

The Cognitive Pathways of Sustainability Decisions

Madalina Vlasceanu

mov@stanford.edu

Stanford University <https://orcid.org/0000-0003-2138-1968>

Yash Patel

Cameron Kay

Union College

Ke Fang

Danielle Goldwert

Wei Ma

New York University <https://orcid.org/0000-0002-9835-9083>

Derek Powell

Robert Hawkins

Stanford University

Article

Keywords: Bayesian networks; cognitive constructs; probabilistic graphical models; belief formation; policy attitudes

Posted Date: October 23rd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11448277/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

The Cognitive Pathways of Sustainability Decisions

Yash Patel¹, Cameron S. Kay², Ke Fang³, Danielle Goldwert⁴, Weiji Ma⁵, Derek Powell⁶, Robert Hawkins⁷, Madalina Vlasceanu^{1*}

¹Department of Environmental Social Sciences, Stanford Doerr School of Sustainability, Stanford, CA, USA

² Department of Psychology, New Mexico State University, Las Cruces, NM, USA

³Department of Psychology, Stanford University, Stanford, CA, USA

⁴Department of Psychology, New York University, New York, NY, USA

⁵Center for Neural Science and Department of Psychology, New York University, New York, NY, USA

⁶GoFundMe, San Francisco, CA, USA

⁷Department of Linguistics, Stanford University, Stanford, CA, USA

*Address correspondence to Madalina Vlasceanu, vlasceanu@stanford.edu

Abstract

What cognitive pathways give rise to sustainability-related decisions? Existing psychological theories typically emphasize a limited set of constructs and pre-specified causal pathways, making it difficult to capture how multiple processes jointly produce sustainable decisions. Across five studies ($N = 11,994$), we addressed this question by estimating a Bayesian network from a broad set of demographic, behavioral, cognitive, affective, and social variables. The recovered network contained two primary pathways to policy support and behavioral commitments, one originating in knowledge and the other in identity, both operating through beliefs. We then tested these pathways in four preregistered experiments. In line with the Bayesian model's prediction, we found that increasing knowledge substantially strengthened beliefs and policy support, whereas activating environmental identity produced smaller effects. Tests of the reverse pathways revealed substantial directional asymmetry: policy information did not alter environmental identity and produced only weak changes in knowledge, although it did influence beliefs. Together, these findings show how probabilistic graphical models, combined with experimental validation, can recover and refine candidate cognitive architectures underlying complex psychological outcomes.

Keywords: Bayesian networks; cognitive constructs; probabilistic graphical models; belief formation; policy attitudes

The authors declare no conflicts of interest.

Introduction

Many of society's most consequential outcomes depend on decisions about sustainability. Whether individuals choose to adopt low-carbon behaviors, support environmental policies, or participate in collective action reflects the operation of numerous interacting psychological processes (Vlasceanu & van der Linden, 2026). Research has shown that sustainability decisions are influenced by factual knowledge (Geiger et al., 2019), beliefs about environmental consequences (van der Linden et al., 2015), social norms (Constantino et al., 2022), emotions (Sabherwal et al., 2026), moral values (Graham et al., 2009), identity processes (Fielding & Hornsey, 2016), political ideology (Hornsey et al., 2016), trust in institutions (Smith & Mayer, 2018), habits (Wood & R nger, 2016), and other cognitive and social factors (Steg, 2023). Environmental psychology has produced several influential theories explaining these processes, including the Theory of Planned Behavior (Ajzen, 1991), Value-Belief-Norm theory (Stern, 2000), social norm frameworks (Cialdini & Goldstein, 2004), and identity-based models (Whitmarsh & O'Neill, 2010). Collectively, these theories have substantially advanced understanding of sustainability behaviors and policy support. However, these theories have typically focused on a subset of cognitive processes or psychological constructs, making it difficult to understand how constructs spanning different theories interact within a broader cognitive system. More generally, existing theories adopt a top-down approach, specifying a limited set of constructs and hypothesized causal pathways before testing them empirically. Although this strategy has generated influential psychological theories, it is not designed to recover the broader architecture through which multiple psychological processes jointly shape judgments and decisions. Addressing this challenge requires complementary bottom-up approaches capable of inferring cognitive pathways directly from data. Here, we investigate the cognitive architecture underlying sustainability-related outcomes using a data-driven, bottom-up approach.

Existing theories of sustainability decisions

One influential framework of sustainability decisions is the Value-Belief-Norm (VBN) theory, which proposes that environmental behavior emerges from a sequence of psychological processes linking values, environmental beliefs, personal norms, and behavioral commitments (Stern, 2000; Stern et al., 1999). In this framework, individuals who endorse biospheric or altruistic values are more likely to perceive environmental threats and develop moral obligations to act. VBN theory has been widely used to explain support for environmental policies and participation in pro-environmental behaviors (Steg, 2023; Stern, 2000). Another widely applied model is the Theory of Planned Behavior, which proposes that behavior is driven by behavioral intentions shaped by attitudes, subjective norms, and perceived behavioral control (Ajzen, 1991). This model has been extensively applied to environmental behaviors including recycling, transportation choices, and energy conservation (Bamberg & M ser, 2007; Steg & Vlek, 2009). Meta-analytic evidence suggests that these constructs reliably predict environmentally relevant

behaviors across contexts, highlighting the importance of attitudes and social expectations in shaping decision making (Bamberg & Möser, 2007).

Research has also emphasized the importance of social influence to sustainable decisions. Norm-based models show that individuals frequently adjust their behavior based on perceptions of what others are doing or approving (Cialdini & Goldstein, 2004; Schultz et al., 2007; Constantino et al., 2022). Field experiments demonstrate that information about normative behavior can significantly influence behaviors such as energy conservation, hotel towel reuse, and recycling (Goldstein, Cialdini, & Griskevicius, 2008; Schultz et al., 2007). Dynamic norms, information about how behaviors are changing over time, have also been shown to influence behavioral intentions by signaling emerging social trends (Sparkman & Walton, 2017).

Identity processes represent another major line of research. Studies show that individuals are more likely to engage in pro-environmental behavior when such behavior aligns with their personal or social identities (Whitmarsh & O'Neill, 2010; van der Linden et al., 2015). Environmental identities and prosocial identities can motivate individuals to adopt behaviors consistent with those identities, while ideological identities may influence how environmental information is interpreted (Fielding & Hornsey, 2016). Identity processes can therefore shape not only behavior directly but also how individuals evaluate evidence and social signals.

One longstanding question in this literature concerns the relative roles of informational and identity-based processes in belief formation. Information-based accounts propose that attitudes change as individuals acquire knowledge that alters their beliefs about the consequences or effectiveness of different actions (Coppock, 2023; Pennycook & Rand, 2019). Motivated reasoning accounts instead emphasize that beliefs are shaped by prior identities and motivations, which influence how individuals interpret and evaluate information (Kunda, 1990). These perspectives offer distinct but potentially complementary accounts of belief formation, illustrating a broader challenge for theories that consider psychological processes largely in isolation.

Emotions also play a critical role in sustainability related decision making. Emotional responses such as concern, hope, guilt, and pride can influence engagement with environmental issues and support for environmental policies (Bao & Vlasceanu, 2026; Smith & Leiserowitz, 2014; Lerner, Li, Valdesolo, & Kassam, 2015; Sabherwal, Goldwert, & Vlasceanu, 2026; Vlasceanu et al., 2024). Research on climate communication suggests that emotional engagement influences how individuals process information about environmental risks and evaluate policy responses (van der Linden et al., 2015; Clayton et al., 2015). Emotional responses may therefore function as evaluative signals that interact with beliefs and identities when individuals form policy attitudes.

Beyond these psychological mechanisms, sustainability attitudes are also shaped by broader ideological and moral frameworks. Political ideology strongly predicts climate policy preferences in many countries, particularly in the United States (McCright & Dunlap, 2011; Hornsey, Harris, Bain, & Fielding, 2016; Berkebile-Weinberg et al., 2024). Moral foundations research suggests that individuals evaluate environmental issues partly through moral lenses

emphasizing care, fairness, authority, loyalty, or purity (Graham, Haidt, & Nosek, 2009; Graham et al., 2013). These ideological and moral orientations influence how individuals interpret environmental information and which policies they find acceptable (Mason, Vlasceanu, & Jost, 2024).

From top down to bottom up

Taken together, this literature suggests that sustainability attitudes and decisions arise from a wide range of interacting psychological processes spanning beliefs, norms, emotions, identities, values, and ideological commitments (Clayton et al., 2015; Steg, 2023). However, most empirical analyses examine only a subset of these processes at a time. Studies often focus on variables derived from a single theoretical framework, such as attitudes and norms in the Theory of Planned Behavior, values and personal norms in VBN theory, or identity processes in social identity models (Ajzen, 1991; Stern, 2000; Whitmarsh & O'Neill, 2010). This fragmentation reflects a broader limitation of top-down theoretical models. Each framework proposes a specific causal pathway linking a limited set of constructs to behavior, but the relationships among constructs from different frameworks are rarely examined simultaneously. As a result, it remains difficult to understand how psychological processes identified across theoretical traditions interact within broader cognitive systems.

Moreover, despite the complexity of sustainability decision-making, most psychological models remain fundamentally top-down, specifying a limited set of constructs and hypothesized causal pathways a priori. Regression and structural equation models are well suited for testing these theoretical predictions by estimating the independent effects of variables such as attitudes, norms, values, and identity on policy support and behavior (Ajzen, 1991; Stern, 2000; Bamberg & Möser, 2007; Klöckner, 2013; Steg, 2023). However, because they are designed to test pre-specified relationships, these approaches are not intended to recover the broader architecture of interacting psychological constructs. By contrast, bottom-up approaches from psychological network science infer systems of interacting cognitive components directly from data, allowing researchers to identify how multiple psychological processes jointly shape judgments and behavior (Dalege et al., 2016; Borsboom, 2017).

Indeed, top-down models necessarily impose assumptions about the organization of psychological systems. In practice, judgments and behaviors often emerge from the interaction of multiple cognitive, affective, and social processes rather than from a small set of variables linked in simple linear chains (Kahneman, 2011; Lerner et al., 2015; van der Linden et al., 2015). Beliefs, emotions, identities, social perceptions, motivations, and habits dynamically influence one another, forming interconnected systems in which changes to one process propagate throughout the network (Converse, 1964; Dalege et al., 2016; Wood & Rüniger, 2016). Accordingly, recent work in network psychology has argued that many psychological phenomena may be better understood as systems of interacting components than as collections of isolated predictors (Borsboom, 2017; Boutyline & Vaisey, 2017; Fried & Cramer, 2017).

This challenge reflects a broader limitation in psychological theory: the continued reliance on verbal models. Although verbal theories organize findings and motivate hypotheses, they often leave assumptions about cognitive structure underspecified, allowing multiple interpretations and limiting precise predictions (Urai et al., 2025). Consequently, many scholars have argued for more formal models that require explicit assumptions, enable rigorous comparison between competing theories, and generate sharper predictions (Farrell & Lewandowsky, 2018; Guest & Martin, 2021; Smaldino, 2017).

Computational approaches provide one path toward such formalization. Over the past two decades, probabilistic models have become central to understanding perception, category learning, language, causal reasoning, and decision making (Griffiths et al., 2010; Tenenbaum et al., 2011). Bayesian approaches, in particular, model cognition as probabilistic inference under uncertainty and have successfully explained a wide range of cognitive phenomena (Griffiths et al., 2010; Powell et al., 2023; Tenenbaum et al., 2011). Their principal contribution, however, is not simply predictive accuracy but the ability to formalize assumptions about cognitive structure and mechanism (Guest & Martin, 2021; Jones & Love, 2011; Sanborn & Chater, 2016). Despite these advances, computational approaches have been applied primarily to lower-level cognitive processes, such as perception, language, category learning, and causal reasoning, rather than to higher-level systems of interacting psychological processes. Consequently, less work has examined whether constructs central to social psychology, including beliefs, emotions, identity, social norms, and moral values, can be represented within a formal cognitive pathway model capable of explaining complex social judgments and behavioral commitments (Guest & Martin, 2021; Jones & Love, 2011).

Bayesian Networks

One promising way to formalize such systems is through probabilistic graphical models. In these models, variables are represented as nodes in a graph and their statistical dependencies are represented by edges connecting them (Pearl, 1988; Koller & Friedman, 2009). Among these approaches, Bayesian networks are particularly useful because they represent variables within a directed acyclic graph whose structure encodes conditional dependencies (Powell et al., 2023). This makes it possible to represent both which variables are related, and how information about one variable changes expectations about others. Such models have been widely used in machine learning, epidemiology, and systems biology to study complex systems composed of many interacting components (Pearl, 1988; Koller & Friedman, 2009).

Within psychology, algorithms for learning Bayesian network structure provide a framework for moving beyond purely top-down verbal models toward formal representations of interacting processes and constructs. Rather than specifying the entire causal structure in advance, researchers can allow aspects of the network architecture to be learned from data while imposing only broad theoretical constraints, such as temporal ordering or logical precedence among variables (Peters, Janzing, & Schölkopf, 2017; Spirtes, Glymour, & Scheines, 2000). Accordingly, bottom-up probabilistic modeling can complement traditional top-down theorizing

by identifying candidate architectures of interacting cognitive processes and psychological constructs that can subsequently be tested through experimental and longitudinal research (Guest & Martin, 2021; Smaldino, 2017).

Bayesian networks are distinct from other network approaches increasingly used to model belief systems. Undirected partial-correlation networks characterize patterns of conditional association among beliefs (Turner-Zwinkels & Brandt, 2022; Tang et al., 2025; de Gourville et al., 2025), whereas energy-based models characterize the dynamics through which interconnected beliefs settle into coherent states (Brandt & Sleegers, 2021; Dalege et al., 2025). Bayesian networks are particularly suited to recovering candidate cognitive pathways because they combine three features: directed dependencies, empirical structure discovery, and the ability to incorporate theory-guided constraints on possible pathways. This combination allows the model to discover psychological structure from data while retaining theoretically meaningful restrictions on how psychological constructs can be organized.

We argue that many psychological phenomena, especially complex social judgments, may be better understood as emergent properties of interconnected cognitive systems than as outcomes of isolated predictors or pre-specified linear pathways. Accordingly, here, we aim to recover a candidate pathway model of interacting psychological constructs critical to sustainability decisions. Recent work provides evidence that this approach is feasible. For instance, Powell and colleagues (2023) used Bayesian networks to model intuitive theories underlying vaccine attitudes, showing how beliefs about safety, effectiveness, and social influence form interconnected systems that shape intentions. Their work demonstrates that probabilistic graphical models can capture belief structures that would be difficult to represent adequately as isolated predictors. The present paper builds on these developments proposing Bayesian networks as a general framework for modeling interacting constructs involved in sustainable decisions.

Study 1: Recovering the cognitive pathway of sustainability decisions bottom up

We estimated a Bayesian network from survey data that captured a broad range of psychological variables related to sustainability related policy attitudes and behavioral commitments. The purpose of this analysis was to estimate a candidate pathway of relationships among psychological constructs, a probabilistic structure that can inform future theoretical and experimental work. Bayesian networks provide a formal representation of conditional dependencies among variables (Pearl, 1988; Koller & Friedman, 2009). In these models, variables are represented as nodes in a directed acyclic graph (DAG), and edges represent statistical dependencies between variables conditional on other nodes in the network. Importantly, the resulting graph represents a factorization of the joint probability distribution over variables. As a result, Bayesian networks can capture how information about one variable probabilistically changes expectations about others. Although directed edges in such networks can sometimes correspond to causal relationships, causal interpretation requires additional assumptions beyond observational data (Pearl, 2009; Peters, Janzing, & Schölkopf, 2017). For

this reason, the network estimated here should be interpreted primarily as a representation of conditional dependencies among cognitive constructs, rather than as definitive evidence of causal relationships. The value of the model lies in identifying a structured architecture of interactions that can guide theoretical interpretation and experimental testing.

Methods Study 1

Participants. A total of 4,061 participants were recruited through the online research platform Prolific. After excluding participants who failed an attention check or provided incomplete responses, the final sample consisted of 3,993 participants. The research protocol was reviewed and approved by the Stanford University Institutional Review Board.

Design and Procedure. Participants were randomly assigned to consider one of four policy-relevant behavioral domains involving sustainability-related decisions: reducing red meat consumption, adopting electric vehicles, reducing air travel, or incorporating renewable energy sources into the electricity supply. These domains were selected because they represent distinct forms of behavioral and policy decisions while engaging similar cognitive constructs, including beliefs about consequences, perceptions of social norms, identity alignment, emotional responses, and motivations. Across these domains, participants completed a series of psychological measures designed to capture a broad set of cognitive constructs commonly studied in social and cognitive psychology. These included measures of media exposure, domain-specific informational exposure, behavioral habits, factual knowledge, beliefs about environmental impacts, perceptions of static and dynamic social norms, emotional responses, intrinsic and extrinsic motivations, identity-related constructs, trust in institutions, social dominance orientation, and moral values. Participants also reported their support for policy interventions relevant to the behavioral domain as well as their willingness to commit to related behavioral changes. Although these measures were collected within the sustainability domain, they were selected primarily because they represent general psychological constructs implicated in belief formation, social perception, moral evaluation, and motivation. The network therefore provides an illustration of how diverse cognitive constructs may interact within a broader psychological system.

Materials. Participants completed the measures listed in Table S1. All multi-item constructs were averaged into a single composite score per participant and rescaled between 0 and 1 before network estimation; reverse-keyed items (marked R) were recoded first. Item wording was matched across the four behavioral domains, with domain-specific substitutions in the bracketed slots (e.g. “reducing red meat consumption”, “switching to an electric vehicle”).

Network Construction. To estimate the structure of relationships among these variables, we constructed Bayesian networks using the Directed Acyclic Graph with Acyclicity

Characterization (DAGMA) algorithm (Bello et al., 2022). DAGMA is a structure-learning method that estimates directed graphs by minimizing a least-squares loss function while enforcing an acyclicity constraint. The algorithm searches for a graph structure that best captures statistical dependencies among variables while ensuring that the resulting network contains no feedback cycles. Structure learning in graphical models involves identifying a graph that approximates the conditional dependency structure of the observed data (Spirtes, Glymour, & Scheines, 2000; Peters et al., 2017). Because many graph structures can represent the same statistical dependencies, such models should be interpreted as representing one member of an equivalence class of possible causal graphs. The goal is therefore not to recover a uniquely correct causal structure but to identify a plausible architecture of dependencies consistent with both the data and theoretical constraints. To incorporate theoretical knowledge about temporal and causal plausibility, we introduced minimal structural constraints during model estimation. Specifically, variables were grouped into three partially ordered tiers. The first tier included demographic variables and established behavioral habits. These variables represent relatively stable characteristics that plausibly precede the formation of attitudes and beliefs. The second tier included psychological constructs such as knowledge, beliefs, norms, emotions, identity processes, trust, and moral values. The third tier included outcome variables representing policy support and behavioral commitments. Edges were allowed to form within tiers and from earlier tiers to later tiers but not from later tiers to earlier tiers. For example, demographic characteristics could influence beliefs or motivations, but policy attitudes were not allowed to influence demographic variables. These constraints reflect broad assumptions about temporal ordering while still allowing complex relationships among psychological variables to emerge from the data. Such partially constrained approaches are commonly used in structure learning when prior theoretical knowledge about variable ordering exists (Spirtes et al., 2000; Peters et al., 2017). Importantly, these constraints were intentionally minimal in order to preserve the exploratory nature of the model. To reduce the risk of overfitting, the dataset was divided into a training set comprising 80% of observations and a validation set comprising the remaining 20%. The training data were used to estimate the network structure, and the validation data were used to assess predictive performance.

Parameter Estimation. Once the network structure had been learned, the strength of relationships between nodes was estimated using linear Gaussian conditional probability distributions (Koller & Friedman, 2009). In this framework, each variable is modeled as a linear function of its parent variables plus Gaussian noise. Linear Gaussian models were chosen for two primary reasons. First, most variables in the dataset were continuous or approximately continuous measures. Second, linear Gaussian conditional distributions provide interpretable parameter estimates while remaining computationally tractable for networks with multiple nodes. Parameter estimation was performed using ordinary least squares regression for each node conditional on its parents. Although these local models capture relationships between parent and child nodes, the global structure of the Bayesian network reflects dependencies across the entire

system. As a result, the network encodes how information about one variable probabilistically propagates through the system via multiple pathways.

Results Study 1

The resulting network consisted of 25 nodes connected by 31 directed edges (Figure 1). Each edge represents a conditional dependency between variables, with the direction indicating the orientation of the dependency within the DAG. Edge weights reflect the magnitude and polarity of these relationships.

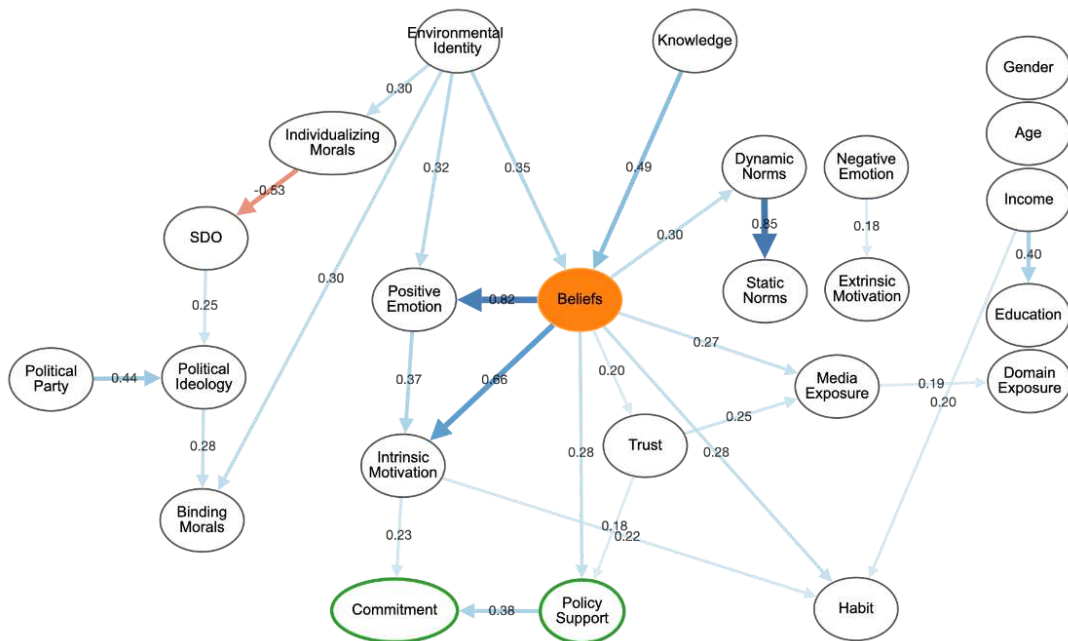


Figure 1. A Bayesian network model of interacting cognitive constructs. Nodes represent demographic, psychological, and outcome variables; directed edges represent conditional dependencies estimated from the data under partial ordering constraints; we only display edges with strength greater than 0.1. Blue edges indicate positive relations and red edges indicate negative relations, with color intensity reflecting edge magnitude. The network should be interpreted as a candidate probabilistic set of pathways of interacting cognitive constructs rather than a definitive causal model. “SDO” stands for Social Dominance Orientation. The ‘Beliefs’ node is colored orange to represent its centrality in the network. Policy support and Commitments are colored in green to represent they are the outcome variables of interest.

Several structural patterns emerged from the network. Most prominently, beliefs occupy a central position linking upstream cognitive constructs to policy attitudes, with a betweenness centrality of 0.027, the largest in the network, and 3.864 times bigger than the second highest node’s betweenness centrality.

Moreover, in the estimated architecture, beliefs arise primarily from two sources: knowledge and identity-related processes. Indeed, both knowledge (edge weight = 0.490, $R^2 = 0.139$) and environmental identity (edge weight = 0.347, $R^2 = 0.106$) predicted beliefs, which in turn, predicted support for related policies (edge weight = 0.282, $R^2 = .432$), which also predicted corresponding behavioral commitments (edge weight = 0.379, $R^2 = .377$).

Figure 2 depicts the total, non-isolated correlation between observed and predicted values of the model. A higher correlation means more alignment between predicted and observed variables, meaning a better model fit.

To assess how reliably the search recovered these edges, we refit the structure on bootstrap resamples of the training data ($n = 200$). The knowledge-to-beliefs edge was selected in 100% of resamples, whereas the environmental identity-to-beliefs edge was selected in 78%. Both were positively signed in every resample in which they appeared.

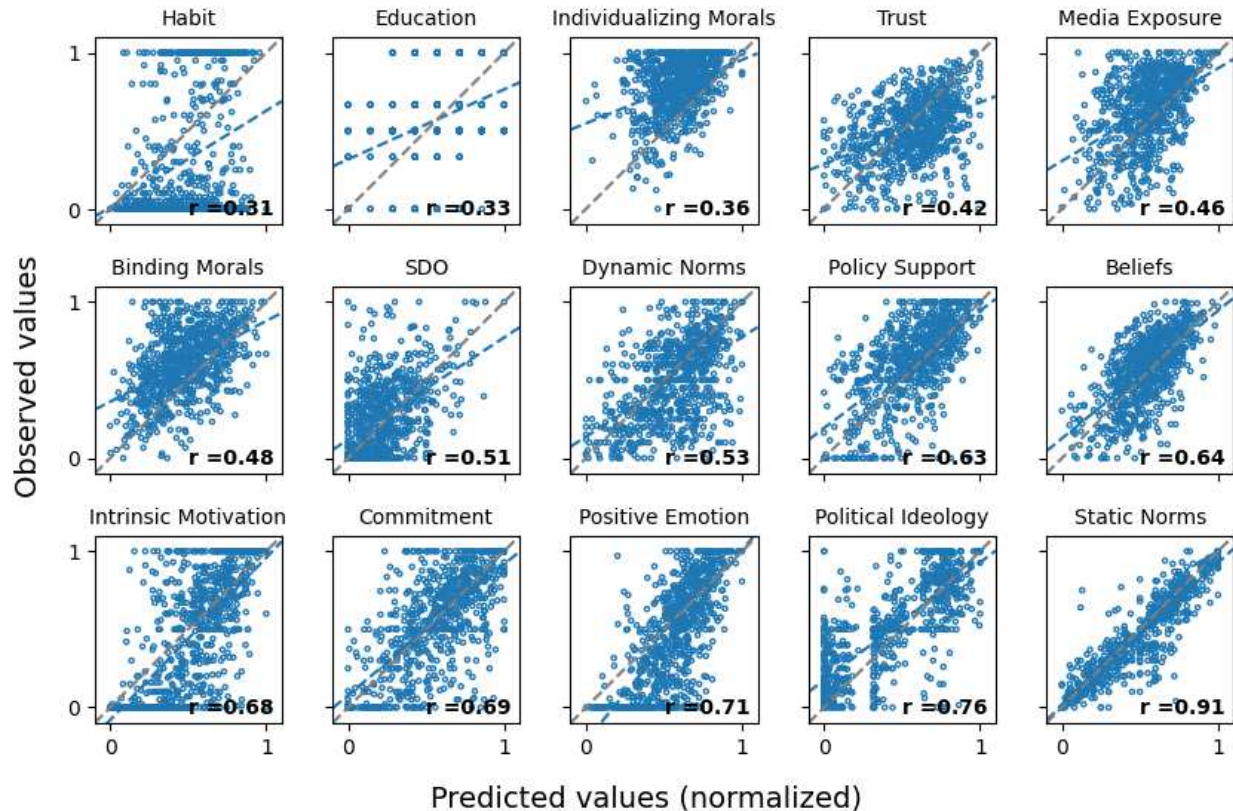


Figure 2. Observed Responses and Cognitive Model Predictions for Each Belief Among Participants in the Testing Split. Correlations between observed and predicted values ranged from 0.31 to 0.91 across the different belief scales.

Discussion Study 1

In Study 1, we used a bottom-up Bayesian network approach to identify relationships among cognitive, affective, and social processes underlying sustainability decisions. Using survey data

from 3,993 participants across four sustainability domains, we recovered a network of 25 variables connected by 31 directed dependencies. The resulting architecture identified two primary pathways originating in knowledge and environmental identity, both converging on beliefs, which occupied the most central position in the network and predicted policy support and subsequent behavioral commitments.

The first pathway suggests that increasing knowledge about the consequences or effectiveness of sustainable behaviors should strengthen related beliefs and, through them, policy support, consistent with an information deficit model of persuasion (Coppock, 2023; Pennycook & Rand, 2019). Accordingly, providing factual information about environmental impacts, technological feasibility, or policy effectiveness should produce changes in beliefs that translate into greater policy support.

The second pathway suggests that identity-related processes shape how individuals evaluate sustainability issues, consistent with motivated reasoning accounts of belief formation (Van Bavel et al., 2024; Kunda, 1990). Activating environmental identity should therefore strengthen positive beliefs about sustainable behaviors, indirectly increasing support for related policies.

Together, these pathways generate experimentally testable predictions about the cognitive processes underlying sustainability decisions. Because the Bayesian network represents conditional dependencies rather than established causal relationships, Studies 2 and 3 experimentally manipulate knowledge and environmental identity to test their predicted effects on beliefs and policy support.

Study 2 and 3

The purpose of experiments 2 and 3 was to test the effect of knowledge (Study 2) and identity (Study 3) on one's beliefs about four sustainable behaviours (e.g., eating less red meat; driving an electric vehicle; reducing air travel; sourcing renewable electricity), as well as policies directed at increasing the adoption of those four sustainable behaviours.

As preregistered on the Open Science Framework (OSF; https://osf.io/wvqam/overview?view_only=d1f00253b0704bf2864f31e3648f810b), we hypothesized that (H1a) gaining knowledge about a sustainable behaviour would increase positive beliefs about the behaviour, (H1b) gaining knowledge about a sustainable behaviour would increase support for policies directed towards increasing adoption of the behaviour, and (H1c) positive beliefs about a sustainable behaviour would mediate the association between knowledge about the behaviour and support for a policy related to the behaviour.

Moreover, we hypothesized that (H2a) priming an environmental identity would increase positive beliefs about a sustainable behaviour, (H2b) priming an environmental identity would increase support for a policy directed towards increasing adoption of a sustainable behaviour, and (H2c) positive beliefs about a sustainable behaviour would mediate the association between one's environmental identity and support for a policy related to the behaviour. The materials,

data, analytic code, and output for the present studies can be found on OSF (https://osf.io/zsh5f/overview?view_only=ff94b89e3f0f4ea4b6dc2db07148ea71).

Methods Study 2 and 3

Participants. We based the sample sizes on an a priori power analysis of the most complex models we aimed to test (i.e., the mediation models). The power analysis (<https://davidakenny.shinyapps.io/MedPower/>; Kenny, 2017) indicated that we would need a sample of 369 participants to detect an indirect effect of .040 or larger 95% of the time that such an effect existed with an alpha level of .05 and assuming a moderate effect ($\beta = .20$) between the predictor (knowledge or identity) and mediator (beliefs), between the mediator (beliefs) and the outcome variable (policy support), and between the predictor (knowledge or identity) and the outcome variable (policy support). To account for approximately 20% of the participants responding carelessly, we increased the targeted sample size to 500 for each study-sample combination. The research protocol was reviewed and approved by the Stanford University Institutional Review Board.

Participants were recruited on Prolific and randomly assigned to complete either Study 2 or Study 3 for one of the four sustainable behaviours. They were paid at a rate equivalent to approximately \$16.50 per hour and were required to be 18 years of age or older and located in the US to participate. We further used quota sampling to demographically match the samples to the 2020 US Census. After implementing our preregistered exclusion criteria, Study 2 included 2,041 participants (506 to 517 participants for each sustainable behaviour) and Study 3 included 1,965 participants (485 to 497 participants for each sustainable behaviour). Exclusion counts and select demographic information can be found in Table 1.

Table 1

Select demographic information for the participants in Study 2 (Knowledge) and Study 3 (Identity).

	Study 2 (Knowledge)					Study 3 (Identity)				
	All	Diet	EVs	Air Travel	Renewa bles	All	Diet	EVs	Air Travel	Renewa bles
<i>Sample Size</i>										
Total	2050	508	511	520	511	1979	499	493	487	500
Exclusions – Instructed										
Response Items	8	2	1	2	3	14	2	3	2	7
Exclusions – Duration	0	0	0	0	0	0	0	0	0	0
Excluded – LLM										
Detection Items	1	0	0	1	0	0	0	0	0	0
Final	2041	506	510	517	508	1965	497	490	485	493
<i>Age Mean (SD)</i>	44.29 (15.94)	44.79 (15.86)	44.92 (15.86)	44.59 (16.39)	42.82 (15.60)	44.19 (15.78)	43.53 (15.56)	44.72 (15.88)	44.63 (16.22)	43.89 (15.49)
<i>Gender Count (%)</i>										
Women	1026 (50.27%)	250 (49.41%)	251 (49.22%)	268 (51.84%)	257 (50.59%)	1003 (51.04%)	257 (51.71%)	254 (51.84%)	241 (49.69%)	251 (50.91%)
Men	968 (47.43%)	244 (48.22%)	246 (48.24%)	236 (45.65%)	242 (47.64%)	931 (47.38%)	234 (47.08%)	224 (45.71%)	233 (48.04%)	240 (48.68%)
Non-binary	27 (1.32%)	8 (1.58%)	6 (1.18%)	8 (1.55%)	5 (0.98%)	11 (0.56%)	4 (0.80%)	1 (0.20%)	6 (1.24%)	0 (0.00%)
Genderfluid	9 (0.44%)	2 (0.40%)	3 (0.59%)	3 (0.58%)	1 (0.20%)	8 (0.41%)	0 (0.00%)	6 (1.22%)	1 (0.21%)	1 (0.20%)
Other	1 (0.05%)	0 (0.00%)	1 (0.20%)	0 (0.00%)	0 (0.00%)	3 (0.15%)	1 (0.20%)	2 (0.41%)	0 (0.00%)	0 (0.00%)
Preferred not to answer	10 (0.49%)	2 (0.40%)	3 (0.59%)	2 (0.39%)	3 (0.59%)	9 (0.46%)	1 (0.20%)	3 (0.61%)	4 (0.82%)	1 (0.20%)
<i>Cultural/Ethnic Identity Count (%)</i>										
White or Caucasian	1268 (62.13%)	312 (61.66%)	313 (61.37%)	330 (63.83%)	313 (61.61%)	1253 (63.77%)	314 (63.18%)	306 (62.45%)	314 (64.74%)	319 (64.71%)
Black or African-American	235 (11.51%)	58 (11.46%)	54 (10.59%)	63 (12.19%)	60 (11.81%)	227 (11.55%)	55 (11.07%)	64 (13.06%)	56 (11.55%)	52 (10.55%)
Asian or Asian-American	129 (6.32%)	34 (6.72%)	37 (7.25%)	28 (5.42%)	30 (5.91%)	95 (4.83%)	21 (4.23%)	26 (5.31%)	24 (4.95%)	24 (4.87%)

Hispanic/Latinx	157 (7.69%)	32 (6.32%)	52 (10.20%)	30 (5.80%)	43 (8.46%)	137 (6.97%)	41 (8.25%)	35 (7.14%)	28 (5.77%)	33 (6.69%)
Multiple cultures/ethnicities	185 (9.06%)	56 (11.07%)	38 (7.45%)	48 (9.28%)	43 (8.46%)	204 (10.38%)	53 (10.66%)	49 (10.00%)	50 (10.31%)	52 (10.55%)
Other cultural/ethnic identity	54 (2.65%)	13 (2.57%)	12 (2.35%)	15 (2.90%)	14 (2.76%)	37 (1.88%)	9 (1.81%)	8 (1.63%)	10 (2.06%)	10 (2.03%)
Preferred not to answer	13 (0.64%)	1 (0.20%)	4 (0.78%)	3 (0.58%)	5 (0.98%)	12 (0.61%)	4 (0.80%)	2 (0.41%)	3 (0.62%)	3 (0.61%)
<i>Political Affiliation Count (%)</i>										
Democratic	833 (40.81%)	208 (41.11%)	218 (42.75%)	215 (41.59%)	192 (37.80%)	781 (39.75%)	201 (40.44%)	195 (39.80%)	193 (39.79%)	192 (38.95%)
Republican	458 (22.44%)	105 (20.75%)	111 (21.76%)	117 (22.63%)	125 (24.61%)	467 (23.77%)	115 (23.14%)	119 (24.29%)	113 (23.30%)	120 (24.34%)
Independent	599 (29.35%)	154 (30.43%)	139 (27.25%)	159 (30.75%)	147 (28.94%)	585 (29.77%)	145 (29.18%)	149 (30.41%)	141 (29.07%)	150 (30.43%)
Other political affiliation	38 (1.86%)	13 (2.57%)	9 (1.76%)	10 (1.93%)	6 (1.18%)	37 (1.88%)	16 (3.22%)	8 (1.63%)	9 (1.86%)	4 (0.81%)
Undeclared	113 (5.54%)	26 (5.14%)	33 (6.47%)	16 (3.09%)	38 (7.48%)	95 (4.83%)	20 (4.02%)	19 (3.88%)	29 (5.98%)	27 (5.48%)

Note. Additional demographic information (e.g., states of residence; educational attainment) can be found in the results document provided on OSF (https://osf.io/zsh5f/overview?view_only=ff94b89e3f0f4ea4b6dc2db07148ea71). Participants were excluded for failing both of the instructed response items embedded in the survey (Curran, 2016; Kay & Saucier, 2023); completing the survey in less than one-third of the median response time (Bedford-Petersen & Saucier, 2021); and failing three or more of the four ECLAIR LLM-detection items embedded in the survey (Kay & Vlasceanu, 2026).

Materials Study 2 and 3

Knowledge Manipulation. In Study 2, we manipulated participants' levels of knowledge about the sustainable behaviours by showing them five to six pieces of mundane trivia (e.g., "Tasmania was formerly known as Van Diemen's Land"; the control condition) or five to six pieces of information about the sustainable behaviour (e.g., "Renewable energy sources accounted for around 20% of U.S. electricity generation in 2023"; the treatment condition). The stimuli for the control condition was drawn from Miller et al. (in prep) (see also Galli et al., 2018) and was purposefully selected to not be curiosity provoking, so as not to introduce additional variability into the data. The stimuli for the treatment condition was drawn from the knowledge items in Study 1.

To test the effectiveness of the knowledge manipulation, participants completed a six-item knowledge test (e.g., "Renewable energy sources accounted for around what percentage of U.S. electricity generation in 2023?") related to the sustainable belief they were assigned. The knowledge test was found to be reliable across the four contexts (α s = .69 to .74; ω s = .74 to .81) and indicated that those who were shown information about a sustainable behaviour (i.e., those in the treatment condition) were more knowledgeable ($t(2031) = 47.24, p < .001, d = 2.09$) about the sustainable behaviour than those shown mundane trivia (i.e., those in the control condition) (Table 2)

Identity Manipulation. In Study 3, we primed participants with an environmental identity by having them either list three things they like about being indoors, visualize being indoors, and write three ways that being indoors can benefit humanity (the control condition) or list three things they like about being in nature, visualize being in nature, and write three ways that being in nature can benefit humanity (the treatment condition).

To test the effectiveness of the environmental identity manipulation, participants completed the 14-item *Revised Environmental Identity Scale* (e.g., "I think of myself as a part of nature, not separate from it"; Clayton et al., 2021) at the end of the survey. The scale was shown to be reliable across the four contexts (α s = .90 to .91; ω s = .90 to .91) and indicated that those who completed the nature induction (i.e., those in the treatment condition) had a greater environmental identity ($t(1875) = 6.25, p < .001, d = 0.28$) than those who thought about the indoors (i.e., those in the control condition) (Table 2).

Table 2

The effect of the knowledge manipulation on knowledge in Study 2 and identity manipulation on identity in Study 3.

Study 2: Did the knowledge manipulation increase knowledge?	Study 3: Did the identity manipulation increase environmental identity?
--	--

	Treatment						Treatment					
	Control: $M_1 (SD_1)$	: $M_2 (SD_2)$	t	df	p	d	Control: $M_1 (SD_1)$	: $M_2 (SD_2)$	t	df	p	d
All	0.46 (0.18)	0.85 (0.19)	47.24	2031	<.001	2.09	1.69 (0.95)	1.93 (0.77)	6.25	1875	<.001	0.28
Diet	0.41 (0.19)	0.81 (0.22)	21.63	493	<.001	1.92	1.67 (1.01)	1.90 (0.78)	2.87	460	.004	0.26
EV	0.48 (0.17)	0.88 (0.16)	27.57	503	<.001	2.44	1.58 (0.95)	1.92 (0.78)	4.29	469	<.001	0.39
Air Travel	0.47 (0.18)	0.88 (0.17)	25.89	511	<.001	2.28	1.81 (0.85)	1.93 (0.77)	1.66	474	.099	0.15
Renewable s	0.49 (0.17)	0.84 (0.20)	21.29	481	<.001	1.89	1.68 (0.95)	1.96 (0.77)	3.60	469	<.001	0.32

Beliefs. In Study 2 and Study 3, we assessed baseline and post-manipulation beliefs about the four sustainable behaviours by having participants respond to four items (e.g., “To what extent do you think using renewable energy sources will improve the environment?”). (We administered six items for each sustainable behaviour for consistency with a prior study, but, as preregistered, we planned to analyze only four of them.) Participants responded to the four items using a 101-point slider scale (0 = “Not at all”; 100 = “Extremely so”). The four items were shown to be reliable across the sixteen study-time-context combinations (α s = .80 to .90; ω s = .81 to .91).

Policy Support. In Study 2 and Study 3, we assessed baseline and post-manipulation support for policies related to the four sustainable behaviours by having participants complete two items (e.g., “To what degree do you support the following policies? Investing in renewable energy technologies.”). (We administered three items for each sustainable behaviour for consistency with a prior study, but, as preregistered, we planned to analyze only two of them.) Participants responded to the two items using a 101-point slider scale (0 = “Not at all”; 100 = “Extremely so”). The two items were shown to be fairly reliable across the sixteen study-time-context combinations (α s = .65 to .96; ω s = .66 to .96).

Procedure. After enrolling in the study on Prolific, participants were randomly assigned to one of eight surveys on Qualtrics. Four of the surveys were part of Study 2 and included the knowledge manipulation; the other four surveys were part of Study 3 and included the identity manipulation. The four surveys for each study focused on one of the four sustainable behaviours: (a) reducing consumption of red meat, (b) increasing adoption of electric vehicles, (c) reducing air travel, or (d) increasing use of renewable electricity. In all surveys, participants were first asked to provide informed consent, confirm that they were 18 years of age or older, and confirm that they currently live in the US. Participants who did not provide consent were unable to progress past the first page of the study. Participants who reported that they were not 18 years of age or older or who reported that they live outside of the US, were redirected to the end of the

survey. After completing the qualification portion of the surveys, the participants completed the scales assessing their baseline beliefs and levels of policy support. Participants in Study 2 then underwent the knowledge manipulation, while participants in Study 3 underwent the identity manipulation. Participants in both studies then completed the scales assessing their post-treatment beliefs and levels of policy support, followed by the manipulation checks for their respective manipulations. Participants ended the survey by providing additional demographic information and being debriefed. They were then redirected to the Prolific platform for payment.

Study 2 and 3 Results

The results of Study 2 indicated that, regardless of the sustainable behaviour in question, knowledge had a sizeable impact on a person's beliefs and levels of policy support (Table 3). Specifically, as a person became more knowledgeable about a given sustainable behaviour they were both more likely to hold positive beliefs about that behaviour ($\beta = 0.37$, 95% CI [0.33, 0.41], $p < .001$) and support policies intended to encourage that behaviour ($\beta = 0.19$, 95% CI [0.16, 0.22], $p < .001$). Furthermore, beliefs mediated the effect of knowledge on policy support (mediation supported by an indirect effect $\beta=0.05$, $p<.001$; Table 4), such that knowledge appeared to contribute to policy support because it increased a person's positive beliefs about the sustainable behaviour.

Similarly, the results of Study 3 indicated that inducing an environmental identity influenced a person's beliefs and levels of policy support (Table 3). Namely, when a person was primed with an environmental identity, they appeared to hold more positive beliefs about sustainable behaviours overall ($\beta = 0.06$, 95% CI [0.03, 0.09], $p < .001$) and were more likely to support policies intended to encourage those behaviours ($\beta = 0.05$, 95% CI [0.02, 0.08], $p = .003$). Importantly, however, these effects were weaker than those observed for knowledge (β s = .02 to .08 versus β s = .13 to .56) and did not increase beliefs about reducing meat consumption, nor did it increase policy support about reducing meat consumption, driving electric vehicles, or reducing air travel. We did find that beliefs mediated the effect of identity on policy support in the combined data, but the effect was negligible ($\beta < .01$) (Table 4).

Table 3

The effect of the knowledge manipulation (Study 2) and identity manipulation (Study 3) on beliefs and levels of policy support.

	Study 2: Does knowledge affect sustainable beliefs and levels of policy support?						Study 3: Does identity affect sustainable beliefs and levels of policy support?					
	Beliefs			Policy Support			Beliefs			Policy Support		
	β	95% CI	p	β	95% CI	p	β	95% CI	p	β	95% CI	p
All	0.37	[0.33, 0.41]	< .001	0.19	[0.16, 0.22]	< .001	0.06	[0.03, 0.09]	< .001	0.05	[0.02, 0.08]	.003
Diet	0.30	[0.23, 0.37]	< .001	0.22	[0.15, 0.29]	< .001	0.05	[-0.01, 0.11]	.078	0.00	[-0.01, 0.11]	.13

EV	0.4	[0.36,	<	0.2	[0.15,	<				0.0	[-0.01,	.09
	3	0.49]	.001	1	0.26]	.001	0.06	[0.01, 0.11]	.028	5	0.10]	9
Air Travel	0.5	[0.47,	<	0.2	[0.16,	<				0.0	[-0.06,	.62
	6	0.65]	.001	4	0.31]	.001	0.08	[0.01, 0.16]	.030	2	0.11]	0
Renewables	0.2	[0.18,	<	0.1	[0.08,	<				0.0		.02
	6	0.33]	.001	3	0.19]	.001	0.08	[0.01, 0.14]	.024	7	[0.01, 0.13]	6

Note. Baseline beliefs were included as covariates in the belief models, and baseline levels of policy support were included as covariates in the policy support models.

Table 4

Beliefs as a mediator of the effect of knowledge on policy support (Study 2) and identity on policy support (Study 3)

	Study 2:			Study 3:		
	Does belief mediate the effect of knowledge on policy support?			Does belief mediate the effect of identity on policy support?		
	β	95% CI	<i>p</i>	β	95% CI	<i>p</i>
All	0.05	[0.04, 0.06]	< .001	0.00	[0.00, 0.01]	.001
Diet	0.04	[0.03, 0.06]	< .001	0.00	[-0.00, 0.01]	.119
EV	0.06	[0.04, 0.09]	< .001	0.00	[0.00, 0.01]	.093
Air Travel	0.10	[0.06, 0.13]	< .001	0.01	[0.00, 0.02]	.078
Renewables	0.04	[0.02, 0.06]	< .001	0.01	[0.00, 0.02]	.056

Note. Beta is the standardized product of the (a) knowledge/identity to beliefs path and (b) beliefs to policy support path. Standard errors were estimated using bootstrapping with 10,000 iterations. Baseline beliefs were added as predictors of post-treatment beliefs, and baseline levels of policy support were added as predictors of post-treatment levels of policy support.

Discussion Study 2 and 3

In Studies 2 and 3, we experimentally tested the two pathways identified by the Bayesian network in Study 1 by manipulating knowledge (Study 2, $N = 2,041$) and environmental identity (Study 3, $N = 1,965$) across four sustainability domains. Consistent with the Bayesian model's predictions, increasing knowledge strengthened beliefs about sustainable behaviors and support for related policies, with belief change partially mediating the effect on policy support. Activating environmental identity also increased beliefs and policy support, although these effects were substantially smaller and less consistent across domains, with negligible mediation through beliefs. Together, these findings provide experimental support for the proposed pathways while revealing an asymmetry in their relative influence, suggesting that informational processes may be more responsive to experimental manipulation than identity-related processes. Studies 4 and 5 extended this investigation by examining the potential for reciprocal influences among these psychological constructs.

Study 4 and Study 5

Studies 4 and 5 examined whether the pathways identified in Study 1 and experimentally tested in Studies 2 and 3 could also operate in the reverse direction. Because Bayesian networks impose an acyclic structure, the recovered model represents directional dependencies even when the underlying psychological processes may be reciprocal. Thus, the pathways identified in Study 1 leave open the possibility that policy attitudes may also influence upstream psychological constructs. To investigate this possibility, we experimentally manipulated information about the effectiveness of policies promoting four sustainable behaviors (reducing red meat consumption, adopting electric vehicles, reducing air travel, and using renewable electricity) and examined its effects on beliefs and knowledge (Study 4) and on beliefs and environmental identity (Study 5).

Although reciprocal influences were theoretically possible, we preregistered the expectation that these reverse pathways would be weaker or absent. Specifically, as preregistered on OSF (https://osf.io/mpjkg/overview?view_only=49d52c16fbca4229856322eb3a534dbd), we hypothesized that exposure to information about an effective sustainability policy would (H3a) not increase positive beliefs about the corresponding sustainable behavior, (H3b) not increase knowledge about that behavior, and (H3c) not produce an indirect effect on knowledge through beliefs. Similarly, we hypothesized that exposure to information about an effective sustainability policy would (H4a) not increase positive beliefs about the corresponding behavior, (H4b) not increase environmental identity, and (H4c) not produce an indirect effect on environmental identity through beliefs. The materials, data, analytic code, and output for these studies are available on OSF (https://osf.io/zsh5f/overview?view_only=ff94b89e3f0f4ea4b6dc2db07148ea71).

Method Study 4 and 5

Participants. The sample sizes targeted for Study 4 and Study 5 were identical to those targeted for Study 2 and Study 3. Participants were recruited from Prolific and were randomly assigned to complete either Study 4 or Study 5 for one of the four sustainable behaviours. They were paid at a rate equivalent to approximately \$16.50 per hour and were required to be 18 years of age or older and located in the US to participate. We used quota sampling to demographically match the samples to the 2020 US Census. After implementing our preregistered exclusion criteria, Study 4 included 1,993 participants (465 to 535 participants for each sustainable behaviour) and Study 5 included 2,002 participants (465 to 535 participants for each sustainable behaviour). Exclusion counts and select demographic information can be found in Table 5. The research protocol was reviewed and approved by the Stanford University Institutional Review Board.

Table 5

Select demographic information for the participants in Study 4 (Knowledge) and Study 5 (Identity).

	Study 4 (Knowledge)					Study 5 (Identity)				
	All	Diet	EVs	Air Travel	Renewa bles	All	Diet	EVs	Air Travel	Renewa bles
<i>Sample Size</i>										
Total	2004	466	537	507	494	2013	538	470	493	512
Exclusions – Instructed										
Response Items	11	1	2	4	4	11	3	5	2	1
Exclusions – Duration	0	0	0	0	0	0	0	0	0	0
Excluded – LLM										
Detection Items	0	0	0	0	0	0	0	0	0	0
Final	1993	465	535	503	490	2002	535	465	491	511
<i>Age Mean (SD)</i>	44.26 (15.69)	44.36 (15.26)	44.46 (16.09)	43.08 (15.71)	45.16 (15.61)	43.87 (15.66)	43.86 (15.84)	43.74 (15.61)	44.05 (15.60)	43.82 (15.60)
<i>Gender Count (%)</i>										
Women	1027 (51.53%)	241 (51.83%)	267 (49.91%)	268 (53.28%)	251 (51.22%)	974 (48.65%)	265 (49.53%)	219 (47.10%)	241 (49.08%)	249 (48.73%)
Men	922 (46.26%)	214 (46.02%)	251 (46.92%)	227 (45.13%)	230 (46.94%)	991 (49.50%)	262 (48.97%)	237 (50.97%)	236 (48.07%)	256 (50.10%)
Non-binary	29 (1.46%)	7 (1.51%)	11 (2.06%)	4 (0.80%)	7 (1.43%)	23 (1.15%)	5 (0.93%)	5 (1.08%)	9 (1.83%)	4 (0.78%)
Genderfluid	7 (0.35%)	1 (0.22%)	4 (0.75%)	2 (0.40%)	0 (0.00%)	4 (0.20%)	1 (0.19%)	1 (0.22%)	2 (0.41%)	0 (0.00%)
Other	5 (0.25%)	1 (0.22%)	2 (0.37%)	1 (0.20%)	1 (0.20%)	3 (0.15%)	0 (0.00%)	1 (0.22%)	1 (0.20%)	1 (0.20%)
Preferred not to answer	3 (0.15%)	1 (0.22%)	0 (0.00%)	1 (0.20%)	1 (0.20%)	7 (0.35%)	2 (0.37%)	2 (0.43%)	2 (0.41%)	1 (0.20%)
<i>Cultural/Ethnic Identity Count (%)</i>										
White or Caucasian	1210 (60.71%)	275 (59.14%)	327 (61.12%)	288 (57.26%)	320 (65.31%)	1235 (61.69%)	341 (63.74%)	275 (59.14%)	295 (60.08%)	324 (63.41%)
Black or African-American	235 (11.79%)	51 (10.97%)	53 (9.91%)	74 (14.71%)	57 (11.63%)	239 (11.94%)	58 (10.84%)	56 (12.04%)	63 (12.83%)	62 (12.13%)
Asian or Asian-American	108 (5.42%)	26 (5.59%)	28 (5.23%)	22 (4.37%)	32 (6.53%)	107 (5.34%)	27 (5.05%)	29 (6.24%)	22 (4.48%)	29 (5.68%)

Hispanic/Latinx	157 (7.88%)	43 (9.25%)	38 (7.10%)	50 (9.94%)	26 (5.31%)	153 (7.64%)	42 (7.85%)	42 (9.03%)	36 (7.33%)	33 (6.46%)
Multiple cultures/ethnicities	224 (11.24%)	49 (10.54%)	72 (13.46%)	57 (11.33%)	46 (9.39%)	202 (10.09%)	48 (8.97%)	49 (10.54%)	53 (10.79%)	52 (10.18%)
Other cultural/ethnic identity	48 (2.41%)	18 (3.87%)	14 (2.62%)	9 (1.79%)	7 (1.43%)	58 (2.90%)	17 (3.18%)	13 (2.80%)	19 (3.87%)	9 (1.76%)
Preferred not to answer	11 (0.55%)	3 (0.65%)	3 (0.56%)	3 (0.60%)	2 (0.41%)	8 (0.40%)	2 (0.37%)	1 (0.22%)	3 (0.61%)	2 (0.39%)
<i>Political Affiliation Count (%)</i>										
Democratic	847 (42.50%)	196 (42.15%)	231 (43.18%)	226 (44.93%)	194 (39.59%)	822 (41.06%)	207 (38.69%)	197 (42.37%)	215 (43.79%)	203 (39.73%)
Republican	441 (22.13%)	112 (24.09%)	112 (20.93%)	115 (22.86%)	102 (20.82%)	465 (23.23%)	130 (24.30%)	112 (24.09%)	95 (19.35%)	128 (25.05%)
Independent	607 (30.46%)	129 (27.74%)	167 (31.21%)	139 (27.63%)	172 (35.10%)	594 (29.67%)	167 (31.21%)	129 (27.74%)	149 (30.35%)	149 (29.16%)
Other political affiliation	31 (1.56%)	8 (1.72%)	12 (2.24%)	6 (1.19%)	5 (1.02%)	36 (1.80%)	11 (2.06%)	10 (2.15%)	7 (1.43%)	8 (1.57%)
Undeclared	67 (3.36%)	20 (4.30%)	13 (2.43%)	17 (3.38%)	17 (3.47%)	85 (4.25%)	20 (3.74%)	17 (3.66%)	25 (5.09%)	23 (4.50%)

Note. Additional demographic information (e.g., states of residence; educational attainment) can be found in the results document provided on OSF (https://osf.io/zsh5f/overview?view_only=ff94b89e3f0f4ea4b6dc2db07148ea71). Participants were excluded for failing both of the instructed response items embedded in the survey (Curran, 2016; Kay & Saucier, 2023); completing the survey in less than one-third of the median response time (Bedford-Petersen & Saucier, 2021); and failing three or more of the four ECLAIR LLM-detection items embedded in the survey (Kay & Vlasceanu, 2026).

All	0.72 (1.67)	1.16 (1.63)	5.96	1990	<.001	0.27	0.80 (1.74)	1.23 (1.64)	5.63	1979	<.001	0.25
Diet	0.30 (1.92)	0.59 (1.84)	1.64	461	.102	0.15	0.21 (1.85)	0.77 (1.74)	3.62	520	<.001	0.31
EV	0.57 (1.58)	1.00 (1.53)	3.20	533	.001	0.28	0.78 (1.66)	1.07 (1.68)	1.89	463	.060	0.18
Air Travel	0.81 (1.61)	1.28 (1.58)	3.29	496	.001	0.29	0.94 (1.71)	1.23 (1.59)	1.96	489	.050	0.18
Renewables	1.18 (1.43)	1.76 (1.37)	4.54	484	<.001	0.41	1.29 (1.56)	1.83 (1.34)	4.17	480	<.001	0.37

Beliefs. Beliefs were assessed in Study 4 and Study 5 in the same manner that they were assessed in Study 2 and Study 3. The four items were, again, shown to be reliable across the sixteen study-time-context combinations (α s = .80 to .90; ω s = .81 to .91).

Knowledge. In Study 4, we assessed baseline and post-manipulation knowledge about the four sustainable behaviours using the six-item manipulation check knowledge test used in Study 2. The six items were not shown to be reliable across the eight time-context combinations (α s = .35 to .52; ω s = .44 to .56). All of the results related to the knowledge should, therefore, be interpreted with caution.

Identity. In Study 5 we assessed baseline and post-manipulation environmental identity levels using the 14-item *Revised Environmental Identity Scale* (Clayton et al., 2021) we used as a manipulation check in Study 3. The 14 items were shown to be reliable across the eight time-context combinations (α s = .91 to .94; ω s = .91 to .93)

Procedure. The procedure used in Study 4 and Study 5 closely mirrored that of Study 2 and Study 3. However, instead of being randomly assigned to complete one of four surveys containing a knowledge manipulation (Study 2) or one of four surveys containing an identity manipulation (Study 3), participants were randomly assigned to complete one of four surveys including the baseline and post-treatment knowledge measure (Study 4) or one of the four surveys including the baseline and post-treatment identity measure (Study 5).

Consequently, participants in Study 4, completed, in this order, the baseline belief measure, baseline knowledge measure, policy-support manipulation induction, post-treatment belief measure, post-treatment knowledge measure, and policy-support induction manipulation checks. Participants in Study 5, in contrast, completed the baseline belief measure, baseline identity measure, policy-support manipulation induction, post-treatment belief measure, post-treatment identity measure, and policy-support induction manipulation checks. As in the prior studies, the four surveys for each study focused on one of four sustainable behaviours: (a) reducing consumption of red meat, increasing adoption of electric vehicles, reducing air travel, or increasing use of renewable electricity.

Results Study 4 and 5

The results of Study 4 suggest that people who were exposed to an effective policy were more likely to hold positive beliefs about the behaviours in the combined data ($\beta = 0.10$, 95% CI [0.06, 0.13], $p < .001$), with this result holding when tested specifically in the context of increasing adoption of EVs, reducing air travel, and increasing reliance on renewable energy sources (Table 7). Moreover, people who were exposed to an effective policy also became slightly more knowledgeable about sustainable behaviours ($\beta = 0.06$, 95% CI [0.02, 0.11], $p = .007$), although this effect only held in the specific context of increasing adoption of electric vehicles, and not for the other three contexts tested, suggesting a much weaker causal direction from policy to knowledge. Positive beliefs about these behaviours also mediated the effect of policy effectiveness on knowledge in the combined data (mediation supported by a significant indirect effect; Table 8), such that policy support appeared to contribute to knowledge because it increased a person's positive beliefs about the sustainable behaviours.

Study 5 further showed that, despite influencing a person's beliefs, policy support has no impact on one's environmental identity (Table 7). Namely, people who were exposed to an effective policy were more likely to hold positive beliefs about sustainable behaviours in the combined data ($\beta = 0.10$, 95% CI [0.07, 0.12], $p < .001$), with this result holding when each sustainable behaviour was analyzed separately, but exposure to an effective policy had no effect on environmental identity levels ($\beta = 0.00$, 95% CI [-0.02, 0.03], $p = .779$). Moreover, we found no evidence that positive beliefs mediated the impact of policy effectiveness on identity (Table 8).

Table 7

The effect of the policy manipulation on beliefs and knowledge in Study 4 and on beliefs and identity in Study 5

	Study 4: Does the policy manipulation affect beliefs and knowledge?						Study 5: Does the policy manipulation affect beliefs and identity?					
	Beliefs			Knowledge			Beliefs			Identity		
	β	95% CI	p	β	95% CI	p	β	95% CI	p	β	95% CI	p
All	0.10	[0.06, 0.13]	< .001	0.06	[0.02, 0.11]	.007	0.10	[0.07, 0.12]	< .001	0.00	[-0.02, 0.03]	.779
Diet	0.05	[-0.01, 0.11]	.130	0.00	[-0.08, 0.08]	.99	0.10	[0.05, 0.16]	< .001	0.02	[-0.03, 0.06]	.429
EV	0.07	[0.00, 0.14]	.039	0.10	[0.02, 0.21]	.01	0.10	[0.07, 0.17]	< .001	-0.00	[-0.05, 0.04]	.881
Air Travel	0.06	[0.08, 0.25]	< .001	0.00	[-0.03, 0.15]	.21	0.10	[0.07, 0.20]	< .001	-0.00	[-0.05, 0.05]	.954
Renewables	0.01	[0.04, 0.19]	.003	0.00	[-0.01, 0.18]	.09	0.00	[0.01, 0.13]	.015	-0.00	[-0.05, 0.05]	.965

Note. Baseline beliefs were included as covariates in the belief models, baseline levels of knowledge were included as covariates in the knowledge models, and baseline environmental identity levels were included as covariates in the identity models.

Table 8

Beliefs as a mediator of the effect of policy effectiveness on knowledge (Study 4) and identity (Study 5)

	Study 4: Does belief mediate the effect of policy effectiveness on knowledge?			Study 5: Does belief mediate the effect of policy effectiveness on identity?		
	β	95% CI	<i>p</i>	β	95% CI	<i>p</i>
All	0.01	[0.00, 0.01]	.001	0.00	[0.00, 0.00]	.019
Diet	0.00	[-0.00, 0.01]	.211	0.00	[-0.00, 0.00]	.769
EV	0.00	[0.00, 0.01]	.126	0.00	[-0.00, 0.00]	.221
Air Travel	0.01	[0.00, 0.03]	.036	0.00	[-0.00, 0.01]	.658
Renewables	0.01	[0.00, 0.02]	.034	0.00	[0.00, 0.01]	.035

Note. Beta is the standardized product of the (a) policy to beliefs path and (b) the beliefs to knowledge/identity path. Standard errors were estimated using bootstrapping with 10,000 iterations. Baseline beliefs were added as predictors of post-treatment beliefs, baseline levels of knowledge were added as predictors of post-treatment knowledge, and baseline environmental identity levels were added as predictors of post-treatment identity.

Discussion Study 4 and 5

In Studies 4 and 5, we experimentally examined the potential for reciprocal relationships within the cognitive pathways identified in Study 1 by manipulating information about policy effectiveness (Study 4, $N = 1,993$; Study 5, $N = 2,002$) across four sustainability domains. In both studies, exposure to information about effective policies increased positive beliefs about the corresponding sustainable behaviors, suggesting that beliefs can be updated through multiple sources of information within the psychological system. Study 4 also found weaker and less consistent effects on knowledge, whereas Study 5 found no significant effect on environmental identity. Together, these findings suggest that beliefs can change in response to information about policy effectiveness, whereas knowledge is less responsive and environmental identity appears relatively stable.

These experimental findings help clarify the distinct roles of knowledge, identity, and beliefs within the cognitive system. Beliefs appear particularly responsive to experimental manipulation, changing following increases in factual knowledge, activation of environmental identity, and exposure to information about policy effectiveness. Knowledge showed substantial effects on beliefs and policy support when experimentally increased, whereas policy-effectiveness information produced comparatively limited changes in knowledge. Environmental identity showed the greatest apparent stability, with small downstream effects following identity activation and no detectable change following exposure to policy-effectiveness information. These findings suggest that knowledge and identity may serve as relatively upstream inputs to belief formation, while beliefs integrate information from multiple psychological and contextual sources.

More broadly, these findings illustrate the value of combining bottom-up probabilistic modeling with experimental validation. Bayesian networks recover candidate pathways of conditional dependence, but their acyclic structure cannot represent reciprocal relationships between variables. By experimentally testing both the predicted and reverse pathways, Studies 2–5 reveal a cognitive system characterized by asymmetric and potentially reciprocal influences. The resulting evidence refines the architecture identified in Study 1, highlighting beliefs as a central and comparatively flexible component connecting informational and identity-related processes to sustainability decisions.

General Discussion

Across five studies ($N = 11,994$), we combined bottom-up Bayesian network modeling with preregistered experiments to investigate how interacting psychological processes shape sustainability decisions. In Study 1 ($N = 3,993$), we estimated a Bayesian network incorporating a broad range of cognitive, affective, and social constructs across four sustainability domains, identifying two primary pathways linking knowledge and environmental identity to policy support and behavioral commitments through beliefs. In Studies 2 ($N = 2,041$) and 3 ($N = 1,965$), we experimentally tested these pathways, finding that increasing knowledge substantially strengthened beliefs and policy support, whereas activating environmental identity produced smaller and less consistent effects. In Studies 4 ($N = 1,993$) and 5 ($N = 2,002$), we examined potential reverse pathways, finding that information about policy effectiveness changed beliefs, produced weaker effects on knowledge, and had no detectable effect on environmental identity. Together, these findings identify beliefs as a central and responsive component of sustainability decisions and demonstrate how probabilistic graphical models can generate candidate cognitive architectures that can be refined through experimental testing.

One implication of these findings concerns the longstanding debate over informational and identity-based explanations of attitude formation. Information deficit accounts argue that policy attitudes change when individuals acquire more accurate knowledge, which alters their beliefs about the consequences or effectiveness of proposed actions (Coppock, 2023; Pennycook & Rand, 2019). By contrast, motivated reasoning accounts emphasize that beliefs are filtered through prior identities and motivations, leading individuals to interpret information in identity-consistent ways (Kunda, 1990; Van Bavel et al., 2024). The cognitive pathways recovered here predicted pathways consistent with both perspectives: knowledge and identity each influenced policy support indirectly through belief formation. Experimental validation further showed that interventions targeting these pathways differed substantially in their effectiveness. Increasing knowledge produced larger changes in beliefs and policy support than activating environmental identity. The magnitude of an intervention's downstream effects depends both on the influence of the targeted construct and on how successfully that construct can be manipulated. Indeed, manipulation checks indicated that the knowledge intervention produced a larger shift in its targeted construct than the identity intervention, making it difficult to isolate differences in pathway strength from differences in manipulability. Thus, the

experiments provide stronger evidence that knowledge represents a more readily actionable intervention target in this context, yielding greater downstream change under the manipulations tested, rather than that knowledge necessarily exerts greater causal influence than identity when equivalent changes in the two constructs are achieved. More broadly, the results illustrate how bottom-up computational models can move theoretical debates beyond qualitative comparisons by generating explicit predictions about competing psychological pathways that can be prospectively evaluated through experimentation, while also highlighting the importance of distinguishing the strength of a psychological pathway from the effectiveness of interventions designed to engage it.

More generally, the present work demonstrates how bottom-up probabilistic modeling can complement traditional top-down approaches to psychological theory. Rather than specifying a cognitive pathway a priori, the Bayesian network recovered the organization of relationships among a broad set of psychological variables while incorporating theoretically motivated constraints. The resulting architecture identified beliefs as a central hub linking multiple upstream cognitive constructs to downstream policy support and behavioral commitments. Importantly, the value of the framework lies in recovering candidate cognitive pathways capable of generating novel theoretical predictions. Accordingly, computational modeling shifts from describing psychological systems to proposing mechanistic hypotheses that can subsequently be subjected to experimental evaluation.

The four preregistered experiments illustrate this iterative process of theory development. The architecture recovered from observational data predicted that knowledge and environmental identity would influence policy support through belief formation, with a stronger pathway originating in knowledge. Experimental tests supported this general architecture while also revealing important asymmetries in the direction of influence. Increasing knowledge consistently produced substantial changes in beliefs and policy support, whereas activating environmental identity produced smaller and less consistent downstream effects. Reversing these pathways showed that policy information could also change beliefs, indicating that beliefs are responsive to influence from multiple points in the system. However, evidence for further upstream effects was considerably weaker: policy information produced only small and inconsistent changes in knowledge and no detectable change in environmental identity. Thus, the network captured the dominant direction of influence while the experiments revealed that the underlying architecture is not strictly unidirectional. Together, these findings illustrate how Bayesian networks can identify psychologically meaningful candidate structures from observational data while experimental tests can establish causality, distinguish stronger from weaker pathways, and reveal where reciprocal influence occurs.

Importantly, the network also generated predictions about the organization of indirect relationships among psychological constructs. Social norms and negative emotional responses were connected to policy support through beliefs and other upstream cognitive constructs, suggesting that their influence may be mediated by these processes. This interpretation builds on prior research demonstrating causal effects of norms and emotions on environmental attitudes

and behavior (Cialdini & Goldstein, 2004; Schultz et al., 2007; Smith & Leiserowitz, 2014; Lerner et al., 2015). The present framework extends these findings by identifying potential cognitive pathways through which norms and emotions shape policy preferences, including their relationships with beliefs about consequences, social expectations, and behavioral effectiveness (Constantino et al., 2022; Lerner et al., 2015; van der Linden et al., 2015). More broadly, Bayesian networks generate explicit hypotheses about both direct and mediated relationships among psychological constructs, providing a basis for experimentally investigating how these processes jointly shape decisions.

Our findings also contribute to the broader movement toward formal psychological theory. Psychological theories are often expressed through conceptual diagrams or verbal descriptions that leave important assumptions underspecified (Urai et al., 2025). Probabilistic graphical models instead require those assumptions to be represented explicitly within a computational framework, allowing alternative cognitive architectures to be compared, refined, and falsified. Rather than replacing traditional theory, Bayesian networks complement it by providing a bottom-up method for discovering candidate pathways that can subsequently be incorporated into cumulative theory development. Moreover, the framework integrates constructs that have traditionally been studied within separate literatures, including beliefs, emotions, identity, social norms, trust, moral values, and political ideology, within a single probabilistic system. This systems perspective shifts attention away from identifying the single strongest predictor of behavior and toward understanding how multiple psychological processes interact to generate attitudes and behavioral commitments.

Several limitations to the present work should be acknowledged. First, although Bayesian networks recover conditional dependency structures consistent with candidate causal architectures, observational data alone cannot establish causal relationships. Experimental and longitudinal evidence therefore remain essential for evaluating the pathways suggested by the recovered pathways. Second, the architecture identified here reflects sustainability decisions in a U.S. sample and may differ across domains (Vlasceanu et al., 2024), cultures (Goldwert et al., 2024), or political contexts (Berkebile-Weinberg et al., 2024). Future work should examine the extent to which cognitive pathways generalize across populations and decision domains, or whether distinct models emerge under different informational and social environments. Finally, the present model represents a static cognitive model. Future work could extend this framework to incorporate temporal dynamics (Vlasceanu et al., 2020; Vlasceanu et al., 2021), reciprocal feedback processes, and multilevel systems linking individual cognition with interpersonal, organizational, and institutional processes (Vlasceanu & Constantino, 2026).

In conclusion, the present findings illustrate a framework for developing psychological theories of interacting cognitive processes. Bottom-up probabilistic models allow candidate cognitive architectures to be recovered from observational data, translated into experimentally testable predictions, and iteratively refined through causal evidence. By integrating computational modeling with experimental validation, this approach provides a foundation for

developing more formal, cumulative, and mechanistic theories of how interacting psychological processes give rise to complex judgments and behaviors.

References

- Ajzen, I. (1991). The theory of planned behavior. *Organizational Behavior and Human Decision Processes*, 50(2), 179–211. [https://doi.org/10.1016/0749-5978\(91\)90020-T](https://doi.org/10.1016/0749-5978(91)90020-T)
- Bamberg, S., & Möser, G. (2007). Twenty years after Hines, Hungerford, and Tomera: A new meta-analysis of psycho-social determinants of pro-environmental behaviour. *Journal of Environmental Psychology*, 27(1), 14–25. <https://doi.org/10.1016/j.jenvp.2006.12.002>
- Bao, E., & Vlasceanu, M. (2026). The Impact of Emotional and Cognitive Pathways on the Support and Priority of Climate Policy. *Journal of Experimental Psychology: Applied*.
- Bedford-Petersen, C., & Saucier, G. (2021). Identifying contrasting themes that orchestrate personality expression across situations. *Personality and Individual Differences*, 171, 110495. <https://doi.org/10.1016/j.paid.2020.110495>
- Bello, K., Aragam, B., & Ravikumar, P. (2022). DAGMA: Learning DAGs via M-matrices and a log-determinant acyclicity characterization. *Advances in Neural Information Processing Systems*, 35, 8226–8239.
- Berkebile-Weinberg, M., Goldwert, D., Doell, K. C., Van Bavel, J. J., & Vlasceanu, M. (2024). The differential impact of climate interventions along the political divide in 60 countries. *Nature Communications*, 15, 3885. <https://doi.org/10.1038/s41467-024-48112-8>
- Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry*, 16(1), 5–13. <https://doi.org/10.1002/wps.20375>
- Boutyline, A., & Vaisey, S. (2017). Belief network analysis: A relational approach to understanding the structure of attitudes. *American Journal of Sociology*, 122(5), 1371–1447. <https://doi.org/10.1086/691274>
- Brandt, M. J., & Sleegers, W. W. A. (2021). Evaluating belief system networks as a theory of political belief system dynamics. *Personality and Social Psychology Review*, 25(2), 159–185. <https://doi.org/10.1177/1088868321993751>
- Cialdini, R. B., & Goldstein, N. J. (2004). Social influence: Compliance and conformity. *Annual Review of Psychology*, 55, 591–621. <https://doi.org/10.1146/annurev.psych.55.090902.142015>
- Clayton, S., Czellar, S., Nartova-Bochaver, S., Skibins, J. C., Salazar, G., Tseng, Y.-C., Irkhin, B., & Monge-Rodriguez, F. S. (2021). Cross-cultural validation of a revised environmental identity scale. *Sustainability*, 13(4), 2387. <https://doi.org/10.3390/su13042387>
- Clayton, S., Devine-Wright, P., Stern, P. C., Whitmarsh, L., Carrico, A., Steg, L., Swim, J., & Bonnes, M. (2015). Psychological research and global climate change. *Nature Climate Change*, 5(7), 640–646. <https://doi.org/10.1038/nclimate2622>
- Constantino, S. M., Sparkman, G., Kraft-Todd, G. T., Bicchieri, C., Centola, D., Shell-Duncan, B., Vogt, S., & 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. <https://doi.org/10.1177/15291006221105279>

- Converse, P. E. (2006). The nature of belief systems in mass publics (1964). *Critical Review*, 18(1–3), 1–74. <https://doi.org/10.1080/08913810608443650>
- Coppock, A. (2023). *Persuasion in parallel: How information changes minds about politics*. University of Chicago Press.
- Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology*, 66, 4–19. <https://doi.org/10.1016/j.jesp.2015.07.006>
- Dalege, J., Borsboom, D., van Harreveld, F., van den Berg, H., Conner, M., & van der Maas, H. L. J. (2016). Toward a formalized account of attitudes: The Causal Attitude Network (CAN) model. *Psychological Review*, 123(1), 2–22. <https://doi.org/10.1037/a0039802>
- Dalege, J., Galesic, M., & Olsson, H. (2025). Networks of beliefs: An integrative theory of individual- and social-level belief dynamics. *Psychological Review*, 132(2), 253–290. <https://doi.org/10.1037/rev0000494>
- de Gourville, D., Douglas, K. M., & Sutton, R. M. (2026). Denialist vs. warmist climate change conspiracy beliefs: Ideological roots, psychological correlates and environmental implications. *British journal of psychology*, 117, S33-S64. <https://doi.org/10.1111/bjop.70035>
- Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316272503>
- Fielding, K. S., & Hornsey, M. J. (2016). A social identity analysis of climate change and environmental attitudes and behaviors: Insights and opportunities. *Frontiers in Psychology*, 7, 121. <https://doi.org/10.3389/fpsyg.2016.00121>
- Fried, E. I., & Cramer, A. O. J. (2017). Moving forward: Challenges and directions for psychopathological network theory and methodology. *Perspectives on Psychological Science*, 12(6), 999–1020. <https://doi.org/10.1177/1745691617705892>
- Galli, G., Sirota, M., Gruber, M. J., Ivanof, B. E., Ganesh, J., Materassi, M., Thorpe, A., Loaiza, V., Cappelletti, M., & Craik, F. I. M. (2018). Learning facts during aging: The benefits of curiosity. *Experimental Aging Research*, 44(4), 311–328. <https://doi.org/10.1080/0361073X.2018.1477355>
- Geiger, S. M., Geiger, M., & Wilhelm, O. (2019). Environment-specific vs. general knowledge and their role in pro-environmental behavior. *Frontiers in Psychology*, 10, 718. <https://doi.org/10.3389/fpsyg.2019.00718>
- Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of Consumer Research*, 35(3), 472–482. <https://doi.org/10.1086/586910>
- Goldwert, D., Bao, Y. E., Doell, K. C., Van Bavel, J. J., & Vlasceanu, M. (2024). The effects of climate action interventions along cultural individualism-collectivism. *PsyArXiv*. <https://doi.org/10.31234/osf.io/cv3n4>

- Graham, J., Haidt, J., Koleva, S., Motyl, M., Iyer, R., Wojcik, S. P., & Ditto, P. H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism. *Advances in Experimental Social Psychology*, 47, 55–130. <https://doi.org/10.1016/B978-0-12-407236-7.00002-4>
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96(5), 1029–1046. <https://doi.org/10.1037/a0015141>
- Griffiths, T. L., Chater, N., Kemp, C., Perfors, A., & Tenenbaum, J. B. (2010). Probabilistic models of cognition: Exploring representations and inductive biases. *Trends in Cognitive Sciences*, 14(8), 357–364. <https://doi.org/10.1016/j.tics.2010.05.004>
- Guest, O., & Martin, A. E. (2021). How computational modeling can force theory building in psychological science. *Perspectives on Psychological Science*, 16(4), 789–802. <https://doi.org/10.1177/1745691620970585>
- Hornsey, M. J., Harris, E. A., Bain, P. G., & Fielding, K. S. (2016). Meta-analyses of the determinants and outcomes of belief in climate change. *Nature Climate Change*, 6(6), 622–626. <https://doi.org/10.1038/nclimate2943>
- Jones, M., & Love, B. C. (2011). Bayesian fundamentalism or enlightenment? On the explanatory status and theoretical contributions of Bayesian models of cognition. *Behavioral and Brain Sciences*, 34(4), 169–188. <https://doi.org/10.1017/S0140525X10003134>
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.
- Kay, C. S., & Saucier, G. (2023). The Comprehensive Infrequency/Frequency Item Repository (CIFR): An online database of items for detecting careless/insufficient-effort responders in survey data. *Personality and Individual Differences*, 205, 112073. <https://doi.org/10.1016/j.paid.2022.112073>
- Kay, C. S., & Vlasceanu, M. (2026). A scale for detecting LLM-generated responses in online survey research. *PsyArXiv*. https://doi.org/10.31234/osf.io/4p7ns_v5
- Kenny, D. A. (2017). MedPower: An interactive tool for the estimation of power in tests of mediation. <https://davidakenny.shinyapps.io/MedPower/>
- Klößner, C. A. (2013). A comprehensive model of the psychology of environmental behaviour—A meta-analysis. *Global Environmental Change*, 23(5), 1028–1038. <https://doi.org/10.1016/j.gloenvcha.2013.05.014>
- Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: Principles and techniques*. MIT Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology*, 66, 799–823. <https://doi.org/10.1146/annurev-psych-010213-115043>
- Mason, K. A., Vlasceanu, M., & Jost, J. T. (2024). Effects of system-sanctioned framing on climate awareness and environmental action in the United States and beyond.

- Proceedings of the National Academy of Sciences, 121(38), e2405973121.
<https://doi.org/10.1073/pnas.2405973121>
- McCright, A. M., & Dunlap, R. E. (2011). The politicization of climate change and polarization in the American public's views of global warming, 2001–2010. *The Sociological Quarterly*, 52(2), 155–194. <https://doi.org/10.1111/j.1533-8525.2011.01198.x>
- Miller, A. L., Whatley, M., Chen, A., & Castel, A. D. (in prep). Curiosity as a cognitive shield: Reducing mind-wandering's harmful effects on learning and memory.
- Pearl, J. (1988). Probabilistic reasoning in intelligent systems: Networks of plausible inference. Morgan Kaufmann. <https://doi.org/10.1016/C2009-0-27609-4>
- Pearl, J. (2009). Causality: Models, reasoning, and inference (2nd ed.). Cambridge University Press. <https://doi.org/10.1017/cbo9780511803161>
- Pennycook, G., & Rand, D. G. (2019). Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition*, 188, 39–50. <https://doi.org/10.1016/j.cognition.2018.06.011>
- Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of causal inference: Foundations and learning algorithms. MIT Press.
- Powell, D., Weisman, K., & Markman, E. M. (2023). Modeling and leveraging intuitive theories to improve vaccine attitudes. *Journal of Experimental Psychology: General*, 152(5), 1379–1395. <https://doi.org/10.1037/xge0001324>
- Sabherwal, A., Goldwert, D., & Vlasceanu, M. (2026). Emotional complexity drives collective climate action. PsyArXiv. https://osf.io/preprints/psyarxiv/24jpe_v2
- Sanborn, A. N., & Chater, N. (2016). Bayesian brains without probabilities. *Trends in Cognitive Sciences*, 20(12), 883–893. <https://doi.org/10.1016/j.tics.2016.10.003>
- Schultz, P. W., Nolan, J. M., Cialdini, R. B., Goldstein, N. J., & Griskevicius, V. (2007). The constructive, destructive, and reconstructive power of social norms. *Psychological Science*, 18(5), 429–434. <https://doi.org/10.1111/j.1467-9280.2007.01917.x>
- Smaldino, P. E. (2017). Models are stupid, and we need more of them. In R. R. Vallacher, S. J. Read, & A. Nowak (Eds.), *Computational social psychology* (pp. 311–331). Routledge.
- Smith, E. K., & Mayer, A. (2018). A social trap for the climate? Collective action, trust and climate change risk perception in 35 countries. *Global Environmental Change*, 49, 140–153. <https://doi.org/10.1016/j.gloenvcha.2018.02.014>
- Smith, N., & Leiserowitz, A. (2014). The role of emotion in global warming policy support and opposition. *Risk Analysis*, 34(5), 937–948. <https://doi.org/10.1111/risa.12140>
- Sparkman, G., & Walton, G. M. (2017). Dynamic norms promote sustainable behavior, even if it is counternormative. *Psychological Science*, 28(11), 1663–1674. <https://doi.org/10.1177/0956797617719950>
- Spirtes, P., Glymour, C., & Scheines, R. (2001). Causation, prediction, and search (2nd ed.). MIT Press.
- Steg, L. (2023). Psychology of climate change. *Annual Review of Psychology*, 74, 391–421. <https://doi.org/10.1146/annurev-psych-032720-042905>

- Steg, L., & Vlek, C. (2009). Encouraging pro-environmental behaviour: An integrative review and research agenda. *Journal of Environmental Psychology*, 29(3), 309–317. <https://doi.org/10.1016/j.jenvp.2008.10.004>
- Stern, P. C. (2002). New environmental theories: Toward a coherent theory of environmentally significant behavior. *Journal of Social Issues*, 56(3), 407–424. <https://doi.org/10.1111/0022-4537.00175>
- Stern, P. C., Dietz, T., Abel, T., Guagnano, G. A., & Kalof, L. (1999). A value-belief-norm theory of support for social movements: The case of environmentalism. *Human ecology review*, 81-97.
- Tenenbaum, J. B., Kemp, C., Griffiths, T. L., & Goodman, N. D. (2011). How to grow a mind: Statistics, structure, and abstraction. *Science*, 331(6022), 1279–1285. <https://doi.org/10.1126/science.1192788>
- Turner-Zwinkels, F. M., & Brandt, M. J. (2022). Belief system networks can be used to predict where to expect dynamic constraint. *Journal of Experimental Social Psychology*, 100, 104279. <https://doi.org/10.1016/j.jesp.2021.104279>
- Urai, A. E., Dablander, F., & Bolderdijk, J. W. (2025). Impactful environmental psychology needs formal theories. *Current Opinion in Behavioral Sciences*, 66, 101619.
- Van Bavel, J. J., Rathje, S., Vlasceanu, M., & Pretus, C. (2024). Updating the identity-based model of belief: From false belief to the spread of misinformation. *Current Opinion in Psychology*, 56, 101787. <https://doi.org/10.1016/j.copsyc.2023.101787>
- van der Linden, S., Maibach, E., & Leiserowitz, A. (2015). Improving public engagement with climate change: Five “best practice” insights from psychological science. *Perspectives on Psychological Science*, 10(6), 758–763. <https://doi.org/10.1177/1745691615598516>
- Vlasceanu, M., & Constantino, S. M. (2026). Mapping psychological processes governing the sociopolitical climate system. https://doi.org/10.31234/osf.io/93rm4_v2
- Vlasceanu, M., Doell, K. C., Bak-Coleman, J. B., Todorova, B., Berkebile-Weinberg, M. M., Grayson, S. J., Patel, Y., Goldwert, D., Pei, Y., Chakroff, A., Pronizius, E., van den Broek, K. L., Vlasceanu, D., Constantino, S., Morais, M. J., Schumann, P., Rathje, S., Fang, K., Aglioti, S. M., . . . Van Bavel, J. J. (2024). Addressing climate change with behavioral science: A global intervention tournament in 63 countries. *Science Advances*, 10(6), ead5778. <https://doi.org/10.1126/sciadv.ad5778>
- Vlasceanu, M., Morais, M. J., & Coman, A. (2021). Network structure impacts the synchronization of collective beliefs. *Journal of Cognition and Culture*, 21(5), 431–448. <https://doi.org/10.1163/15685373-12340120>
- Vlasceanu, M., Morais, M. J., Duker, A., & Coman, A. (2020). The synchronization of collective beliefs: From dyadic interactions to network convergence. *Journal of Experimental Psychology: Applied*, 26(3), 453–464. <https://doi.org/10.1037/xap0000265>
- Vlasceanu, M., & van der Linden, S. (Eds.). (2026). *The psychological foundations of climate solutions: From individual minds to collective climate action*. Cambridge University Press.

- Whitmarsh, L., & O'Neill, S. (2010). Green identity, green living? The role of pro-environmental self-identity in determining consistency across diverse pro-environmental behaviours. *Journal of Environmental Psychology*, 30(3), 305–314. <https://doi.org/10.1016/j.jenvp.2010.01.003>
- Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, 67, 289–314. <https://doi.org/10.1146/annurev-psych-122414-033417>

Supplement

Table S1. Constructs measured in Study 1, with example wording and response scales.

Construct (network node)	Example item	Response scale
Media Exposure	“To what degree do you consume mass media?”; “To what degree do you use social media?”	0–100 slider (Not at all – Extremely frequently)
Domain Exposure	“To what degree do you encounter information about [domain]?”	0–100 slider (Not at all – Extremely frequently)
Habit	“To what degree do you eat red meat (e.g. beef, lamb, pork)?”	0–100 slider; EV/Flying/Renewables item binary
Knowledge	“The production of 1 Calorie from beef takes how much more or less land compared to the production of 1 Calorie from peas?” (see Table S2)	Domain-specific sliders; scored for accuracy
Beliefs	“To what extent do you think [behavior] will improve the environment?”	0–100 slider (Not at all – Extremely so)
Static Norms	“To what extent do you think Americans are [performing behavior]?”	0–100 slider (0% – 100% of Americans)
Dynamic Norms	“To what extent do you think Americans are starting to [perform behavior] more?”	0–100 slider (0% – 100% of Americans)
Negative Emotion	“When it comes to [behavior], how strongly do you experience the following emotions?:” Anger, Sadness, Fear, Guilt, Disappointment, Anxiety, Disgust	0–100 slider (Not at all – Extremely)
Positive Emotion	Same stem; Hope, Pride, Joy	0–100 slider (Not at all – Extremely)
Intrinsic Motivation	“I attempt to [perform behavior] because it is personally important to me.”	0–100 slider (Strongly disagree – Strongly agree)
Extrinsic Motivation	“I [perform behavior] because of pressure from others.”	0–100 slider (Strongly disagree – Strongly agree)
Environmental Identity	“To what degree do you identify with someone who cares about the natural environment?”	0–100 slider (Not at all – Very much so)
Social Identity	“To what degree are the following factors important to who you are:” similarity with in-group; family nationality; sense of difference; rebelliousness	0–100 slider (Not at all – Very much so)
Political Ideology	Political orientation on social issues; on economic issues	0–100 slider (Extremely liberal – Extremely conservative)

Social Dominance Orientation (SDO)	“Some groups of people are simply inferior to other groups”; “Group equality should be our ideal” (R); “No one group should dominate in society” (R)	0–100 slider (Not at all – Very much so)
Trust	Trust in mass media; in government; in scientists	0–100 slider (Not at all – Very much so)
Individualizing Morals	“Compassion for those who are suffering is the most crucial virtue”; “When the government makes laws, the number one principle should be ensuring that everyone is treated fairly”	0–100 slider (Not at all – Very much so)
Binding Morals	“It is more important to be a team player than to express oneself”; “Respect for authority is something all children need to learn”; “People should not do things that are disgusting, even if no one is harmed”	0–100 slider (Not at all – Very much so)
Policy Support	Diet: farmed animal tax; vegetarian options in public cafeterias; investment in protein crops. EV: fossil fuel vehicle tax; charging infrastructure; battery technology. Flying: carbon tax on air travel; alternatives to business flights; train infrastructure. Renewables: carbon tax; solar tax credits; renewable technology investment	0–100 slider (Not at all – Extremely so)
Commitment	Diet: reducing red meat consumption; substituting plant proteins. EV: choosing EV rideshare; switching to an EV. Flying: one less flight a year; train or bus over flying. Renewables: installing solar panels; opting into renewable electricity	0–100 slider (Definitely not – Absolutely yes)
Age	“How old are you?”	Open numeric
Gender	“What is your gender?”	Male / Female / Non-binary or other / Prefer not to say
Education	“How many years of formal education have you completed?”	4 levels (0–6 to 17+ years)
Income	“What is your total yearly family/household income?”	8 levels (<\$10,000 to \$200,000+)
Political Party	“What political party do you identify with?”	Democratic / Republican / Other