Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

Ideology and predictive processing: coordination, bias, and polarization in socially constrained error minimization

Nathan E Wheeler[1], Suraiya Allidina[1], Elizabeth U Long[1], Stephen P Schneider[2], Ingrid J Haas[3] and William A Cunningham[1]

Recent models of cognition suggest that the brain may implement predictive processing, in which top-down expectations constrain incoming sensory data. In this perspective, expectations are updated (error minimization) only if sensory data sufficiently deviate from these expectations (prediction error). Although originally applied to perception, predictive processing is thought to generally characterize cognitive architecture, including the social cognitive processes involved in ideological thinking. Scaling up these simple computational principles to the social sphere outlines a path by which group members may adopt shared ideologies and beliefs to predict behavior and cooperate with each other. Because ideological judgments are of specific interest to others in our political groups, we may increasingly regulate each other's thinking, sharing the process of error minimization. In this paper, we outline how this process of shared error minimization may lead to shared ideologies and beliefs that allow group members to predict and cooperate with each other, and how, as a consequence, political polarization and extremism may result.

Addresses
[1] Department of Psychology, University of Toronto, 100 St. George Street, Toronto, ON, M5S3G3, Canada

[2] Carl R. Woese Institute for Genomic Biology, University of Illinois at Urbana-Champaign, 1206 W. Gregory Dr., Urbana, IL, 61801, USA
[3] Department of Political Science, University of Nebraska, 531 Oldfather Hall, Lincoln, NE, 68588, USA

Corresponding author:
Cunningham, William A (cunningham@psych.utoronto.ca)

Current Opinion in Behavioral Sciences 2020, **34**:192–198

This review comes from a themed issue on **Emotion, motivation, personality and social sciences- *Political ideologies***

Edited by **John Jost**, **Eran Halperin** and **Kristin Laurin**

https://doi.org/10.1016/j.cobeha.2020.05.002

Introduction

For some time, cognitive scientists have suggested that we do not see the world as it is, but rather, interpret reality in a way that meets our needs (e.g. [1]). Because representing the complex perceptual landscape around us requires metabolically expensive cognitive architecture, perceptual representation must find an adaptive balance between accuracy and efficacy. Along these lines, predictive processing provides a computational framework for a neurally-plausible hierarchical architecture that achieves this balance. From this approach, increasingly superordinate layers of a neural network constrain perceptual processing to be consistent with increasingly abstract models or beliefs about the causal structure of the world [2••,3,4]. At the same time, error resulting from these constraints directs action and learning to converge on more accurate and pragmatic models. While originally applied to perception, predictive processing is meant to extend up the cognitive hierarchy, including to social cognitive processes [5]. However, operating within the social world produces a unique paradox, as cooperating with strangers requires us to predict their likely actions, in many cases without prior knowledge of what they are like. To this end, social living requires that strangers adopt common conventions and convictions that can be used to organize their interactions, allowing them to both predict and be predicted by others. The emergence of such abstract beliefs and conventions may be modeled as mutually constraining error minimization processes that produce shared schemas and goals (e.g. stereotypes, social norms), which constrain interpretations of ambiguous social and political events in a manner that supports cooperation (e.g. [6]). At the highest level, ideologies or worldviews may structure and direct social life via this process by creating rules by which people should live. Thus, while perceptual models anticipate the structure of the physical world, ideologies may both prescribe and be prescribed by cultural schemas that support cooperation for outcomes of mutual interest [7].

In this paper, we seek to integrate levels of analysis to understand the role of ideological processes within a contemporary computational framework. As such, our goal is not to replace related theories in social psychology that inform ideological reasoning (and bias) [8–10,11••,12,13]; rather, we attempt to situate principles of these theories within a computational framework. Specifically, we focus on articulating how the same data-reducing cognitive architecture that directs low-level perception and motor actions may be scaled up to describe ideological reasoning and bias. Along these lines, we outline how extant work in the field may be integrated into an understanding of ideology and ideological bias as a product of shared minimization of prediction errors. To this end, we first review how error minimization constrains learning of causal models of the physical world that structure and direct perception and action. Second, we examine how this error minimization may extend to the sociocognitive domain to develop and update ideologies, or higher-order models of the sociopolitical world, which structure and direct social life. In particular, we discuss how, by these computations alone, social error minimization may lead to shared ideologies and beliefs and allow group members to predict and cooperate with each other. Further, we explore how perceptual biases that are a byproduct of these cognitive heuristics may similarly extend to the sociocognitive domain, leading to ideological bias, polarization, and extremism.

Basics of predictive processing and active inference

Before bridging levels of analysis to apply principles of the predictive processing framework to understanding ideological thought, we first review the basic mechanisms of this framework. Central to the predictive processing framework is the idea that perception is the integration of top-down expectations and bottom-up sensory data representing violations of those expectations [3,14]. Specifically, this framework suggests that perception and action rely on hierarchically organized neural systems with reciprocal connections between each layer of the network. Higher layers of the network provide increasingly general models that predict incoming sensory data in lower layers. Lower layers in turn send feedback to higher levels about how well the predictions match the input, only propagating up sensory information that fails to be predicted by top-down expectations [15]. From this perspective, the brain is constantly predicting the incoming flow of sensory data, forming expectations of what the perceiver should observe in the world. These predictions form the crux of what we see, unless the incoming sensory information provides sufficiently reliable error signals indicating that the predictions are inaccurate. If the sensory signals provide enough reliable prediction errors, these error signals will propagate up the hierarchy to affect what we perceive. Critically, this will only happen if alternate theories that explain the surprising sensory data can be formed, even if the perceiver did not originally expect these theories to be true [2••]. *Stated simply, unless there is sufficient evidence that what we expected to be true is not actually true, we experience the world as in line with the predictions of our internal model*. The relative weighting of these prediction errors is governed by their precision, or uncertainty [5]. Sensory prediction errors with greater precision (less uncertainty) are given higher weight, and are thus more likely to influence the resulting percept. Signals with lower precision are instead down weighted, and in these instances, expectations are more likely to drive perception. Importantly, precision is not simply an objective measure of the reliability of sensory data, but may be affected by individual differences, contexts, or motivational factors such as one's confidence in the attainability of relevant goals [16–21].

Within the predictive processing framework, through these bidirectional connections between neural layers, cognitive systems are driven to minimize free energy, conceptualized as the global mismatch between internal representations of the world and the actual sensory data received from the world, such that we adopt representations that most reliably align with sensory data we receive [2••,22]. Thus, low free energy within the system suggests that the top-down expectations provide a good match to the bottom-up sensory information, whereas high free energy likely indicates that expectations are failing to accurately predict the sensory data. Critically, however, the system aims to minimize prediction errors over the long-term and not just in any one situation: while drastically changing one's world model to account for surprising information might decrease prediction errors within the current context, this would result in a sharp increase in prediction errors in other contexts, and thus an undesirable increase in overall free energy. Prediction error minimization therefore needs to be understood as a global process that aims to keep the abstract whole of the generative model consistent in spite of environmental fluctuations. For this reason, incoming prediction errors rarely propagate all the way to the top of the hierarchy and are instead accommodated at lower levels of the model. If a model specifies some set of expectations about the world, violation to a small subset of those expectations will not result in a revision of the entire model — they will more often be accommodated by more moderate revision of more intermediate rungs of the hierarchy or be ignored completely (e.g. [23]). For example, deformations in the physical features of an object (e.g. painting an apple gold) will not immediately lead to updating one's representation of the object's identity (e.g. 'This is a gold apple'). This property of generative models becomes especially relevant when dealing with extremely abstract, comprehensive models such as ideologies.

In addition to the relatively passive process of updating perceptual models to reflect the underlying structure of the world, agents also have the option of changing the environment to reflect their perceptual models. Active inference, an extension of the predictive processing framework, accounts for this by suggesting that in addition to perceptual inference, error is further reduced through action, specifically through inferring (and applying) actions that selectively sample the environment to produce information that matches our expectations [24]. This can be as simple as orienting attention toward (or away from) environmental (ir)regularities to selectively approach (or avoid) situations and experiences that are expected to (mis)match one's expectations [25]. Actions that reduce error in this way have the function of structuring the environment to be more predictable under the current model, reducing the demand on perceptual models to accommodate errors that the agent does not need to learn. Along these lines, perception and action work together such that we actively alter our environment to better match our present expectations (and meet our goals) at the same time as we learn how the world resists our attempts to reduce error in our present beliefs. In this way, error minimization that includes active inference is not driven solely by accuracy motives, but by pragmatic concerns [26,27]. Specifically, within active inference, we represent the world according to a model that minimizes prediction error not just through its accuracy to the world, but through its ability to facilitate actions that make the world predictable (and produce desirable outcomes).

In summary, because higher-level predictions constrain lower-level sensory processing, perceivers see the world according to how well it fits with their existing model of reality (e.g. [28]). Specifically, sensory information from lower levels of the hierarchy is only propagated upwards if it fails to be explained by top-down expectations from higher-level models, which then learn from these errors to update predictions for the future. In essence, this provides a processing advantage by leveraging the forest to see the trees [2••,29]. Critically, *under this perceptual model, we do not see the world as it is, but rather the world that most makes sense of our sensory experiences in a way that most accurately predicts and pragmatically directs our future experiences*. For this reason, such a global error minimization may lead to perceptual illusions or blindspots in which error in processing any single feature will be ignored if doing so makes the most sense of the percept as a whole. Thus, through these bidirectional connections between hierarchically organized neural layers, cognitive systems can dynamically learn about, predict, and structure the physical environment.

Ideology in predictive processing and active inference

To the extent that motivated behavior can be described through the principle of global error minimization, coordinated social living presents a challenge. On its own, the formation of cooperative and often complex groups is a major way in which humans construct more predictable environments for ourselves, bringing the inherent precarity of the natural world more and more into our control [30••]. However, successful cooperation requires that group members accurately predict each other's behaviors, and people are themselves complex entities with their own potentially hidden motivations, goals, and desires. To solve this computational challenge, political systems may have arisen to restrict the possibility space of group living through a shared understanding of duties, rights, and responsibilities [31,32]. These systems not only dictate our expectations for social life and human behavior, but also allow the group to work from a shared set of assumptions about what beliefs and behaviors are acceptable and valued. In predictive processing terms, ideology is the hierarchical generative model specifying a set of predictions about the world and other people that facilitates group cooperation when shared with other members of the group. Although any particular set of rules and regulations may be arbitrary, and different groups of people can choose to set these rules in different ways, having a shared system that all members of a group adhere to reduces the variance in group members' thoughts and behaviors, rendering the social world more predictable [see Ref. 33•]. In doing so, ideologies have likely provided the social minimization of free energy that has allowed for group living.

While shared free energy minimization may function to constrain individuals' behavior to facilitate group-based cooperation, the processes involved in developing this consensus may also lead to polarization and extremism [1]. In particular, ideologies can bias the way people see the world, shaping their interpretations of relevant information such that two individuals with differing ideologies may draw opposite conclusions from the same data. People are likely to interpret such information as consistent with their ideologies, discounting disconfirming evidence in favor of reaffirming their existing beliefs [34–40].

Within a predictive processing perspective, these findings that initially come across as willful ignorance may follow directly from the way in which the brain optimizes the process of integrating expectation with sensation. Because ideologies function as sets of shared expectations that constrain the processing of sensory information, principles of predictive processing can be leveraged to understand why people do not update their ideological beliefs when faced with disconfirming information. First, predictive processing suggests that error alone does not guarantee that predictions will be updated. Rather, error must be sufficiently reliable and must present a sufficiently large deviation from prior beliefs to be incorporated into the model [5]. Failing this, expectations will dominate one's experience of the world, essentially rendering one blind to disconfirming evidence. Thus, someone may be presented with a great deal of disconfirming evidence that appears highly reliable to an outsider, but may nevertheless deem it unreliable due to its source [41] (e.g. a climate change denier ascribes low reliability to information from climate scientists, since they do not trust mainstream science). Along these lines, differential access to reliable information has been shown to account for apparent biases (e.g. overconfidence, [42,43]). Second, even in cases where there is sufficient reliable disconfirming evidence, recall that error must be minimized *globally*, across the entire system. In some cases, it may be better to allow for persistent prediction error in one domain in order to protect the rest of the model from significant disruption [44,45]. This is particularly complicated in the domain of ideological thought by the shared nature of ideology: in order to maintain the error-minimizing benefits of a shared narrative, individuals may need to sacrifice the accuracy of their beliefs in one domain in order to function harmoniously within the group's narrative [11••,33•,37,46,47].

If updating lower-level beliefs about ideologically relevant domains is rare, global ideological change is even rarer [48]. This can be partly understood as a logical consequence of the mechanisms described above, as it is difficult to accumulate enough evidence against any of the beliefs comprising one's ideology. Moreover, because generative models are hierarchical and ideologies are superordinate within that hierarchy, errors generally do not get propagated far enough up the hierarchy to result in a total revision of the ideology, as they may be better accounted for earlier on. In other words, local belief change may be sufficient for accommodating error without the need to revise the entire belief system (e.g. it is sufficient for a climate change denier to infer that climate scientists are untrustworthy, rather than updating their ideology as a whole). Additionally, because error must be minimized globally throughout the entire model, the extent to which an ideological model can be revised is kept in check by the need to avoid producing more error in domains where the ideology is relevant. Ideologies (e.g. social dominance orientation [49]) constrain expectations across a large number of experiential domains (e.g. the hierarchical structure of the family, the workplace, and society as a whole), so although error within a single domain may be reduced by adopting an entirely different ideology, doing so would disrupt the rest of the system and render chaotic all of the domains previously kept predictable by one's ideology. For these reasons, this framework makes the prediction that people will rarely make significant updates to their ideologies unless provided with a viable alternative that will minimize error as well as or better than the one they had before.

In the examples above, we have discussed how predictive processing mechanisms can contribute to biases in how people encode ideological information. Yet as noted previously, people can also minimize free energy through mechanisms of active inference, selectively attending to information and entering environments that are likely to support their underlying model of social behavior [e.g. 50]. As such, in social and political domains, one can minimize prediction error by selecting environments where people believe the same things and act in predictable ways, and selecting information sources that do not challenge one's preconceptions. For example, people are likely to seek out news sources that echo their existing ideological beliefs and frame facts in a way that does not threaten their worldview [51–53]. This is also exemplified in the formation of echo chambers in which little ideological disagreement is enacted or allowed, as selectively associating with others who share your worldview ensures that you are better able to predict their beliefs and behaviors, minimizing the influence of prediction errors on global beliefs [54,55] Further, individuals within an echo chamber will aim to act in highly predictable ways that ensure that they are not challenging the narratives they share with others in their group [56,57]. By taking action to ensure both that they are predictable to others in their group and that others are predictable to them, these shared narratives are reinforced and free energy is minimized. The reduction in prediction error resulting from isolating into echo chambers comes at the expense of accuracy, however, as selectively associating with those who share one's theories prevents the prediction errors that are necessary to update beliefs and maintain an accurate global worldview.

These processes of socially constrained free energy minimization could snowball, leading to polarized models of the world that are based on information selectively gained from within the group (which could be artificially generated or spuriously shared, for example, partisan news, 'fake news' [58] and conspiracy theories [59,60•,61,62]). These polarized models may be unable to adequately interpret social events outside of the echo chambers within which they were created, leading to increasing reliance on the group as a source of information and ideological reinforcement. Indeed, more extreme groups may be especially structured to afford ideological reinforcement, becoming more homogenous [54,63•,64] and authoritarian [12], while eschewing outside contact. The end result of this cycle may be polarized, homogeneous ideological groups that have little contact with ideas that are external to the group. Indeed, this process of social error minimization may explain the social sorting that political scientists have observed in recent decades, whereby social identities have become increasingly aligned with partisan (and ideological) groups (e.g.etc [65•,66]). In this way, society may become increasingly fractured into isolated extremist ideologies that reduce uncertainty in artificial, tightly controlled environments, but are inaccurate outside of those environments.

Although these cognitive mechanisms can promote ideological stability within people and groups, there remains a great deal of variation in the types of ideologies that individual members of groups hold at any given time, driven both by individual differences and contextual factors. For example, conservatives more so than liberals within a culture or community tend to be less open to new ideas and deviant behavior and more likely to demand adherence to social systems [67,68] with at least some of this variance accounted for by genetics [69]. Furthermore, the degree to which people are open to new ideas is critically shaped by how uncertain and threatened they feel in any given moment [70]. Predictive processing can be leveraged to understand such differences in several ways. First, individual differences in openness and tolerance may correspond to individual differences in how cognitive systems estimate precision or the degree to which errors will be accommodated to influence learning and experience. Specifically, in comparison to liberals, conservatives may be influenced by biological (e.g. genetic) or situational (e.g. high arousal resulting in uncertainty or threat) factors that lead to global neuromodulatory responses that tend to underestimate the reliability of incoming error, decreasing the degree to new information will update expectations so as to 'double-down' on prior predictions. Indeed, work in neuroscience shows that neuromodulatory fluctuations can change the global properties of cognitive computation to alter the degree to which incoming information is accommodated [71]. For example, acetylcholine shapes the precision of bottom-up neural signals [72], while MDMA and dopamine (D1) shape the excitability of neurons that encode prediction errors [73]. Indeed, potentially as a result of overestimating the strength of predictions, individual differences in such neuromodulators have been linked to clinical hallucinations and delusions [74], experiences that have direct analogues to biased ideological thought. Given this, work in computational psychology has recently begun to examine how these parameters of global cognition are associated with specific ideological biases [75].

At the same time, rather than general differences in how precision is estimated, individual differences in openness to experience between conservatives and liberals may simply be a product of different needs and experiences that are accommodated through error minimization. In particular, conservatives, in comparison to liberals, may be more sensitive to certain needs over others, such as needs for security, purity and the preservation of traditions [76–78]. Within an active inference framework, it is possible both conservatives and liberals may estimate precision in similar ways yet, due to differences in motivational parameters, still prioritize learning information that addresses different needs and goals, leading to different models in the long run. Providing some support for this, recent work indicates that in some domains, liberals, rather than conservatives, are more sensitive to negative information, tentatively suggesting that to some extent, differences in openness to error-producing information may reflect domain-specific priorities and knowledge [79]. These different possibilities indicate that, although in its infancy, applying-specific computations principles to the processes underlying ideological thought will likely lead to new predictions and competing hypotheses from perspectives that can be disambiguated at a computational level of analysis.

Conclusions

In this paper, we placed ideological thought within the computational framework of predictive processing. In particular, we examine how the basic error minimizing neural dynamics that organize lower-level perception and motor action may scale up to shape the neural organization of ideological representation and decision-making. In this way, ideological reasoning and bias, like perception, may be understood as a product of the pragmatic constraints of error minimization in a complex world. These pragmatic constraints may outline a path for the development of shared ideologies that globally structure and direct social life. Specifically, since group members learn from, motivate, and regulate each other's beliefs, mutually constrained error minimization may align individual experience and behavior with shared beliefs and schemas that allow us to predict and cooperate with others. However, as a byproduct, particularly under conditions of uncertainty, this mutually constrained error minimization may produce increased polarization and extremism. Although it is unclear at this point whether these principles simply provide a computational account for mechanisms already known in the literature, or whether they provide additional mechanisms, moving to a more computational approach to the study of political behavior should allow for more formal tests of competing hypotheses using a common language.

Conflict of interest statement

Nothing declared.

Acknowledgements

We thank Arie Kruglanski, Victoria Davis, and the members of the Social Cognitive Science lab at University of Toronto for helpful comments on earlier versions of this manuscript. This work was partially funded by the Social Sciences and Humanities Research Council [SSHRC-501259] and the Natural Sciences and Engineering Research Council of Canada [RGPIN-2018-05946] to WAC and a SSHRC Doctoral Fellowship to SA.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Bruner JS: **On perceptual readiness**. *Psychol Rev* 1957, **64**:123-152.

2. Clark A: **Whatever next? Predictive brains, situated agents, and the future of cognitive science**. *Behav Brain Sci* 2013, **36**:181-204.
•• This paper provides a comprehensive account of the predictive processing framework, synthesizing across decades of work in neuroscience, computer science, and philosophy. The potential implications of this framework for understanding perception, action, agency, and consciousness are also discussed.

3. Clark A: **Perceiving as predicting**. *Percept Modal* 2014:23-43 http://dx.doi.org/10.1093/acprof:oso/9780199832798.003.0002.

4. Friston K: **The free-energy principle: a unified brain theory?** *Nat Rev Neurosci* 2010, **11**:127-138.

5. Clark A: *Surfing Uncertainty: Prediction, Action, and the Embodied Mind*. Oxford University Press; 2015.

6. Jost JT: **The end of the end of ideology**. *Am Psychol* 2006, **61**:651-670.

7. Durkheim E: **What is a social fact?** In *The Rules of Sociological Method: Contemporary Social Theory*. Edited by Luke S. Palgrave; 1982:50-59.

8. Heine SJ, Proulx T, Vohs KD: **The meaning maintenance model: on the coherence of social motivations**. *Pers Soc Psychol Rev* 2006, **10**:88-110.

9. Solomon S, Greenberg J, Pyszczynski T: **A terror management theory of social behavior: the psychological functions of self-esteem and cultural worldviews**. *Adv Exp Soc Psychol* 1991, **24**:93-159 http://dx.doi.org/10.1016/s0065-2601(08)60328-7.

10. Jost JT, Hunyady O: **The psychology of system justification and the palliative function of ideology**. *Eur Rev Soc Psychol* 2003, **13**:111-153.

11. •• Jost JT, Ledgerwood A, Hardin CD: **Shared reality, system justification, and the relational basis of ideological beliefs**. *Soc Pers Psychol Compass* 2008, **2**:171-186.

The authors describe how relational motives shape the adoption and maintenance of different political ideologies and how these pressures disproportionately lend themselves to the transmission of more conservative ideologies that reinforce the status quo.

12. Hogg MA, Adelman J: **Uncertainty-identity theory: extreme groups, radical behavior, and authoritarian leadership**. *J Soc Issues* 2013, **69**:436-454.

13. Jern A, Chang KK, Kemp C: **Belief polarization is not always irrational**. *Psychol Rev* 2014, **121**:206-224.

14. Huang Y, Rao RPN: **Predictive coding**. *Wiley Interdiscip Rev Cogn Sci* 2011, **2**:580-593.

15. Rao RPN, Ballard DH: **Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects**. *Nat Neurosci* 1999, **2**:79-87.

16. Ward J: **Individual differences in sensory sensitivity: a synthesizing framework and evidence from normal variation and developmental conditions**. *Cogn Neurosc* 2019, **10**:139-157.

17. Lawson RP, Mathys C, Rees G: **Adults with autism overestimate the volatility of the sensory environment**. *Nat Neurosci* 2017, **20**:1293-1299.

18. Sterzer P, Adams RA, Fletcher P, Frith C: **The predictive coding account of psychosis**. *Biol Psychiatry* 2018, **84**:634-643.

19. Ainley V, Apps MA, Fotopoulou A, Tsakiris M: **'Bodily precision': a predictive coding account of individual differences in interoceptive accuracy**. *Philos Trans R Soc B Biol Sci* 2016, **371**:20160003.

20. Lutz A, Mattout J, Pagnoni G: **The epistemic and pragmatic value of non-action: a predictive coding perspective on meditation**. *Curr Opin Psychol* 2019, **28**:166-171.

21. Friston KJ, Schwartenbeck P, FitzGerald T, Moutoussis M, Behrens T, Dolan RJ: **The anatomy of choice: dopamine and decision-making**. *Philos Trans R Soc B Biol Sci* 2014, **369**:20130481.

22. Friston K: **The free-energy principle: a rough guide to the brain?** *Trends Cogn Sci* 2009, **13**:293-301.

23. Rensink RA, O'Regan JK, Clark JJ: **To see or not to see: the need for attention to perceive changes in scenes**. *Psychol Sci* 1997, **8**:368-373.

24. Friston K, Daunizeau J, Kilner J, Kiebel SJ: **Action and behavior: a free-energy formulation**. *Biol Cybern* 2010, **102**:227-260.

25. Brown H, Friston K, Bestmann S: **Active inference, attention, and motor preparation**. *Front Psychol* 2011, **2**.

26. Gelpi R, Cunningham WA, Buchsbaum D: **Belief as a non-epistemic adaptive benefit**. 2020, https://doi.org/10.31234/OSF.IO/6YTHF.

27. Kruglanski AW, Jasko K, Friston K: **All thinking is 'wishful' thinking**. *Trends Cogn Sci* 2020, **24**:413-424.

28. Hohwy J, Roepstorff A, Friston K: **Predictive coding explains binocular rivalry: an epistemological review**. *Cognition* 2008, **108**:687-701.

29. Hochstein S, Ahissar M: **View from the top: hierarchies and reverse hierarchies in the visual system**. *Neuron* 2002, **36**:791-804.

30. •• Veissière SPL, Constant A, Ramstead MJD, Friston KJ, Kirmayer LJ: **Thinking through other minds: a variational approach to cognition and culture**. *Behav Brain Sci* 2019, **43**:1-97 http://dx.doi.org/10.1017/s0140525x19001213.

The authors extend the principles of active inference and free energy minimization to provide an integrative account of how culture shapes and is shaped by the development of shared habits, norms, and expectations, which organize individuals' attention and behavior.

31. Alford JR, Hibbing JR: **The origin of politics: an evolutionary theory of political behavior**. *Perspect Politics* 2004, **2**:707-723.

32. Laswell HD: **Politics: who gets what, when, how**. *Meridian Books*. 1958.

33. • Kitto K, Boschetti F: **Attitudes, ideologies and self-organization: information load minimization in multi-agent decision making**. *Adv Complex Syst* 2013, **16**:1350029.

This paper provides a computational framework for modeling attitude change and the emergence of ideology within groups of interacting agents.

34. Kahan DM: **The politically motivated reasoning paradigm, part 1: what politically motivated reasoning is and how to measure it**. *Emerging Trends in the Social and Behavioral Sciences*. John Wiley & Sons, Inc.; 2016:1-16.

35. Kunda Z: **The case for motivated reasoning**. *Psychol Bull* 1990, **108**:480-498.

36. Taber CS, Lodge M: **Motivated skepticism in the evaluation of political beliefs**. *Am J Polit Sci* 2006, **50**:755-769.

37. Kahan DM: **Ideology, motivated reasoning, and cognitive reflection: an experimental study**. *SSRN Electron J* 2012, **8**:407-424 http://dx.doi.org/10.2139/ssrn.2182588.

38. Lord CG, Ross L, Lepper MR: **Biased assimilation and attitude polarization: the effects of prior theories on subsequently considered evidence**. *J Pers Soc Psychol* 1979, **37**:2098-2109.

39. Redlawsk DP, Civettini AJW, Emmerson KM: **The affective tipping point: do motivated reasoners ever "get it"?** *Polit Psychol* 2010, **31**:563-593.

40. Taber CS, Cann D, Kucsova S: **The motivated processing of political arguments**. *Polit Behav* 2009, **31**:137-155.

41. Nisbet EC, Cooper KE, Garrett RK: **The partisan brain: how dissonant science messages lead conservatives and liberals to (dis) trust science**. *Ann Am Acad Polit Soc Sci* 2015, **658**:36-66.

42. Moore DA, Small DA: **Error and bias in comparative judgment: on being both better and worse than we think we are**. *J Pers Soc Psychol* 2007, **92**:972-989.

43. Moore DA, Healy PJ: **The trouble with overconfidence**. *Psychol Rev* 2008, **115**:502-517.

44. Dalege J, Borsboom D, van Harreveld F, van den Berg H, Conner M, van der Maas HL: **Toward a formalized account of attitudes: the Causal Attitude Network (CAN) model**. *Psychol Rev* 2016, **123**:2-23.

45. Cushman F: **Rationalization is rational**. *Behav Brain Sci* 2019, **43**: E28 http://dx.doi.org/10.1017/s0140525x19001730.

46. Echterhoff G, Higgins ET, Levine JM: **Shared reality: experiencing commonality with others' inner states about the world**. *Perspect Psychol Sci* 2009, **4**:496-521.

47. Jost JT, Hunyady O: **Antecedents and consequences of system-justifying ideologies**. *Curr Direct Psychol Sci* 2005, **14**:260-265.

48. Peterson JC, Smith KB, Hibbing JR: **Do people really become more conservative as they age?** *J Polit* 2020, **82**:600-611 http://dx.doi.org/10.1086/706889.

49. Sidanius J, Pratto F, Martin M, Stallworth LM: **Consensual racism and career track: some implications of social dominance theory**. *Polit Psychol* 1991, **12**:691-721.

50. Karlsson N, Loewenstein G, Seppi D: **The ostrich effect: selective attention to information**. *J Risk Uncertain* 2009, **38**:95-115.

51. Garrett RK: **Echo chambers online?: Politically motivated selective exposure among Internet news users**. *J Comput Mediat Commun* 2009, **14**:265-285.

52. Stroud NJ: **Polarization and partisan selective exposure**. *J Commun* 2010, **60**:556-576.

53. Iyengar S, Hahn KS: **Red media, blue media: evidence of ideological selectivity in media use**. *J Commun* 2009, **59**:19-39.

54. Williams HTP, McMurray JR, Kurz T, Hugo Lambert F: **Network analysis reveals open forums and echo chambers in social media discussions of climate change**. *Glob Environ Change* 2015, **32**:126-138.

55. Colleoni E, Rozza A, Arvidsson A: **Echo chamber or public sphere? Predicting political orientation and measuring political homophily in twitter using big data**. *J Commun* 2014, **64**:317-332.

56. Centola D: **The spread of behavior in an online social network experiment**. *Science* 2010, **329**:1194-1197.

57. Centola D, Macy M: **Complex contagions and the weakness of long ties**. *Am J Sociol* 2007, **113**:702-734.

58. Pennycook G, Rand DG: **Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking**. *J Pers* 2019, **88**:185-200 http://dx.doi.org/10.1111/jopy.12476.

59. Miller JM, Saunders KL, Farhart CE: **Conspiracy endorsement as motivated reasoning: the moderating roles of political knowledge and trust**. *Am J Polit Sci* 2015, **60**:824-844.

60. • Flynn DJ, Nyhan B, Reifler J: **The nature and origins of misperceptions: understanding false and unsupported beliefs about politics**. *Polit Psychol* 2017, **38**:127-150.
Review article summarizes recent work in political science explaining directional motivated reasoning and implications for understanding when people believe unsupported claims. Also discusses implications for measuring misperceptions in survey research and broader implications for democracy.

61. van Prooijen JW, Acker M: **The influence of control on belief in conspiracy theories: conceptual and applied extensions**. *Appl Cogn Psychol* 2015, **29**:753-761.

62. van Prooijen JW, Krouwel APM, Pollet TV: **Political extremism predicts belief in conspiracy theories**. *Soc Psychol Pers Sci* 2015, **6**:570-578.

63. • Boutyline A, Willer R: **The social structure of political echo chambers: variation in ideological homophily in online networks**. *Polit Psychol* 2017, **38**:551-569 http://dx.doi.org/10.1111/pops.12337.
This article examines the relationship between political ideology and political homophily on social media, arguing that conservatives and those with more extreme ideologies prefer certainty and are more likely to have homophilous social networks that reaffirm their views.

64. Schulz-Hardt S, Frey D, Lüthgens C, Moscovici S: **Biased information search in group decision making**. *J Pers Soc Psychol* 2000, **78**:655-669.

65. • Mason L: *Uncivil Agreement: How Politics Became Our Identity*. University of Chicago Press; 2018.
Uses a social identity approach to understand increasing political polarization in the United States, arguing that Americans have become increasingly sorted into political groups that are largely overlapping with nonpolitical groups and identities.

66. Mason L, Wronski J: **One tribe to bind them all: how our social group attachments strengthen partisanship**. *Polit Psychol* 2018, **39**:257-277.

67. Hibbing JR, Smith KB, Alford JR: *Predisposed: Liberals, Conservatives, and the Biology of Political Differences*. Routledge; 2013.

68. Jost JT, Glaser J, Kruglanski AW, Sulloway FJ: **Political conservatism as motivated social cognition**. *Psychol Bull* 2003, **129**:339-375.

69. Alford JR, Funk CL, Hibbing JR: **Are political orientations genetically transmitted?** *Am Polit Sci Rev* 2005, **99**:153-167.

70. Haas IJ, Cunningham WA: **The uncertainty paradox: perceived threat moderates the effect of uncertainty on political tolerance**. *Polit Psychol* 2014, **35**:291-302.

71. Aston-Jones G, Cohen JD: **An integrative theory of locus coeruleus-norepinephrine function: adaptive gain and optimal performance**. *Ann Rev Neurosc* 2005, **28**:403-450.

72. Moran RJ, Campo P, Symmonds M, Stephan KE, Dolan RJ, Friston KJ: **Free energy, precision and learning: the role of cholinergic neuromodulation**. *J Neurosci* 2013, **33**:8227-8236.

73. Kanai R, Komura Y, Shipp S, Friston K: **Cerebral hierarchies: predictive processing, precision and the pulvinar**. *Philos Trans R Soc B Biol Sci* 2015, **370**:20140169.

74. Brown H, Adams RA, Parees I, Edwards M, Friston K: **Active inference, sensory attenuation and illusions**. *Cogn Process* 2013, **14**:411-427.

75. Zmigrod L, Eisenberg I, Bissett P, Robbins TW, Poldrack R: **A data-driven analysis of the cognitive and perceptual attributes of ideological attitudes**. *PsyArXiv* 2020 http://dx.doi.org/10.31234/osf.io/dgaxr.

76. Haidt J, Graham J: **When morality opposes justice: conservatives have moral intuitions that liberals may not recognize**. *Soc Justice Res* 2007, **20**:98-116.

77. Jost JT, Amodio DM: **Political ideology as motivated social cognition: behavioral and neuroscientific evidence**. *Motiv Emot* 2012, **36**:55-64.

78. Smith KB, Oxley DR, Hibbing MV, Alford JR, Hibbing JR: **Linking genetics and political attitudes: reconceptualizing political ideology**. *Polit Psychol* 2011, **32**:369-397.

79. Fiagbenu ME, Proch J, Kessler T: **Of deadly beans and risky stocks: political ideology and attitude formation via exploration depend on the nature of the attitude stimuli**. *Br J Psychol* 2019 http://dx.doi.org/10.1111/bjop.12430.