PHILOSOPHICAL TRANSACTIONS B

royalsocietypublishing.org/journal/rstb

Research

Cite this article: Kashima Y, Perfors A, Ferdinand V, Pattenden E. 2021 Ideology, communication and polarization. *Phil. Trans. R. Soc. B* **376**: 20200133. https://doi.org/10.1098/rstb.2020.0133

Accepted: 2 November 2020

One contribution of 18 to a theme issue 'The political brain: neurocognitive and computational mechanisms'.

Subject Areas:
behaviour, cognition

Keywords:
cultural dynamics, opinion dynamics, social network dynamics, communication

Author for correspondence:
Yoshihisa Kashima
e-mail: ykashima@unimelb.edu.au

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.5252427.

THE ROYAL SOCIETY PUBLISHING

Ideology, communication and polarization

Yoshihisa Kashima, Andrew Perfors, Vanessa Ferdinand and Elle Pattenden

Melbourne School of Psychological Sciences, The University of Melbourne, Parkville, Victoria 3010, Australia

YK, 0000-0003-3627-3273

Ideologically committed minds form the basis of political polarization, but ideologically guided communication can further entrench and exacerbate polarization depending on the structures of ideologies and social network dynamics on which cognition and communication operate. Combining a well-established connectionist model of cognition and a well-validated computational model of social influence dynamics on social networks, we develop a new model of ideological cognition and communication on dynamic social networks and explore its implications for ideological political discourse. In particular, we explicitly model ideologically filtered interpretation of social information, ideological commitment to initial opinion, and communication on dynamically evolving social networks, and examine how these factors combine to generate ideologically divergent and polarized political discourse. The results show that ideological interpretation and commitment tend towards polarized discourse. Nonetheless, communication and social network dynamics accelerate and amplify polarization. Furthermore, when agents sever social ties with those that disagree with them (i.e. structure their social networks by homophily), even non-ideological agents may form an echo chamber and form a cluster of opinions that resemble an ideological group.

This article is part of the theme issue 'The political brain: neurocognitive and computational mechanisms'.

1. Introduction

From Brexit to climate change, polarized political discourse is ubiquitous in democracies across the world. At one level, it is not too surprising that there are diverse opinions about how to meet ecological and geopolitical challenges such as global warming and the changing balance of international powers. After all, humans respond to the natural and human-made environment by shaping their world through collective adaptation [1,2]. When the environment is changing dynamically, sorting out a vision for our collective future via political discourse is an inevitable part of this adaptation process.

But, does vigorous political discourse always polarize? Here we define political discourse widely, to encompass all forms of political communication within a population, ranging from individuals to groups and everyday conversations to elite political discussions. Such discourse is, fundamentally, *human collective information processing* whose goal is to solve collective adaptation problems. Different configurations of collective information processing can generate divergent opinion dynamics (e.g. polarization or consensualization) under different circumstances.

Our main aim is to present a computational behavioural science approach that enables us to evaluate the contributions of different factors on polarized political discourse. We show that polarization is affected by communicative as well as cognitive factors, and that they interact with each other in non-obvious ways.

2. Background to computational modelling of public opinion dynamics

Around half a century ago, a computational behavioural science of collective information processing emerged in an attempt to understand the polarization of public opinion on fluoridation—the use of fluoride to reduce tooth decay

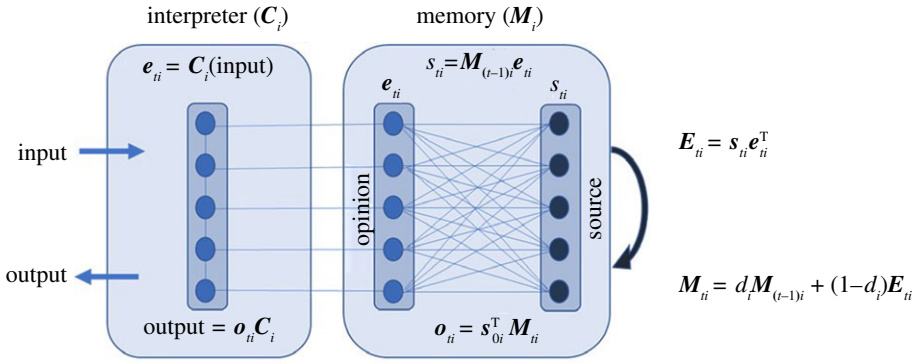

Figure 1. A schematic of the psychological model of ideological thinking. Note. Each agent consists of an interpreter and a memory. A circle is a cognitive unit that represents a proposition; the figure assumes five propositions constitute an ideology. Activation of each unit indicates the level of agreement that the agent has for the corresponding proposition. C_i represents the associations between the propositions for that agent, while $M_{(t-1)i}$ represents the pre-existing associations between opinions and sources. e_{ti} = the agent's interpreted input at time t; s_t = the agent's source attribution at time t; M_{ti} = their memory representation at time t; E_{ti} = their episodic memory that binds the attributed source and interpreted input; o_{ti} = the agent's opinion retrieved by cuing memory by self-representation, s_{0i}; d_i = a weight given to the prior memory; T designates transposition. (Online version in colour.)

in public water supply—in the United States [3,4]. A computational model was developed following the then available theories of social influence [4,5]. It assumed that individuals process each other's expressed opinions, recognize discrepancies in opinion, and modify their opinions to reduce the discrepancies. Ironically, this model showed public opinions should never polarize, but always consensualize. This led Abelson, a pioneer of computational behavioural science, to famously decry 'what on earth one must assume in order to generate the bimodal' [4, p. 153] distribution of public opinion!

Subsequent computational models identified several mechanisms that generate polarized public opinion. Flache and his colleagues' recent review [6] provides a useful overview. For example, clusters of opinions are likely to emerge if interacting individuals have commitments to their initial opinions [7] or if their opinions are only modified by agents whose opinions are sufficiently similar to their own [8]. Polarization is also likely if opposing opinions have a psychologically repulsive impact [9,10]. Although there are numerous variations on the themes, this voluminous work has compellingly showed that cognition and communication matter as critical drivers of discursive dynamics.

However, these computational models have two general drawbacks. First, they are typically about a single attitudinal issue: public opinions are assumed to vary on a single dimension of societal concern. However, much of contemporary public discourse is ideologically informed. That is, multiple issues are conceptually interconnected and thus constitute an overall *system* of beliefs [11]. One example is neoliberalism. It is not a unidimensional proposition about the virtue of a market economy; it also includes political, social and even psychological dimensions [12,13]. A second drawback is that these computational models tend to conflate the cognitive process of interpretation and the motivational process of ego-involvement. As we will show later, these processes can be modelled separately and can shed light on the ongoing discussion about cognitive and motivational accounts of polarization (e.g. [14–16]).

In this paper, we develop a framework that enables us to model ideological cognition and communication on multiple opinion dimensions by integrating a psychologically grounded connectionist approach to cognition [17–19] and a network approach to communication and social influence [20,21]. This allows us to identify the aspects of ideological cognition, communication and socio-technological context that may modulate the process of polarization. We find that polarization occurs most often when individuals interpret ambiguous information through an ideological lens, their identity is tied to their ideology, and they can choose whom to communicate with. Moreover, these factors interact in complex ways.

3. A connectionist model of the ideological mind

Within the cultural dynamics framework [2], we first develop a simple model of ideological thinking using a well-established connectionist approach called the tensor product model [17–19], and place this model of the individual mind within a social network of interacting agents later. We focus on a focal agent i, which encounters an opinion (**input**) expressed by external sources at time t (figure 1). This is represented by a column vector with n elements, which correspond to a set of propositions that constitute an ideology [11]. The activation levels of the n elements reflect the degrees of endorsement of the corresponding propositions expressed in the encountered opinion, and are given by real numbers from +1 (endorsement) to −1 (dis-endorsement). An opinion then can be understood as a point in the n-dimensional opinion space thus defined, and opinion dynamics as movements in the opinion space.

In this model, an agent consists of an interpreter and a memory mechanism. A focal agent i interprets the encountered opinion through the interpreter (C_i), which acts as their cognitive lens. Mathematically, C_i is an $n \times n$ matrix that filters the input and generates its interpretation in the 'opinion' layer in figure 1 via $e_{ti} = C_i$(**input**). Psychologically, C_i indicates agent i's existing associations between the propositions—how much the agent believes that one proposition is related to another. e_{ti} is a column vector representing how agent i has interpreted the encountered opinion's levels of agreement with the propositions.

Each agent also has a pre-existing memory, $M_{(t-1)i}$, which represents the associations between the interpreted opinions ('opinion' in figure 1) and the represented sources ('source' in figure 1). Agent i's interpreted opinion is entered into the memory mechanism, where it generates $s_{ti} = M_{(t-1)i} e_{ti}$. For the reasons we explain later, we regard s_{ti} as an input *source*. The input source and the interpreted input are associated by Hebbian learning to produce an episodic memory at time t,

$\boldsymbol{E}_{ti} = \boldsymbol{s}_{ti}\boldsymbol{e}_{ti}^{\mathrm{T}}$. Hebbian learning is a mechanism often used in the modelling of associative memory in which two simultaneously activated cognitive representations increase their memory association proportionally to the product of the levels of activation (e.g. [18]; see electronic supplementary material). The updated memory is computed as a weighted average of the episodic memory and the pre-existing memory, $\boldsymbol{M}_{(t-1)i}$, as below:

$$\boldsymbol{M}_{ti} = d_i\boldsymbol{M}_{(t-1)i} + (1 - d_i)\boldsymbol{E}_{ti}, \tag{3.1}$$

where $0 < d_i < 1$ controls the rate of learning for agent i, with a larger value implying a slower learning. For all simulations, we kept d_i constant at 0.5. As a new input enters into the system, it is interpreted, its input source is encoded, episodic memories are constructed and memory is updated iteratively.

We postulate that ideological as opposed to non-ideological processing of information can occur separately for interpretation and memory. First, the interpreter can be *ideological* or *unbiased*. Akin to Zmigrod's *doctrinal* [22] ideological thinking, ideological interpreters generate ideologically biased interpretations of inputs and produce doctrinaire perceptions of social reality, whereas unbiased interpreters process information accurately (also see [23]). Second, the memory mechanism may be *ideologically ego-involved* or *uninvolved*. Well aligned with Zmigrod's *relational* or *identity-based* ideological thinking, ideological ego-involvement promotes motivated commitment to an ideological position, while an ideologically uninvolved mechanism does not show ideological commitment (also see [24]). Factorially combining ideological versus unbiased interpreter and ideologically ego-involved versus uninvolved memory mechanisms, we create a 2×2 typology of ideology-relevant information processing styles and study their implications for opinion dynamics.

(a) Interpretation-based ideological mechanism

To model ideological interpretation, let a prototypical ideology be represented as a point $\boldsymbol{a}$ in the n-dimensional opinion space. We assume that, prior to receiving inputs (i.e. $t = 0$), agent i acquires this ideology with some slight variation from its prototype, so that their acquired ideology, $\boldsymbol{e}_{0i}$, approximates the ideology, $\boldsymbol{a}$: $\boldsymbol{e}_{0i} \approx \boldsymbol{a}$. Here, the cognitive mechanisms normalize the vectors (i.e. $|\boldsymbol{e}_{0i}|$, $|\mathbf{input}| = 1$), and the interpreter is modelled as an autoassociator: $\boldsymbol{C}_i = \boldsymbol{e}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}} \approx \boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$, which acts as an ideologically biased filter. When an input enters, the following obtains: $\boldsymbol{C}_i(\mathbf{input}) = (\boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input})\boldsymbol{e}_{0i}$, where $\boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input} = \alpha$ is a scalar. In other words, the interpreted input is a replica of the learned ideology with its length scaled to its cosine similarity with the prototypical ideology. The more similar the input is to the ideology, the more amplified it becomes; however, if the input is contrary to the ideology ($\alpha < 0$), it is encoded as its antithesis. Thus, for an ideological agent, the cognitive associations between the ideological propositions, $\boldsymbol{C}_i$, provide cognitive support for the acquired ideology.

In contrast, an *unbiased* agent has an interpreter $\boldsymbol{C}_i$ corresponding to the identity matrix $\boldsymbol{I}$ whose diagonal elements are 1 and off-diagonals are 0, so that $\boldsymbol{C}_i(\mathbf{input}) = \mathbf{input}$. In other words, agent i interprets their input without distortion because they interpret all propositions independently of each other.

(b) Memory-based ideological mechanism

The memory-based ideological mechanism can be modelled through the initial memory representation, $\boldsymbol{M}_{0i}$, which exists prior to receiving inputs (i.e. $t = 0$). For the *ideologically ego-involved*, the individual's identity is inexorably associated with the ideology [25]. Let agent i's self-representation, i.e. identity, $\boldsymbol{s}_{0i}$ ($|\boldsymbol{s}_{0i}| = 1$), be a pattern of activations over the 'source' units in figure 1. The units represent aspects of agent i's identity, including personal attributes, group memberships [26], or any features that agent i defines themself with. Following Johnson & Eagly [27], we model ideological ego-involvement as a memory association between their self-representation and the ideology as learned by agent i, $\boldsymbol{e}_{0i} \approx \boldsymbol{a} : \boldsymbol{M}_{0i} = \boldsymbol{s}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$.

This mechanism has a profound implication for how the source of the first input, $\boldsymbol{s}_{1i}$, is encoded. If we let $\alpha = \boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input}$, $\boldsymbol{s}_{1i} = \alpha\boldsymbol{s}_{0i}$. Psychologically, the first input source is encoded in terms of the aspects of the agent's identity. If the input resonates with the learned ideology, $\boldsymbol{e}_{0i}$, ($\alpha > 0$), the first input source, $\boldsymbol{s}_{1i} = \alpha\boldsymbol{s}_{0i}$, is regarded as an ideological ally represented similarly to themself. If the first input opposes the learned ideology, $\alpha < 0$, the source becomes their nemesis. To see this, note that for those with an ideological interpreter, $\boldsymbol{s}_{1i} = \boldsymbol{M}_{0i}\boldsymbol{e}_{1i} = \boldsymbol{s}_{0i}(\boldsymbol{e}_{0i}^{\mathrm{T}}\boldsymbol{e}_{1i}) = \boldsymbol{e}_{0i}^{\mathrm{T}}[\boldsymbol{C}_i(\mathbf{input})]\boldsymbol{s}_{0i} = (\boldsymbol{e}_{0i}^{\mathrm{T}}\boldsymbol{e}_{0i})(\boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input})\boldsymbol{s}_{0i} = (\boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input})\boldsymbol{s}_{0i} = \alpha\boldsymbol{s}_{0i}$; for those with unbiased interpreter, $\boldsymbol{s}_{1i} = \boldsymbol{M}_{0i}\boldsymbol{e}_{1i} = \boldsymbol{s}_{0i}(\boldsymbol{e}_{0i}^{\mathrm{T}}\boldsymbol{e}_{1i}) = \boldsymbol{e}_{0i}^{\mathrm{T}}[\boldsymbol{C}_i(\mathbf{input})]\boldsymbol{s}_{0i} = (\boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input})\boldsymbol{s}_{0i} = \alpha\boldsymbol{s}_{0i}$ because $\boldsymbol{C}_i = \mathbf{I}$.

In *ideologically uninvolved* minds, their identity is associated with ideas unrelated to the ideology, $\boldsymbol{e}_{ui}$, such that $\mathrm{EXP}(\boldsymbol{e}_{ui}\boldsymbol{a}^{\mathrm{T}}) = 0$, where EXP stands for expected value. That is, the individual's initial ideas are, on average, orthogonal to the ideology. Here, agent i's initial memory is an association between their identity, $\mathbf{s_{0i}}$, and the unrelated ideas, $\boldsymbol{e}_{ui}$: $\boldsymbol{M}_{0i} = \boldsymbol{s}_{0i}\boldsymbol{e}_{ui}^{\mathbf{T}}$. This means that the initial ideas, $\boldsymbol{e}_{ui}$, of some agents may resonate with the ideology ($\boldsymbol{e}_{ui}\boldsymbol{a}^{\mathrm{T}} \approx \boldsymbol{e}_{ui}\boldsymbol{e}_{0i}^{\mathrm{T}} = \beta > 0$), and the initial ideas of others might oppose it ($\beta < 0$). On the average, $\mathrm{EXP}(\beta) \approx 0$. If the agent has an ideological interpreter, the first input source is encoded as $\boldsymbol{s}_{1i} = \alpha\beta\boldsymbol{s}_{0i}$ ($\boldsymbol{s}_{1i} = \boldsymbol{M}_{0i}\boldsymbol{e}_{1i} = \boldsymbol{s}_{0i}(\boldsymbol{e}_{ui}^{\mathrm{T}}\boldsymbol{e}_{1i}) = \boldsymbol{e}_{ui}^{\mathrm{T}}[\boldsymbol{C}_i(\mathbf{input})]\boldsymbol{s}_{0i} = [(\boldsymbol{e}_{ui}^{\mathrm{T}}\mathbf{e_{0i}})(\boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input})\boldsymbol{s}_{0i}]$). If the agent has unbiased interpreter, $\mathbf{s}_{1i} = \beta\boldsymbol{s}_{0i}$ ($\boldsymbol{M}_{0i}\boldsymbol{e}_{1i} = \boldsymbol{s}_{0i}(\boldsymbol{e}_{ui}^{\mathrm{T}}\boldsymbol{e}_{1i}) = \boldsymbol{e}_{ui}^{\mathrm{T}}[\boldsymbol{C}_i(\mathbf{input})]\boldsymbol{s}_{0i} = (\boldsymbol{e}_{ui}^{\mathrm{T}}\mathbf{input})\boldsymbol{s}_{0i}$).

(c) Output process

So far, we have described the process by which an agent forms an opinion. How does the agent express it? This process is modelled as accessing memory using the agent's self-representation, $\boldsymbol{s}_{0i}$, as a cue: $\boldsymbol{o}_{ti} = \boldsymbol{s}_{0i}^{\mathrm{T}}\boldsymbol{M}_{ti}$. What is retrieved ($\boldsymbol{o}_{ti}$) is then the self's opinion at the time of access. Suppose that the focal person expresses an opinion based on their initial memory, $\boldsymbol{M}_{0i}$, before receiving any inputs. An ideologically ego-involved person retrieves the learned ideology: $\boldsymbol{o}_{0i} = (\boldsymbol{s}_{0i}^{\mathrm{T}}\boldsymbol{s}_{0i})\boldsymbol{e}_{0i}^{\mathrm{T}} = \boldsymbol{e}_{0i}^{\mathrm{T}}$. However, an ideologically uninvolved person retrieves the unrelated initial ideas: $\boldsymbol{o}_{0i} = \boldsymbol{e}_{ui}$. Depending on whether their initial ideas resonate with the learned ideology ($\beta > 0$) or oppose it ($\beta < 0$), they will express opinions accordingly. The retrieved opinion is then put through the interpreter in reverse to generate an output (**output**) based on the notion that interpretation and production involve the same underlying representations [28].

(d) Typology of ideological thinking

Combining the interpretation-based and memory-based ideological mechanisms, we have four different types of ideological thinking (table 1).

Table 1. Typology of ideological thinking. Note: $\boldsymbol{a}$ = prototype of ideology; $\boldsymbol{e}_{0i}$ = learned ideology ($\approx\boldsymbol{a}$); $\boldsymbol{s}_{0i}$ = self; $\boldsymbol{e}_{ui}$ = unrelated initial beliefs; $\boldsymbol{I}$ = identity matrix; $\alpha = \boldsymbol{e}_{0i}^{\mathrm{T}}\mathbf{input}$; $\beta = \boldsymbol{e}_{ui}^{\mathrm{T}}\boldsymbol{e}_{0i}$. Note that $-1 < \alpha, \beta < 1$, but $0 < \alpha^2 < 1$.

type		interpreter (ζ_i)		$\mathbf{e}_{1i}$	memory ($\mathbf{M}_{0i}$)		$\mathbf{s}_{1i}$	$\mathbf{E}_{1i}$
1	fully ideological	ideological	$\boldsymbol{e}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$	$\alpha\boldsymbol{e}_{0i}$	ego-involved	$\boldsymbol{s}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$	$\alpha\boldsymbol{s}_{0i}$	$\alpha^2\boldsymbol{s}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$
2	ideological interpreter	ideological	$\boldsymbol{e}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$	$\alpha\boldsymbol{e}_{0i}$	uninvolved	$\boldsymbol{s}_{0i}\boldsymbol{e}_{ui}^{\mathrm{T}}$	$\alpha\beta\boldsymbol{s}_{0i}$	$\alpha^2\beta\boldsymbol{s}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$
3	ego-involved ideologist	unbiased	$\boldsymbol{I}$	$\mathbf{input}$	ego-involved	$\boldsymbol{s}_{0i}\boldsymbol{e}_{0i}^{\mathrm{T}}$	$\alpha\boldsymbol{s}_{0i}$	$\alpha\boldsymbol{s}_{0i}\mathbf{input}^{\mathrm{T}}$
4	non-ideological	unbiased	$\boldsymbol{I}$	$\mathbf{input}$	uninvolved	$\boldsymbol{s}_{0i}\boldsymbol{e}_{ui}^{\mathrm{T}}$	$\beta\boldsymbol{s}_{0i}$	$\beta\boldsymbol{s}_{0i}\mathbf{input}^{\mathrm{T}}$

Type 1 is *fully ideological*—interpreting inputs through an ideological lens and ideologically ego-involved.
Type 2 is an *ideological interpreter*—ideologically interpreting inputs, but ideologically uninvolved.
Type 3 is an *ego-involved ideologist*, unbiased in interpretation, but committed to an ideology.
Type 4 is *non-ideological*—unbiased in interpretation and ideologically uninvolved.

Arguably, type 1 is most and type 4 is least ideologically minded, with the other types in between.

To illustrate the opinion updating processes, we simulated 100 agents of each type, in which each agent learned 50 inputs with each element randomly generated from uniform distribution [−1, 1]. Here, the agents do not interact with each other, but simply process incoming random inputs. Figure 2 shows the results, with the *y*-axis representing **(output)a**, the extent to which each agent's output is aligned with the ideology, $\boldsymbol{a}$, as defined in §3a.

Type 1 takes any input as a form of confirmation of the learned ideology. This is because an input that opposes their ideology is attributed to an opponent, and the input is treated as confirming their ideology (table 1; electronic supplementary material). Consequently, outputs never change. The fully ideological mind, therefore, remains stable and highly aligned with the ideology no matter what. Thus, a population of fully ideological agents will consensualize if there is only one ideology; however, it will always polarize if there are competing ideologies.

Type 2 remembers an input in terms of its alignment with their initial stance, $\boldsymbol{e}_{ui}$. Inputs are regarded as supporting the ideology if their initial opinions are aligned with the ideology, but as opposing the ideology if their initial opinions are opposing it. As a result, these ideological interpreters change their opinions depending on their starting point. In the long run, those who have initially agreed with an ideology even slightly will eventually become staunch supporters, whereas those who have initially opposed it will become a strong opposition. In contrast to type 1, a population of type 2 agents will *always* polarize even if there is only one ideology (provided that there is enough variability in initial opinions). These processes are akin to 'elective affinities' or 'ideo-affective resonance', in which an ideology 'finds' the minds that resonate with it [29]. This type does not have an ego-involved commitment to an ideology *per se*, but implies doctrinal information processing [22,30]. Growing evidence that those who hold polarized opinions tend to lack reflective reasoning (e.g. [15,30,31]) implies this type of ideological thinking.

Type 3 initially retains a learned ideology, but changes opinion as inputs come in depending on whether they confirm or contradict the ideology. This occurs because, although the initial memory is strongly aligned with the ideology, new episodic memories are largely unbiased and accumulate accordingly. If inputs continue to contradict the learned ideology, type 3 agents eventually change their opinions in line with the preponderance of the inputs. In this sense, the ideologically ego-involved may not show cognitive inflexibility, but exhibit ego-involved commitment. Whether a population of type 3 agents polarizes, therefore, depends on the inputs they receive.

Type 4 changes opinions in accordance with inputs, as one would expect given unbiased and uninvolved processes. Does a population of type 4 agents polarize? As we will show below, it depends on communication and social network dynamics.

4. Modelling communication and social influence on dynamic social networks

(a) Communication and social influence in multidimensional opinion space

Our modelling has so far explored the role of psychological factors on polarization, but does not capture the dynamics arising from communication. In the previous simulation, all agents simply received random inputs, rather than inputs from each other. Let us now place our isolated cognitive agents within a framework that incorporates communication. We adopt Parsegov, Friedkin, and their colleagues' model of multidimensional social influence on social networks [20,21] (see electronic supplementary material for other references). In their model, all agents communicate their opinions, and are all influenced by others' opinions simultaneously. Supposing that there are N agents ($1 \le i, j \le N$), agent j's influence on agent i is captured by a coefficient $0 < w_{ji} < 1$. However, not all agents are influenced equally; agent i is assumed to have a degree of stubbornness, represented by a coefficient, w_{ii}, an influence on itself. Each agent's influence on each other (including themself) is captured by the full influence matrix, $\boldsymbol{W}$, whose element is w_{ji}. If an agent's opinion contains n propositions, the influence of that agent applies to every proposition equally, and the sum of all influences $\sum w_{ij}$ over all other agents j is assumed to be 1 for every agent i.

Let $\boldsymbol{X}$ be an $N \times n$ matrix of the population opinions: each agent i's initial endorsement of each of n propositions is in the ith row. $\boldsymbol{X}(0)$ indicates the set of initial opinions, and after each round of discussion, each agent receives the aggregated input as weighted by the influence matrix, $\boldsymbol{W}$, as modelled by (4.1):

$$\boldsymbol{X}(k+1) = \boldsymbol{A}\boldsymbol{W}\boldsymbol{X}(k)\boldsymbol{C}^{\mathrm{T}} + (\boldsymbol{I} - \boldsymbol{A})\boldsymbol{X}(0), \quad k = 0, 1, \ldots, \tag{4.1}$$

where $\boldsymbol{A}$ is a diagonal $N \times N$ matrix with $1 - w_{ii}$ in its element at ith row and ith column and 0 elsewhere; $\boldsymbol{W}$ is an $N \times N$ matrix

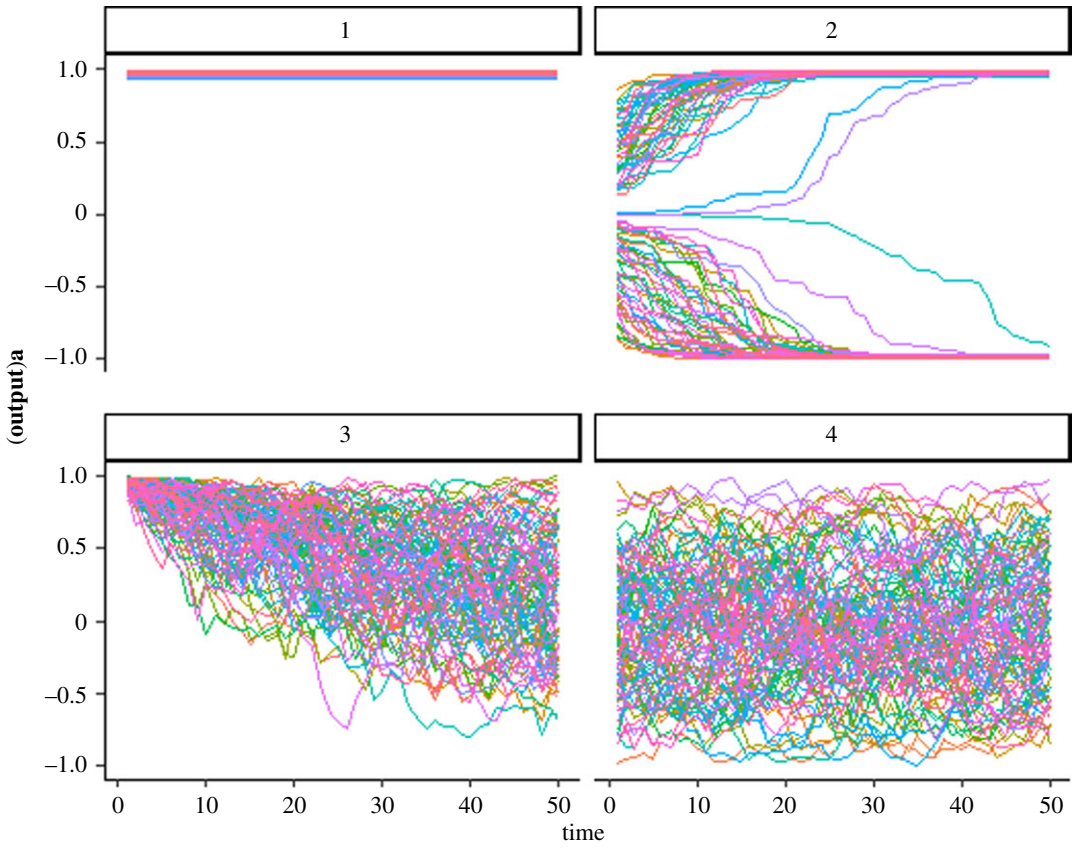

Figure 2. Trajectories of each of the four cognitive styles. Note. The *y*-axis indicates **(output)a**, the dot product of the output with an ideology. Each graph shows the results of 50 random inputs for 100 agents of each type. 1, ideological interpreter + ideologically ego-involved; 2, ideological interpreter; 3, ideologically ego-involved; 4, non-ideological. (Online version in colour.)

with w_{ij}; and $\boldsymbol{I}$ is an identity matrix. Note that an $n \times n$ matrix $\boldsymbol{C}$ has an element c_{ml}, which captures the mth proposition's implication on the lth proposition ($1 \leq m$, $1 \leq n$). In this model, the implication of one proposition for another proposition, $\boldsymbol{C}$, is the same for all agents. Put differently, all agents are assumed to have the same understanding about the conceptual interdependencies between all propositions.

Overall, then, this model represents the input to each agent after a round of communication as a weighted average of that agent's originally expressed opinion and the opinions of all other agents. Rounds of communication continue k times, and the input received by each agent is updated in combination with the other agents' opinions until eventually reaching a stable input for each agent.

In this paper, we elaborate on this model in two main ways. First, we interpret the process modelled here as *grounding*, a process by which communicators establish a mutual understanding about their utterances [2]. Each agent communicates the output of its cognitive system to all other agents, and the communication process aggregates these outputs together to produce new input to each agent's cognitive system. Second, instead of assuming that all agents have the same conceptual interdependencies between propositions, we allow each agent to have their own understanding. This individual variation derives from the cognitive processes captured earlier. Recall that we modelled an ideological interpreter who regards each proposition as having a *psychological* implication ($c_{ml} \neq 0$) for another proposition, and an unbiased interpreter as one who evaluates each proposition independently of other propositions ($c_{ml} = 0$, for all $m \neq l$). Our agents also vary, as before, in whether they are ideologically ego-involved or uninvolved. For this reason, $\boldsymbol{C}$ is not included in our application of equation (4.1) because it is absorbed by our cognitive model.

Figure 3 shows the results of simulations for populations of agents with different cognitive styles, given the communication model. Here, we constructed $\boldsymbol{W}$ based on the assumption that social influence depends on status characteristics (electronic supplementary material). Among all agents, communication removes variability and opinion diversity, generally by exacerbating existing patterns. Full-fledged ideologues (type 1) converge to their interpreted ideology, and ideological interpreters (type 2) polarize even more rapidly. Ego-involved ideologists (type 3) converge to their ideology because they are exchanging ideologically biased communications, basically forming an 'echo chamber' (e.g. [32]). In contrast, unbiased type 4 agents converge on a middle ground.

(b) Communication and influence on dynamic social networks

One simplification made by the communication model so far is that all agents communicate with all other agents within a stable social network. However, the contemporary socio-technological context is characterized by the ability to modify network ties easily. This is because of the social media technology, which makes it easier than ever before to rapidly form or sever social connections at the click of a mouse. Social psychologists call this *relational mobility* [33,34].

The final part of our paper explores the impact of relational mobility on public opinion dynamics. We capture this in the model by permitting agents to modify their influence weights. Specifically, we let each agent modify its tie

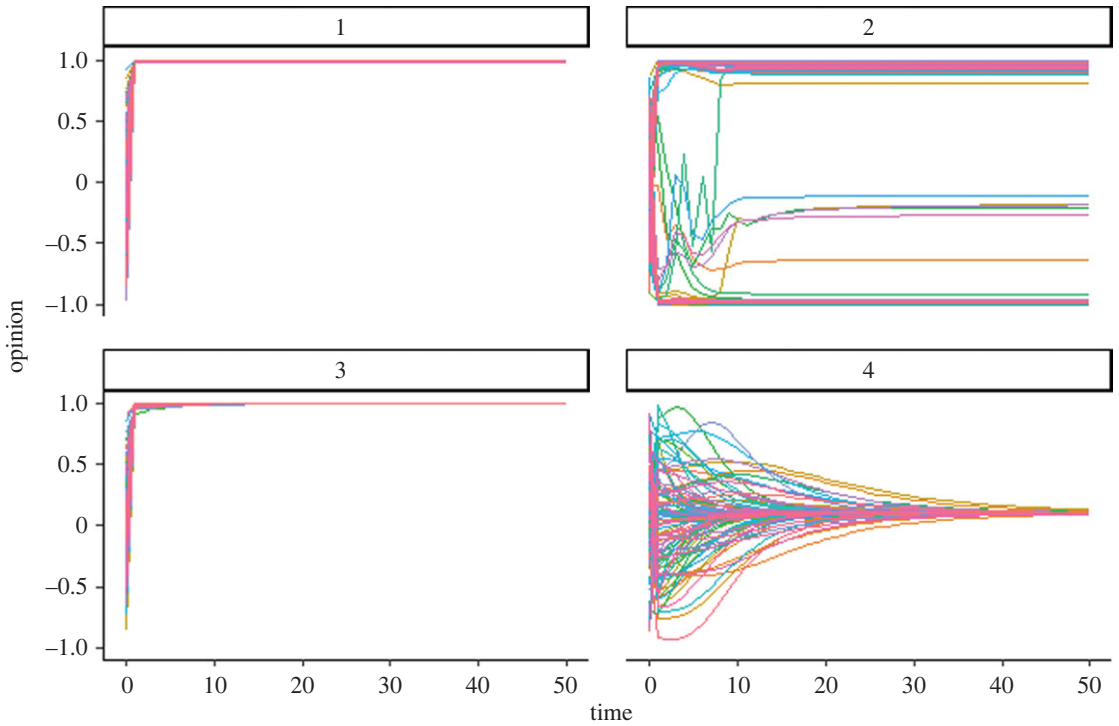

Figure 3. Trajectories of opinion dynamics with communication. Note. The *y*-axis indicates the extent to which each agent's output is aligned with the ideology. In this plot, time 0 is that of the initial random inputs to the system, and times 1–50 plot the opinions expressed by agents: 1, ideological interpreter + ideologically ego-involved; 2, ideological interpreter; 3, ideologically ego-involved; 4, non-ideological.

strength as a function of the degree of agreement between agents. If agent *j*'s opinion resonates with agent *i*'s opinion (i.e. the cosine similarity between two opinion vectors is positive), the tie strength is adjusted upwards as in (4.2). If they are antagonistic to each other (i.e. the cosine similarity is zero or negative), the tie is severed. This reflects homophily—the tendency to affiliate with others who hold similar opinions (e.g. [35]).

$$w_{ij} = \begin{cases} w_{ij}(1 - w_{ii}) / \sum_{j \in S+} w_{ij} & \text{if } s(\boldsymbol{x}(i),\boldsymbol{x}(j)) > 0 \\ w_{ii}, & \text{and } \sum_{j \in S+} w_{ij} \neq 0 \text{ for } j \neq i; \text{ for } j = i; \\ 0, & \text{otherwise.} \end{cases} \tag{4.2}$$

Note that the stubbornness of agent *i*, w_{ii}, is an individual difference and thus remains constant over the course of each simulation. $s(\boldsymbol{x}(i),\boldsymbol{x}(j))$ is a cosine similarity between opinion vectors, $\boldsymbol{x}(i)$ and $\boldsymbol{x}(j)$, for agents *i* and *j*. S+ is the set of agents $j \neq i$ such that $s(\boldsymbol{x}(i),\boldsymbol{x}(j)) > 0$. Note also that equation (4.2) holds unless $\sum_{j \in S+} w_{ij} = 0$. Otherwise, w_{ii} is the only non-zero weight and agent *i* becomes an isolate. Equation (4.2) redistributes the weights for the severed ties to the retained ties proportional to the latter's original weights.

Figure 4 shows the results of simulations with 100 relationally mobile agents and 50 random starting points for each of the four cognitive styles. Relational mobility clearly exacerbates the effects of communication even further. Populations with type 1 and type 3 agents again do not move from their initial ideology. Populations of type 2 agents—interpreting information through the lens of their ideology but ideologically uninvolved—polarize into opposite opinions. Type 4 is particularly interesting. Previously this type always produced a consensus in the presence of communication. When relationally mobile, however, they can now produce divergent opinion clusters sometimes. This happens when initial variations in opinions get reified because agents quickly stop talking to other agents with whom they disagree. Our simulations of type 2 and type 4 cognitive styles with and without relational mobility (electronic supplementary material) confirmed these observations. Type 2 polarized more quickly with mobility than without. Type 4 always consensualized without mobility, but occasionally converged to divergent opinion clusters with mobility.

5. Discussion

The ideological mind is a foundation of polarization. Whether a full-fledged ideologue (type 1), ideological interpreter (type 2), or ego-involved ideologist (type 3), if an agent engages in ideological thinking through biased interpretation or ego-involved commitment, the result is likely to be a polarized discourse. However, communication and social network dynamics interact with ideological cognition and motivation to produce different trajectories. Although those who interpret with an ideological lens (type 1 or 2) always hold extreme views, ego-involved ideologists (type 3) polarize particularly when they form an echo chamber. Communication can influence non-ideological agents (type 4): if they exclusively interact with polarized agents, they too end up with extreme views. Relational mobility can drive even non-ideological agents into diverse and siloed opinion clusters. The contemporary socio-technological context of digital communication with high levels of relational mobility is conducive to polarized discourse.

Our modelling points to the possibility that those who hold polarized opinions can do so as a result of any combination of cognitive, motivational, or network dynamic processes. The interpretation- and memory-based mechanisms of ideological thinking provide psychological mechanisms for assimilative and repulsive social influence processes [6], in

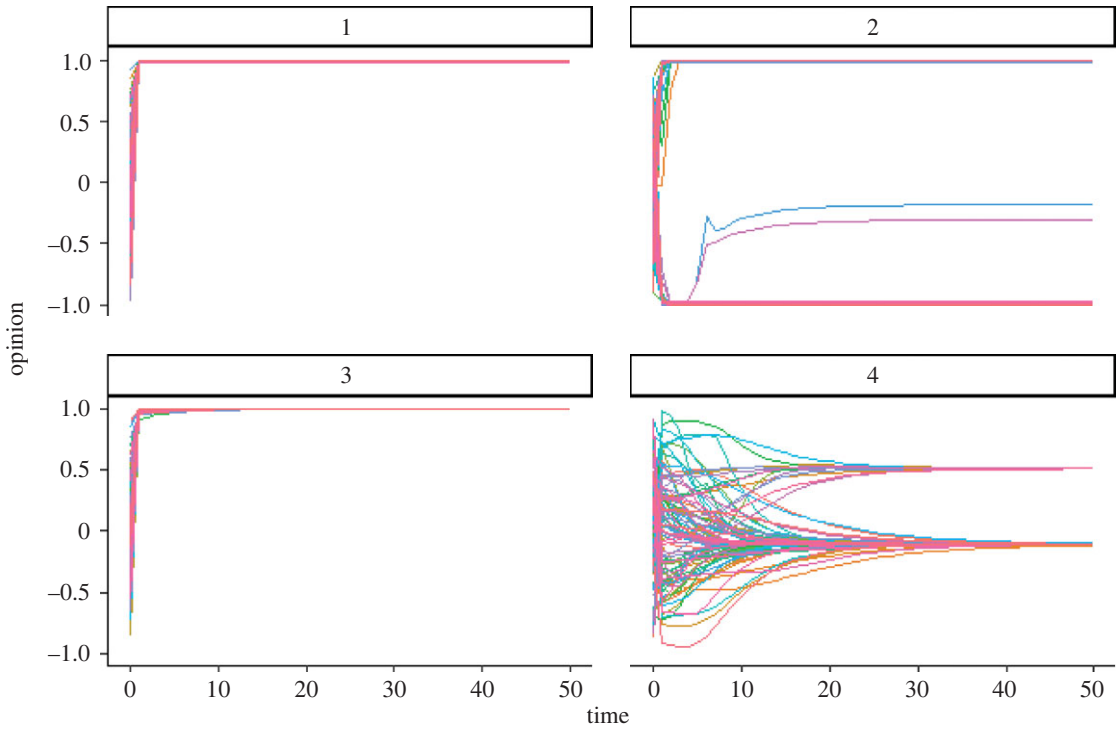

Figure 4. Trajectories of opinion dynamics with relational mobility. Note. The *y*-axis indicates the extent to which each agent's output is aligned with the ideology. Each graph shows the results of an example of 100 interacting agents of each type: 1, ideological interpreter + ideologically ego-involved; 2, ideological interpreter; 3, ideologically ego-involved; 4, non-ideological. (Online version in colour.)

which communication drives opinions to assimilate with or repulse each other. Nonetheless, the network mechanism of homophily generates the social condition in which agents only assimilate similar opinions, thereby increasing the likelihood of opinion clustering [6]. Therefore, even those who are psychologically non-ideological can hold polarized opinions if they are in an ideologically minded closed social circle.

A corollary is that different processes may depolarize different types of ideological agents. Social ties with others who hold diverse opinions could depolarize the non-ideological (type 4) and even the ideologically committed (type 3) in the long run. What about the others who ideologically interpret (types 1 and 2)? One possibility is to introduce new opinion dimensions in the political discourse. By having additional dimensions on which agents can agree, even they may depolarize. Further research is needed to explore this possibility.

Our modelling suggests that the existence of opposing ideologies is not a necessary condition for polarization; even the presence of a single ideology can produce extreme opinions among ideological interpreters (type 2). This points to an intriguing dialectic—if a single ideology structures political discourse, a polarizing opposition can arise in a population. Indeed, neoliberalism may have played such a role in the late 20th and early 21st CE [12,13] (electronic supplementary material). In this historical context, different types of cognitive and motivational styles may have predominated the opposing ends of a political spectrum. Conservatives who support neoliberalism may be type 1, 2 or 3, but those on the left may be largely type 2. Here, there may be a right–left asymmetry in psychological style—the right has been more committed to a single ideology, whereas the left may have had diverse preoccupations (for a review see [16]). Nonetheless, we may now be witnessing the emergence of an alternative opposing ideology. Our modelling suggests that similar psychological styles may be found under these historical circumstances just as Zmigrod and her colleagues recently reported [16].

Computational approaches to cultural dynamics may be able to provide an integrative framework in which to understand not only micro-psychological, but also macro-historical processes at play in the contemporary circumstances of political polarization.

Data accessibility. This article has no additional data.

Authors' contributions. Y.K., A.P. and V.F. wrote the paper; Y.K. and E.P. conducted simulations; and Y.K. and E.P. wrote the electronic supplementary material.

Competing interests. We declare we have no competing interests.

Funding. This study was funded by Defence Science and Technology Group.

References

1. Laland KN, Odling-Smee JF, Feldman MW. 2000 Niche construction, biological evolution, and cultural change. *Behav. Brain Sci.* **23**, 131–146. (doi:10.1017/S0140525X00002417)
2. Kashima Y, Bain P, Perfors A. 2019 The psychology of cultural dynamics. *Annu. Rev. Psychol.* **70**, 499–529. (doi:10.1146/annurev-psych-010418-103112)
3. Abelson RP, Bernstein A. 1963 A computer simulation model of community referendum controversies. *Public Opin. Q.* **27**, 93–122. (doi:10.1086/267152)
4. Abelson RP. 1964 Mathematical models of the distribution of attitudes under controversy. In *Contributions to mathematical psychology* (eds N Freriksen, H Gulliksen), pp. 142–160. New York, NY: Holt, Rinehart & Winston.

5. DeGroot MH. 1974 Reaching a consensus. *J. Am. Stat. Assoc.* **69**, 118–121. (doi:10.1080/01621459.1974.10480137)
6. Flache A, Mäs M, Feliciani T, Chattoe-Brown E, Deffuant G, Huet S, Lorenz J. 2017 Models of social influence. *J. Artif. Soc. Social Simul.* **20**, 31, 2. (doi:10.18564/jasss.3521)
7. Friedkin NE, Johnsen EC. 1990 Social influence and opinions. *J. Math. Sociol.* **15**, 193–205. (doi:10.1080/0022250X.1990.9990069)
8. Hegselmann R, Krause U. 2002 Opinion dynamics and bounded confidence models, analysis, and simulation. *J. Artif. Soc. Social Simul.* **5**, 3. See https://jasss.soc.surrey.ac.uk/5/3/2.html.
9. Baldassarri D, Bearman P. 2007 Dynamics of political polarization. *Am. Sociol. Rev.* **72**, 784–811. (doi:10.1177/000312240707200507)
10. Jager W, Amblard F. 2005 Uniformity, bipolarization and pluriformity captured as generic stylized behavior with an agent-based simulation model of attitude change. *Comput. Math. Org. Theory* **10**, 295–303. (doi:10.1007/s10588-005-6282-2)
11. Converse PE. 1964 The nature of belief systems in mass publics. In *Ideology and discontent* (ed. D Apter), pp. 206–261. New York, NY: Free Press.
12. Bay-Cheng LY, Fitz CC, Alizaga NM, Zucker AN. 2015 Tracking Homo oeconomicus. *J. Soc. Polit. Psychol.* **3**, 71–88. (doi:10.5964/jspp.v3i1.366)
13. Beattie P. 2019 The road to psychopathology. *J. Social Iss.* **75**, 89–112. (doi:10.1111/josi.12304)
14. Kahan DM. 2013 Ideology, motivated reasoning, and cognitive reflection. *Judgm. Decis. Mak.* **8**, 407–424. (doi:10.2139/ssrn.2182588)
15. Pennycook G, Rand DG. 2019 Lazy, not biased. *Cognition* **188**, 39–50. (doi:10.1016/j.cognition.2018.06.011)
16. Zmigrod L, Rentfrow PJ, Robbins TW. 2020 The partisan mind: is extreme political partisanship related to cognitive inflexibility? *J. Exp. Psychol. Gen.* **149**, 407–418. (doi:10.1037/xge0000661)
17. Humphreys MS, Bain JD, Pike R. 1989 Different ways to cue a coherent memory system. *Psychol. Rev.* **96**, 208–233. (doi:10.1037/0033-295X.96.2.208)
18. Kashima Y, Woolcock J, Kashima ES. 2000 Group impressions as dynamic configurations. *Psychol. Rev.* **107**, 914–942. (doi:10.1037/0033-295X.107.4.914)
19. Osth AF, Dennis S. 2015 Sources of interference in item and associative recognition memory. *Psychol. Rev.* **122**, 260–311. (doi:10.1037/a0038692)
20. Friedkin NE, Proskurnikov AV, Tempo R, Parsegov SE. 2016 Network science on belief system dynamics under logic constraints. *Science* **354**, 321–326. (doi:10.1126/science.aag2624)
21. Parsegov SE, Proskurnikov AV, Tempo R, Friedkin NE. 2017 Novel multidimensional models of opinion dynamics in social networks. *IEEE Trans. Autom. Control* **62**, 2270–2285. (doi:10.1109/TAC.2016.2613905)
22. Zmigrod L. 2020 A psychology of ideology. *PsyArXiv*. (doi:10.31234/osf.io/ewy9t).
23. Zmigrod L, Eisenberg IW, Bissett PG, Robbins TW, Poldrack RA. 2021 The cognitive and perceptual correlates of ideological attitudes: a data-driven approach. *Phil. Trans. R. Soc. B* **376**, 20200424. (doi:10.1098/rstb.2020.0424)
24. Bélanger JJ. 2021 The sociocognitive processes of ideological obsession: review and policy implications. *Phil. Trans. R. Soc. B* **376**, 20200144. (doi:10.1098/rstb.2020.0144)
25. Sherif M, Hovland CI. 1961 *Social judgment*. New Haven, CT: Yale University Press.
26. Lau T. 2021 Reframing social categorization as latent structure learning for understanding political behaviour. *Phil. Trans. R. Soc. B* **376**, 20200136. (doi:10.1098/rstb.2020.0136)
27. Johnson BT, Eagly AH. 1989 Effects of involvement on persuasion. *Psychol. Bull.* **106**, 290–314. (doi:10.1037/0033-2909.106.2.290)
28. Pickering MJ, Garrod S. 2013 An integrated theory of language production and comprehension. *Behav. Brain Sci.* **36**, 329–347. (doi:10.1017/S0140525X12001495)
29. Jost JT, Federico CM, Napier JL. 2009 Political ideology. *Annu. Rev. Psychol.* **60**, 307–337. (doi:10.1146/annurev.psych.60.110707.163600)
30. Zmigrod L. 2021 The cognitive and perceptual correlates of ideological attitudes: a data-driven approach. *Phil. Trans. R. Soc. B* **376**, 20200424. (doi:10.1098/rstb.2020.0424)
31. Gampa A, Wojcik SP, Motyl M, Nosek BA, Ditto PH. 2019 (Ideo)Logical reasoning: ideology impairs sound reasoning. *Soc. Psychol. Personal. Sci.* **10**, 1075–1083. (doi:10.1177/1948550619829059)
32. Colleoni E, Rozza A, Arvidsson A. 2014 Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *J. Commun.* **64**, 317–332. (doi:10.1111/jcom.12084)
33. Thomson R *et al.* 2018 Relational mobility predicts social behaviors in 39 countries and is tied to historical farming and threat. *Proc. Natl Acad. Sci. USA* **115**, 7521–7526. (doi:10.1073/pnas.1713191115)
34. Oishi S. 2010 The psychology of residential mobility. *Perspect. Psychol. Sci.* **5**, 5–21. (doi:10.1177/1745691609356781)
35. McPherson M, Smith-Lovin L, Cook JM. 2001 Birds of a feather. *Annu. Rev. Sociol.* **27**, 415–444. (doi:10.1146/annurev.soc.27.1.415)