablerow.

Bergquist, M. (2020). Most people think they are more pro-environmental than others: A demonstration of the better-than-average effect in perceived pro-environmental behavioral engagement. *Basic and Applied Social Psychology*, *42*(1), 50–61.

Bicchieri, C. (2005). *The grammar of society: The nature and dynamics of social norms.* Cambridge University Press.

Bolderdijk, J. W., & Cornelissen, G. (2022). "how do you know someone's vegan?" they won't always tell you. an empirical test of the do-gooder's dilemma. *Appetite*, *168*, 105719.

Bolderdijk, J. W., & Jans, L. (2021). Minority influence in climate change mitigation. *Current Opinion in Psychology*, *42*, 25–30.

Bollinger, B., & Gillingham, K. (2012). Peer effects in the diffusion of solar photovoltaic panels. *Marketing Science*, *31*(6), 900–912.

Bonabeau, E. (2002). Agent-based modeling: Methods and techniques for simulating human systems. *Proceedings of the National Academy of Sciences*, *99*, 7280–7287.

Bouman, L., Antosz, P., Jager, W., Polhill, G., Salt, D., Scalco, A., ... Rodríguez-Arias, A. (2021). *SMARTEES: Reports on scenario development and experiments for selected cases.* Retrieved from https://local-social-innovation.eu/fileadmin/user_upload/Deliverables/

SMARTEES_D7.4_Scenario_development_and_experiments_R2_1_.pdf

Bouman, T., & Steg, L. (2022). A spiral of (in) action: Empowering people to translate their values in climate action. *One Earth*, *5*(9), 975–978.

Bouman, T., Steg, L., & Zawadzki, S. J. (2020). The value of what others value: When perceived biospheric group values influence individuals' pro-environmental engagement. *Journal of Environmental Psychology*, *71*, 101470.

Bouman, T., van der Werff, E., Perlaviciute, G., & Steg, L. (2021). Environmental values and identities at the personal and group level. *Current Opinion in Behavioral Sciences*, *42*, 47–53.

Brick, C., Sherman, D. K., & Kim, H. S. (2017). "green to be seen" and "brown to keep down": Visibility moderates the effect of identity on pro-environmental behavior. *Journal of Environmental Psychology*, *51*, 226–238.

CBS. (2022, April). *Sociaal-economische status; scores per wijk en buurt, regio-indeling 2021.* https://opendata.cbs.nl/#/CBS/nl/dataset/85163NED/table.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*(3), 702–734.

Centola, D., Willer, R., & Macy, M. (2005). The emperor's dilemma: A computational model of self-enforcing norms. *American Journal of Sociology*, *110*(4), 1009–1040.

Chater, N., & Loewenstein, G. (2022). The i-frame and the s-frame: How focusing on individual-level solutions has led behavioral public policy astray. *Behavioral and Brain Sciences*, 1–60.

Cialdini, R. B., Reno, R. R., & 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.

Constantino, S. M., Sparkman, G., Kraft-Todd, G. T., Bicchieri, C., Centola, D., Shell-Duncan, B., . . . Weber, E. U. (2022). Scaling up change: A critical review and practical guide to harnessing social norms for climate action. *Psychological Science in the Public Interest*, *23*(2), 50–97.

Drews, S., Savin, I., Van Den Bergh, J. C., & Villamayor-Tomás, S. (2022). Climate concern and policy acceptance before and after covid-19. *Ecological Economics*, *199*, 107507.

Epley, N., & Caruso, E. M. (2008). Perspective taking: Misstepping into others' shoes. In *Handbook of imagination and mental simulation* (pp. 295–309). Psychology Press.

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

Feddes, F., de Lange, M., & te Brömmelstroet, M. (2020). Hard work in paradise. the contested making of amsterdam as a cycling city. In *The politics of cycling infrastructure* (pp. 133–156). Policy Press.

French Jr, J. R. P. (1956). A Formal Theory of Social Power. *Psychological Review*, *63*(3), 181-194.

Frey, E., & Rogers, T. (2014). Persistence: How treatment effects persist after interventions stop. *Policy Insights from the Behavioral and Brain Sciences*, *1*(1), 172–179.

Gehl, J. (1987). *Life between buildings: Using public space.* New York: Van Nostrand Reinhold.

Geiger, N., & Swim, J. K. (2016). Climate of silence: Pluralistic ignorance as a barrier to climate change discussion. *Journal of Environmental Psychology*, *47*, 79–90.

Gifford, R. (2011). The dragons of inaction: psychological barriers that limit climate change mitigation and adaptation. *American Psychologist*, *66*(4), 290.

Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, *9*(9), 424–430.

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

Greenebaum, J. B. (2012). Managing impressions: "face-saving" strategies of vegetarians and vegans. *Humanity & Society*, *36*(4), 309–325.

Griskevicius, V., Tybur, J. M., & Van den Bergh, B. (2010). Going green to be seen: status, reputation, and conspicuous conservation. *Journal of Personality and*

Social Psychology, *98*(3), 392.

Hamann, K. R., Reese, G., Seewald, D., & Loeschinger, D. C. (2015). Affixing the theory of normative conduct (to your mailbox): Injunctive and descriptive norms as predictors of anti-ads sticker use. *Journal of Environmental Psychology*, *44*, 1–9.

Hanel, P. H., Wolfradt, U., Lins de Holanda Coelho, G., Wolf, L. J., Vilar, R., Monteiro, R. P., … Maio, G. R. (2018). The perception of family, city, and country values is often biased. *Journal of Cross-Cultural Psychology*, *49*(5), 831–850.

Henrich, J. (2009). The evolution of costly displays, cooperation and religion: Credibility enhancing displays and their implications for cultural evolution. *Evolution and human behavior*, *30*(4), 244–260.

Henriksen, I. M., & Tjora, A. (2014). Interaction pretext: Experiences of community in the urban neighbourhood. *Urban Studies*, *51*(10), 2111–2124.

IJzerman, H., Lewis Jr, N. A., Przybylski, A. K., Weinstein, N., DeBruine, L., Ritchie, S. J., … others (2020). Use caution when applying behavioural science to policy. *Nature Human Behaviour*, *4*(11), 1092–1094.

Ito, K., Ida, T., & Tanaka, M. (2018). Moral suasion and economic incentives: Field experimental evidence from energy demand. *American Economic Journal: Economic Policy*, *10*(1), 240–267.

JASP Team. (2023). *JASP (Version 0.17)[Computer software].* Retrieved from `https://jasp-stats.org/`

Jones, M. S., & Niemiec, R. M. (2023). Motivating relational organizing behavior for biodiversity conservation. *Conservation Science and Practice*, e12880.

Kashima, Y., Wilson, S., Lusher, D., Pearson, L. J., & Pearson, C. (2013). The acquisition of perceived descriptive norms as social category learning in social networks. *Social Networks*, *35*(4), 711–719.

Kraft-Todd, G. T., Bollinger, B., Gillingham, K., Lamp, S., & Rand, D. G. (2018). Credibility-enhancing displays promote the provision of non-normative public goods. *Nature*, *563*(7730), 245–248.

Lenton, T. M., Benson, S., Smith, T., Ewer, T., Lanel, V., Petykowski, E., … Sharpe, S. (2022). Operationalising positive tipping points towards global sustainability. *Global Sustainability*, *5*.

Leviston, Z., & Uren, H. V. (2020). Overestimating one's "green" behavior: Better-than-average bias may function to reduce perceived personal threat from climate change. *Journal of Social Issues*, *76*(1), 70–85.

Leviston, Z., Walker, I., & Morwinski, S. (2013). Your opinion on climate change might not be as common as you think. *Nature Climate Change*, *3*(4), 334–337.

Lorenzoni, I., Nicholson-Cole, S., & Whitmarsh, L. (2007). Barriers perceived to engaging with climate change among the uk public and their policy implications. *Global Environmental Change*, *17*(3-4), 445–459.

Maibach, E., Leiserowitz, A., Rosenthal, S., Roser-Renouf, C., & Cutler, M. (2016). Is there a climate "spiral of silence" in america. *Yale Program on Climate Change Communication.*

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Merkelbach, I., Dewies, M., & Denktas, S. (2021). Committing to keep clean: Nudging complements standard policy measures to reduce illegal urban garbage disposal in a neighborhood with high levels of social cohesion. *Frontiers in Psychology*, 2894.

Mildenberger, M., & Tingley, D. (2019). Beliefs about climate beliefs: the importance of second-order opinions for climate politics. *British Journal of Political Science*, *49*(4), 1279–1307.

Minson, J. A., & Monin, B. (2012). Do-gooder derogation: Disparaging morally motivated minorities to defuse anticipated reproach. *Social Psychological and Personality Science*, *3*(2), 200–207.

Moinet, A., Barrat, A., & Pastor-Satorras, R. (2018). Generalized voterlike model on activity-driven networks with attractiveness. *Physical Review E*, *98*, 022303.

Nielsen, K. S., Brick, C., Hofmann, W., Joanes, T., Lange, F., & Gwozdz, W. (2022). The motivation–impact gap in pro-environmental clothing consumption. *Nature*

Sustainability, *5*(8), 665–668.

Nolan, J. M. (2021). Social norm interventions as a tool for pro-climate change. *Current Opinion in Psychology*, *42*, 120–125.

Noppers, E. H., Keizer, K., Bolderdijk, J. W., & Steg, L. (2014). The adoption of sustainable innovations: Driven by symbolic and environmental motives. *Global Environmental Change*, *25*, 52–62.

Nyborg, K., Anderies, J. M., Dannenberg, A., Lindahl, T., Schill, C., Schlüter, M., ... others (2016). Social norms as solutions. *Science*, *354*(6308), 42–43.

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

Perra, N., Gonçalves, B., Pastor-Satorras, R., & Vespignani, A. (2012). Activity driven modeling of time varying networks. *Scientific Reports*, *2*(1), 469. doi: 10.1038/srep00469

Peyton Young, H. (1993). The evolution of conventions. *Econometrica*, *61*(1), 57. doi: 10.2307/2951778

Poortinga, W., Whitmarsh, L., & Suffolk, C. (2013). The introduction of a single-use carrier bag charge in wales: Attitude change and behavioural spillover effects. *Journal of Environmental Psychology*, *36*, 240–247.

Prentice, D. A., & Miller, D. T. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In M. P. Zanna (Ed.), (Vol. 28, p. 161-209). Academic Press.

Proskurnikov, A. V., & Tempo, R. (2017). A tutorial on modeling and analysis of dynamic social networks. Part I. *Annual Reviews in Control*, *43*, 65-79.

R Core Team. (2022). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `https://www.R-project.org/`

Reese, G., Loeschinger, D. C., Hamann, K., & Neubert, S. (2013). Sticker in the box! object-person distance and descriptive norms as means to reduce waste.

Ecopsychology, *5*(2), 146–148.

Rogers, E. M. (1962). *Diffusion of Innovations.* New York: The Free Press of Glencoe.

Ross, L. (1977). The intuitive psychologist and his shortcomings: Distortions in the attribution process. In *Advances in experimental social psychology* (Vol. 10, pp. 173–220). Elsevier.

Rozin, P. (1999). The process of moralization. *Psychological Science*, *10*(3), 218–221.

Salmivaara, L., Lombardini, C., & Lankoski, L. (2021). Examining social norms among other motives for sustainable food choice: The promise of descriptive norms. *Journal of Cleaner Production*, *311*, 127508.

Sharpe, S., & Lenton, T. M. (2021). Upward-scaling tipping cascades to meet climate goals: plausible grounds for hope. *Climate Policy*, *21*(4), 421–433.

Sparkman, G., Geiger, N., & Weber, E. U. (2022). Americans experience a false social reality by underestimating popular climate policy support by nearly half. *Nature Communications*, *13*(1), 1–9.

Steentjes, K., Kurz, T., Barreto, M., & Morton, T. A. (2017). The norms associated with climate change: Understanding social norms through acts of interpersonal activism. *Global Environmental Change*, *43*, 116–125.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: Multivariate imputation by chained equations in r. *Journal of Statistical Software*, *45*(3), 1-67. doi: 10.18637/jss.v045.i03

van Doorn, J., van den Bergh, D., Böhm, U., Dablander, F., Derks, K., Draws, T., . . . others (2021). The jasp guidelines for conducting and reporting a bayesian analysis. *Psychonomic Bulletin & Review*, *28*(3), 813–826.

Wolske, K. S., Gillingham, K. T., & Schultz, P. (2020). Peer influence on household energy behaviours. *Nature Energy*, *5*(3), 202–212.

Zino, L., Ye, M., & Cao, M. (2022). Facilitating innovation diffusion in social networks using dynamic norms. *PNAS Nexus*, *1*(5), pgac229.

Appendix A

Survey Data

Procedure

We obtained ethical approval from the faculty's review board before collecting data. The data was collected as part of a course at the University of Groningen. Students recruited participants by making door-to-door visits in eight different neighbourhoods. A flyer with the survey URL was left in their mailbox if they were not home. The survey was programmed with the software Maptionnaire, and a pen-and-paper questionnaire was available for those who lacked computer literacy. Each neighbourhood was surveyed on one pro-environmental behaviour (i.e., avoiding car use, saving water and energy, installing solar panels, or urban greening).

After providing informed consent, participants answered questions about their neighbourhood, such as the perceived size of the neighbourhood and the number of contacts they had. Next, respondents answered questions measuring their pro-environmental convictions, their expectations about neighbours' pro-environmental convictions (i.e., normative expectations) and neighbours' pro-environmental behaviour (i.e., empirical expectations), and their own pro-environmental behaviour. They also marked on a map where their neighbourhood contacts (i.e., family members, friends, acquaintances) lived and provided demographic information. Lastly, they indicated whether they had conversations with neighbours about problems related to the environment. Not all measures in the questionnaire were relevant to the present paper but were collected for a course assignment and are therefore not reported in the materials section. The questionnaire was offered in Dutch and English. All study items are available at the online repository associated with this paper (`https://osf.io/w8pjk/`).

Data exclusion

We excluded all data on resource conservation due to poor internal consistency of the survey items. For the solar panels data set, we excluded participants who already owned solar panels because the questionnaire was designed for residents who did not have solar panels yet (n = 101). Across all the data, we excluded participants who failed to answer at least one of the questions for each of the variables of interest (i.e., convictions, normative expectations, empirical expectations, pro-environmental behaviour) for this study (n = 38). Lastly, we excluded one participant who indicated to have 100 family members, one who indicated to have more than 300 acquaintances in their neighbourhood, and one who indicated to live for 1250 years in their neighbourhood.

Additional Analysis

Is there pluralistic ignorance? In an exploratory analysis, we compared respondents' pro-environmental convictions with their normative expectations for each of the behaviours separately; again using a Bayesian Wilcoxon signed-rank test. For the data on installing solar panels, the Bayes factor in favour of the hypothesis that convictions and normative expectations differ is 0.219, which means that the observed data are approximately 4.5 times more likely under the null hypothesis. There is thus moderate evidence for no difference between convictions ($M = 4.205$) and normative expectations ($M = 4.136$) for the solar panels data. For this sub-sample, residents thus seem to have rather accurate normative expectations. The accuracy of their expectations is likely due to the high visibility of solar panels and the fact that many residents in the respective neighbourhoods have solar panels on their roofs. On the other hand, there is extreme evidence ($BF_{10} = 540000$) for a difference in means between convictions ($M = 4.245$) and normative expectations ($M = 3.183$) for the greening data. There is similarly extreme evidence ($BF_{10} = 221000$) for a difference in means between pro-environmental convictions ($M = 3.711$) and normative expectations ($M = 2.850$) for the travel behaviour data.

Do social expectations predict respondents' pro-environmental behaviour? We used a Bayesian linear regression to analyse the extent to which normative expectations and empirical expectations predict pro-environmental behaviour. The data was mean-centred for each questionnaire separately. Table A1 shows that the model which only contains normative expectations as a predictor has the best predictive adequacy. After observing data, the odds in favour of the model containing only normative expectations as a predictor have increased by a factor of $BF_M = 8.761$. The observed data are $BF_{01} = 5.272$ times more likely under the model containing only normative expectations as a predictor compared to the model that includes both normative and empirical expectations as predictors.

Table A1

Model Comparison

Model	P(M)	P(M|data)	BF_M	BF_{10}	R^2
Null model	0.250	0.091	0.299	1.000	0.000
Normative Expectations (NE)	0.250	0.745	8.761	8.229	0.033
Empirical Expectations (EE)	0.250	0.023	0.071	0.257	0.005
NE + EE	0.250	0.141	0.494	1.561	0.033

After observing the data, the probability of including empirical expectations falls from 0.5 to 0.165, while the posterior probability of including normative expectations equals 0.886. Including normative expectations in the model produces a $BF_{inclusion} = 7.789$ — the data have increased the prior odds for including normative expectations as a predictor by a factor of approximately 8. This can be considered strong evidence for including normative expectations in the regression model. The point estimate for normative expectations has a posterior mean of 0.273 where the coefficient is 95% probable to lie between 0.000 and 0.477.

Appendix B

Agent-Based Model

Network Formation Process

The ADN formation process occurs at each time-step, as follows.

1. At the start of time-step t, both layers are empty and no links are present ($\mathcal{E}^b(t) = \mathcal{E}^c(t) = \emptyset$).

2. Each individual $i \in \mathcal{V}$ establishes two types of interactions. In the first type, they establish m_b directed links with m_b other individuals selected uniformly at random from the population. These links belong to the behaviour layer. In the second type, the same individual i establishes m_c directed links with m_c other individuals selected uniformly at random from the population. These links belong to the convictions layer. The selections are made independent of selections from previous time-steps, and m_b and m_c are both integers.

3. If individual i is connected to individual j via a link on the behaviour layer, then link (i, j) is added to $\mathcal{E}^b(t)$, which means that individual i is able to obtain information on the behaviour of individual j at time t, that is, $x_j(t)$. Similarly, if individual i is connected to individual k via a link on the convictions layer, then link (i, k) is added to $\mathcal{E}^c(t)$, which means that individual i is able to obtain information on the conviction of individual k, that is, z_k.[8]

4. Using the obtained information of others' behaviours and convictions, each individual i updates $x_i(t)$ and $y_i(t)$ (with the precise update to be defined in the sequel).

[8] The link interactions are directed (unilateral). This means, for instance, that if $j \in \mathcal{N}_i^b(t)$ but $i \neq \mathcal{N}_j^b(t)$, then individual i can learn of $x_j(t)$, but individual j does not know $x_i(t)$. Similarly, for a given individual i, the set of contacts on the behaviour and conviction layers are not automatically the same. It is possible for $\mathcal{N}_i^c(t)$ and $\mathcal{N}_i^b(t)$ to be disjoint, have overlapping elements, or coincide, since selections for link creation are made uniformly at random and independently.

5. Then, all links are removed from the network. Both layers are thus again empty and new links are established at the next time-step, $t+1$ ($\mathcal{E}^b(t+1) = \mathcal{E}^c(t+1) = \emptyset$).

Increasing the Visibility of Pro-Environmental Behaviour

Mathematically, we implement the second intervention as follows: We replace the uniformly random selection of each of the m_b interactions that i establishes at time t on the behaviour layer (described in item 2. of the network formation process description in the above) with the following probabilistic rule:

$$\mathbb{P}[i \text{ interacts with } j] = \begin{cases} \dfrac{1+\sigma}{n-1+\sigma\sum_{k\in\mathcal{V}\setminus\{i\}} x_j(t)} & \text{if } x_j(t) = 1, \\ \dfrac{1}{n-1+\sigma\sum_{k\in\mathcal{V}\setminus\{i\}} x_j(t)}, & \text{if } x_j(t) = 0, \end{cases} \tag{4}$$

where the parameter $\sigma \geq 0$ captures the increase in visibility of people who adopt the pro-environmental behaviour due to, for instance, the use of stickers. Note that $\sigma = 0$ captures the scenario in which the intervention is absent (coinciding with the standard network formation process), while positive values of the parameters capture the presence of such intervention. The magnitude of σ should be interpreted as the average increase in visibility at the population level and accounts to some extent for the fact that only a portion of the population may decide to increase their visibility. Thus, the network formation process on the behaviour layer ensures that each individual has a greater chance of observing other individuals with pro-environmental behaviour than those who are not adopting it.