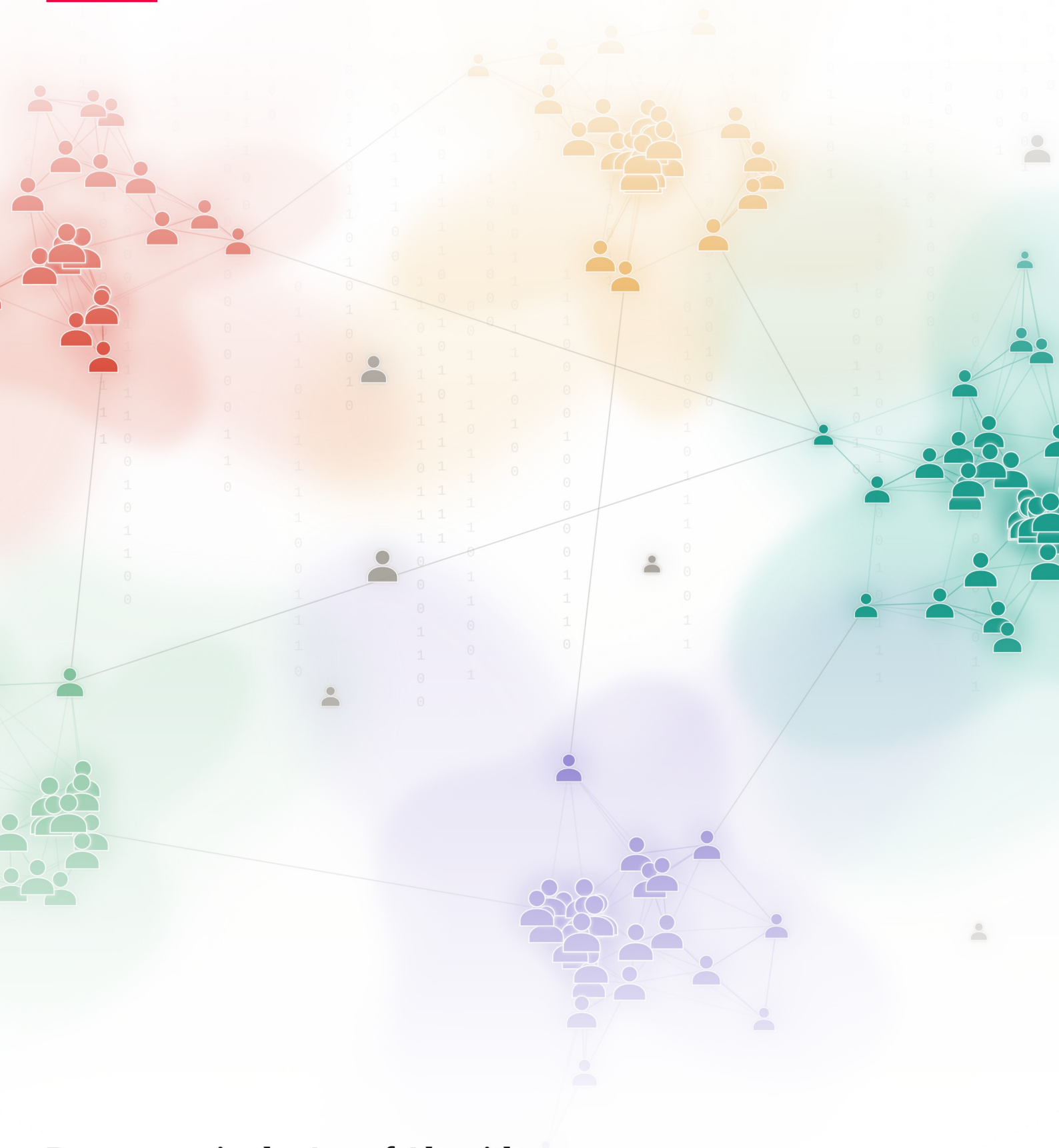




Democracy in the Age of Algorithms

Why Reliable Information Is Not Enough to Sustain Public Debate

Anna Knorr & Małgorzata Kossowska

EDITORIAL NOTE

The following text is an adapted version of Anna Knorr's article "[The Crisis of Informed Decisions: Our Brains in the Epistemic Architecture of Digital Spaces](#)", published on the blog of the European Commission's Joint Research Centre (JRC).

Contemporary democracy rests on the assumption that citizens make better decisions when they have access to reliable information. Credible knowledge allows them to evaluate the actions of those in power, understand the consequences of proposed changes, and make informed choices about their representatives. This goal - at least in theory - was meant to be achieved by the digital revolution. Never before has access to information been so easy, fast, and widespread.

Yet the paradox of our times is that more information does not always mean a better understanding of reality. Although we have access to vast amounts of knowledge every day, reaching shared conclusions has become increasingly difficult. Information increasingly fails to fulfil its basic function – instead of helping us better understand reality, it merely confirms what we already believe. New facts do not always prompt us to change our views. On the contrary, when they contradict our beliefs, they tend to be interpreted in ways that reinforce – and sometimes even strengthen – those beliefs. As a result, practices such as fact-checking rarely lead to lasting changes in opinion, while social and political divisions continue to deepen.

KEY TAKEAWAY

The effective functioning of a democratic state requires not only free elections but also trust in institutions, open debate, and a shared understanding of basic facts. When agreement about what is true begins to erode, so do the mechanisms that sustain democratic coexistence.

For decades, the prevailing belief was that the root cause of mistaken views was primarily a lack of knowledge – that providing people with reliable information would be enough for them to, guided by reason, verify their beliefs and change their minds. This approach, known in the literature as the information deficit model, still underlies many educational campaigns, science communication initiatives, and public-institution communication strategies today.

In practice, however, reality proves far more complex. Numerous studies show that although disinformation undoubtedly deepens social divisions, merely correcting false information rarely leads to lasting changes in belief. Exposure to well-documented facts can even reinforce a recipient's prior opinions. Two people may read the same article, watch the same material, or review the same scientific report, and yet draw entirely different conclusions from it.

Where does this paradox come from? Increasingly, the cognitive sciences provide answers. From their perspective, the human brain does not function like a computer that first gathers data and only then draws conclusions. On the contrary, the brain continuously generates predictions about the surrounding world, interpreting incoming information through the lens of prior experience, knowledge, and beliefs. This means

that facts never reach an “empty mind” – they are always confronted with an already-existing model of reality, which shapes how they are interpreted.

In this text, we show how predictive coding theory helps explain why it is so difficult to change our beliefs in the face of new information, and how the contemporary digital environment may reinforce the mechanisms that govern our cognition.

This is not a marginal issue. Three-quarters of Europeans – especially young people – now obtain information about the world primarily from the internet, where digital platforms compete for our attention. Rather than fostering a shared understanding of reality, their architecture may reinforce existing beliefs, deepen social divisions, and shape what we come to accept as true.

The Digital Information Environment

The way we perceive the world is shaped not only by how the human mind operates, but also by the information environment itself. The internet is often portrayed as a vast library or a modern marketplace of ideas, where different viewpoints compete freely. There is considerable truth to this image, but it does not fully capture how contemporary digital platforms actually function.

Increasingly, the internet resembles a personal editor that decides each day which content will appear on our screens and which will remain out of view. As a result, most of the information we encounter on social media or in search engines is not the outcome of a deliberate choice on the user's part. What we read, watch, or hear is increasingly determined by recommendation systems that analyse our previous behaviour.

Every click, every pause while scrolling, every like or comment becomes a signal that algorithms use to predict which content will most effectively capture our attention. Personalisation itself is not inherently negative – it makes it easier to find a film, a book, or a restaurant that matches our interests. Difficulties arise, however, when the same mechanism begins to determine which information about politics, health, science, or social issues reaches individual audiences. In such an environment, the reach of information increasingly depends not on its credibility but on its ability to capture and hold attention. This is precisely why content that provokes strong emotions, controversy, or outrage tends to spread faster than calm, well-documented explanations.

Social media operate according to the logic of the attention economy. Their primary goal is to keep users on the platform for as long as possible, since time spent on screen translates directly into advertising revenue. For this reason, algorithms favour content that provokes strong emotions – surprise, outrage, fear, or anger. Calm, balanced, and nuanced material tends to lose out to more clear-cut and controversial messaging. A self-reinforcing mechanism emerges: the more often we react to a given type of content, the more often the platform shows it to us. And the more often we see it, the more typical, widespread, and credible it appears.

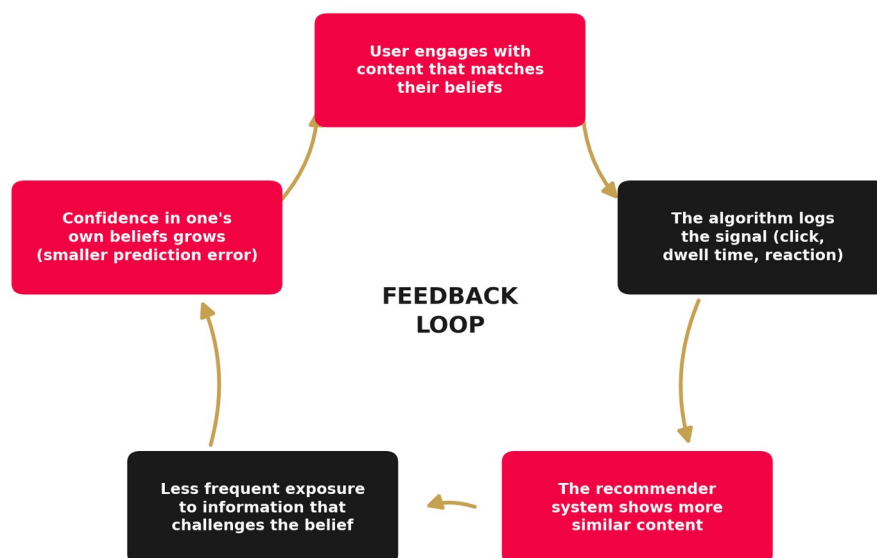
At the same time, the internet makes it easier to form communities of people with similar views. For many users, this is of great value – it allows them to find others with shared interests or experiences. But it also has a less obvious consequence: even beliefs

held by a small group can appear dominant if nearly all the opinions one encounters confirm the same way of thinking.

In addition, social media equip every piece of content with a set of popularity signals: numbers of likes, shares, comments, or views. Although these metrics primarily reflect user activity, they are easily mistaken for a measure of credibility. Popularity begins to be perceived as evidence of truth.

The danger lies in the fact that all these mechanisms operate simultaneously. Algorithms display content similar to what has already captured our attention. Online communities reinforce the sense that our views are shared by the majority. Popularity indicators reassure us that, since many people think alike, we are unlikely to be wrong. As a result, the digital environment does not merely convey information – it increasingly co-determines which information we consider worth our attention and which we are willing to trust.

How algorithms reinforce beliefs: a self-reinforcing cycle



Prepared by Behaviour in Crisis Lab, Jagiellonian University, based on the report "Democracy in the Age of Algorithms"

Figure 1. A self-reinforcing cycle: user reactions train the algorithm, which in turn shapes further user reactions.

How the Brain Processes Information

Prediction

At every moment, an enormous number of sensory stimuli reach our brain. To navigate this constant stream of information, the brain does not analyse every signal from scratch. Instead, it draws on prior experience to generate predictions about what we are about to see, hear, or feel. These predictions are based on our existing model of the world – accumulated knowledge, experience, and beliefs. When new information arrives, it is compared against these expectations, and the brain evaluates whether the existing model of reality needs to be revised.

This means that perceiving the world is not a passive reception of incoming stimuli. It is a process of actively constructing reality, in which prior beliefs continually collide with new experiences. If what we experience matches our expectations, the existing model of the world is confirmed – beliefs not only remain unchanged but are often further reinforced.

The situation changes when incoming information contradicts what we expected. At that point, the brain registers what is known as a prediction error. The occurrence of a prediction error does not, however, automatically lead to a change of view. The brain must first determine whether to revise its own model of the world or, instead, to reinterpret – or even reject – the new information in order to preserve its existing understanding of reality. This choice is shaped primarily by two factors: the credibility of the new information and the strength of one's conviction in the correctness of the existing worldview.

If prior beliefs are not strongly entrenched and the new information appears credible, we typically update our understanding of reality. But if we are deeply convinced of our own correctness and the incoming information is ambiguous or doubtful, prior beliefs retain the upper hand, and new evidence has little influence on our conclusions. Put simply: the greater our confidence in our own beliefs, the harder they are to change. Crucially, this sense of confidence does not arise out of nowhere – it is shaped, to a large extent, by the information we have previously encountered.

It is precisely here that the contemporary digital environment takes on particular significance. The internet does not deliver information neutrally. Online platform algorithms primarily promote content similar to what has previously captured our attention. As a result, we are far more likely to encounter messages that confirm our existing views than information that might challenge them. This asymmetry gradually reinforces the certainty of our own beliefs. The more often we encounter a similar message, the more we become convinced that this particular picture of the world best describes reality. Over time, our views become increasingly resistant to change.

This mechanism is well illustrated by the example of climate change. If a user repeatedly watches material that challenges the scientific consensus on this issue, the recommendation system will identify this interest and begin suggesting further content in a similar vein. Over time, the user's information feed will become dominated by material that reinforces their prior beliefs, while content presenting scientific findings will appear less and less frequently. As a result, information that contradicts the user's existing worldview may come to seem less credible, or even suspect. The algorithm does not merely reflect the user's prior preferences – it gradually entrenches them.

The shaping of beliefs, however, is influenced not only by algorithms. Other features of the contemporary information environment are equally significant, such as the functioning of online communities of like-minded individuals, the double-counting bias – encountering the same information multiple times as though it were independent confirmation – and the power of social signals such as the number of likes, shares, and comments. Being surrounded by people who think alike can lend an appearance of credibility even to beliefs with little factual grounding.

Active Inference

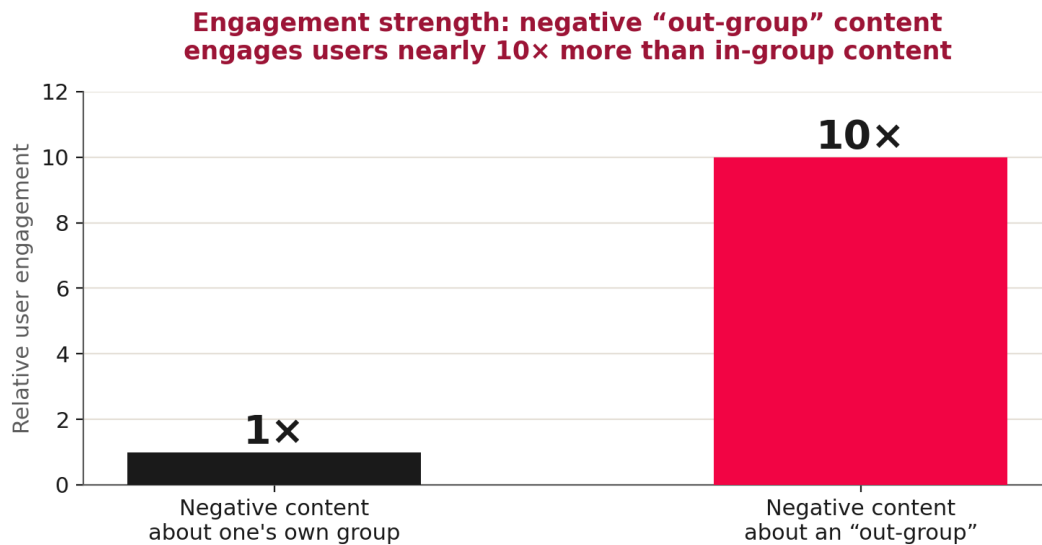
If predictive coding theory explains why it is so difficult for us to change our minds, the next concept – active inference – takes this a step further. It shows that human beings are not passive recipients of information. We do not merely interpret reality through the lens of our own beliefs; we actively select environments that help sustain those beliefs. In other words, we do not wait for the world to surprise us. Far more often, we seek out places, people, and information that make the world feel more predictable.

From a psychological standpoint, this is entirely understandable. Uncertainty is costly, both cognitively and emotionally. It requires continually revising one's beliefs, making new decisions, and coming to terms with the possibility of error. The brain, like any efficient system, seeks to minimise this effort – which is why we favour environments in which our predictions are most often confirmed. This is not, in itself, a flaw. On the contrary, it allows us to quickly recognise patterns, learn, and function effectively in a complex world. The problem arises only when this natural tendency meets the architecture of contemporary digital platforms.

Social media do not create our need for confirmation. They simply learn, very quickly, how to exploit it. Every user reaction – a click, a comment, a share, or even a few extra seconds of lingering attention – becomes information for the recommendation system. The algorithm does not evaluate why a piece of content caught our interest; all that matters is that it did. As a result, we receive an ever-growing supply of similar material.

If we updated our internal model of the world without reflection, treating all incoming information as equally credible, we would quickly lose the ability to distinguish genuine patterns from random noise. The pursuit of certainty is therefore not a flaw of the human mind – it is one of its fundamental features. The problem arises specifically within a digital information environment where content is not evenly distributed. Under these conditions, the natural tendency to seek confirmation of one's own beliefs can lead to their gradual entrenchment, and in extreme cases, to radicalisation.

This is well illustrated by findings from a recent study, which showed that content portraying groups perceived as “outsiders” in a negative light engages users nearly ten times more strongly than negative content about one's own group. From the perspective of active inference theory, this phenomenon can be interpreted as a reflection of the brain's drive to minimise prediction error – we are more inclined to engage with content that confirms our prior expectations.



Source: Rasmussen, Rathje, Bavel & Robertson (2025); prepared by Behaviour in Crisis Lab, Jagiellonian University

Figure 2. Content portraying out-groups negatively engages users nearly ten times more strongly than negative content about one's own group.

Epistemic Motivation

The process, however, does not end there. Whether we accept new information depends not only on its quality but also on what we are actually looking for. Psychologists refer to this as epistemic motivation – the drive that guides our pursuit of knowledge. Not everyone seeks information for the same reason. Some primarily want to find the correct answer, even if that means changing their own mind; for them, new evidence represents an opportunity to better understand the world. Others seek mainly confirmation of what they already believe to be true; information that challenges their views may be perceived not as a learning opportunity but as a threat to their identity or their sense of belonging to a particular group.

What is more, something even more surprising can occur: confrontation with opposing arguments can actually strengthen prior beliefs. The more strongly someone identifies with a particular worldview, the greater the temptation to treat critical information as further proof that “the other side” is wrong. It is therefore hardly surprising that researchers argue that “all thinking is, to some extent, wishful thinking.” Our reasoning is almost always intertwined with emotional needs, a sense of security, and social identity.

Can We Still Make Fact-Based Decisions?

If the way we perceive the world depends not only on facts but also on the information environment in which we operate, a more fundamental question arises: what conditions must be met for democracy to function well?

Since the Enlightenment, it has been assumed that access to reliable information fosters more informed decision-making, and that open debate allows positions to gradually converge and consensus to be built. Contemporary research shows, however, that access to information alone is not sufficient. How we understand reality is also shaped by the way information reaches us and by the cognitive mechanisms through which we interpret it.

This does not mean that democracy is ceasing to function, or that digital technologies are inherently a threat. It does mean, however, that its success increasingly depends on the quality of the information environment. When citizens inhabit divergent information worlds and encounter a diminishing range of perspectives, it becomes harder to conduct debate grounded in shared facts, to build trust, and to make decisions about matters of common concern.

The challenge, therefore, is not merely to combat disinformation. Equally important is the creation of an information environment that encourages the confrontation of different perspectives, enables the critical evaluation of information, and helps preserve a shared space for public debate.

The most important – and, at the same time, most optimistic – conclusion to emerge from the research discussed in this report is that today's information landscape is not an inevitable consequence of technological progress. It is the result of specific design, business, and regulatory decisions. Given this, the same systems can be designed differently.

This is not about abandoning personalisation or restricting freedom of expression. It is, rather, about creating an information environment that does not systematically reinforce our cognitive tendency to retreat into our own beliefs. Both digital platform designers and public institutions have an increasingly important role to play in this effort. European regulations – such as the Digital Services Act and the AI Act – show that the functioning of recommendation systems is becoming a subject of debate not only in technology, but in politics and ethics as well.

At the same time, regulation alone will not be enough. A better understanding of the mechanisms that govern human cognition is also required. Knowledge from cognitive psychology and neuroscience should become one of the foundations for the design of future digital environments. Democracy does not need a society that agrees on everything. It needs a society capable of disagreeing on the basis of shared, commonly accepted facts. In a world increasingly shaped by algorithms, this capacity may prove to be one of the most valuable assets a democratic state possesses.

References

- Friston, K. J., FitzGerald, T., Rigoli, F., Schwartenbeck, P. & Pezzulo, G. (2017). Active Inference: A Process Theory. *Neural Computation*, 29(1), 1–49.
- Huang, Y. & Rao, R.P.N. (2011). Predictive coding. *WIREs Cognitive Science*, 2, 580–593.
- Kossowska, M. & Sroczyńska, K. (2026). Czy ten świat się kończy [Is This the End of the World]. *Smak Słowa*.
- Kossowska, M., Czarnek, G. & Szwed, P. (2023). Political ideology and belief change in the face of counterevidence. *European Journal of Social Psychology*, 53(6), 1157–1171.
- Kossowska, M. & Knorr, A. (2025). Facts won't save us. *ISSEP*.
- Kruglanski, A., Jasko, K. & Friston, K. (2020). All Thinking is 'Wishful' Thinking. *Trends in Cognitive Sciences*, 24, 413–424.

- Marino, M., Iacono, R. & Mollerstrom, J. (2024). (Mis-)Perceptions, information, and political polarization: A survey and a systematic literature review. *European Journal of Political Economy*, 85, 102578.
- McCoy, J., Press, B., Somer, M. & Tuncel, O. (2022). Reducing pernicious polarization: A comparative historical analysis of depolarization. *Carnegie Endowment for International Peace*.
- Parr, T., Pezzulo, G. & Friston, K. J. (2022). *Active inference: the free energy principle in mind, brain, and behavior*. MIT Press.
- Rasmussen, J., Rathje, S., Bavel, J. & Robertson, C. (2025). Negativity and Identity Language Have Additive Effects on Online News Consumption.
- Tipping, M. E. (2003). Bayesian inference: An introduction to principles and practice in machine learning. In *Summer school on machine learning* (pp. 41–62). Springer.
- Walter, N., Cohen, J., Holbert, R. L. & Morag, Y. (2020). Fact-Checking: A Meta-Analysis of What Works and for Whom. *Political Communication*, 37(3), 350–375.
- Weinstein, N. D. (1980). Unrealistic optimism about future life events. *Journal of Personality and Social Psychology*, 39(5), 806–820.
- Williams, D. (2020). Predictive coding and thought. *Synthese*, 197, 1749–1775.

CONTACT

Behaviour in Crisis Lab, Jagiellonian University, 6 Ingardena St., 30-060 Kraków, Poland • Phone: +48 12 663 11 98

Email: malgorzata.kossowska@uj.edu.pl (Lab Lead) • florentyna.zakrzewska@student.uj.edu.pl (Media contact) • wojciech.lesniakiewicz@uj.edu.pl (Editorial contact) **Website:** becrilab.id.uj.edu.pl