Title: Evidencing the impact of misinformed and disinformed beliefs on individual and group behaviors.

Magda Osman[1,2]

1. Judge Business School, University of Cambridge, Trumpington St, Cambridge CB2 1AG.

2. Centre for Decision Research, Leeds Business School, University of Leeds, Leeds, West Yorkshire, LS2 9JT, UK.

Correspondence: m.osman@jbs.cam.ac.uk

Keywords: Beliefs, Misinformation and Disinformation, Causal analysis, Situational analysis, Explanatory and Predictive modelling

Funding details. There was no financial support in preparation for this article.

Disclosure statement. The author reports there are no competing interests to declare.

Introduction

How much causal impact do beliefs have on our behaviors? The task of understanding beliefs and, what if any, consequences follow from them can unravel quickly in the absence of a framework, taxonomy, and predictions. Granados Samayoa and Albarracín's target article "Understanding Belief-Behavior Correspondence: Beliefs and Belief-to-Behavior Inferences" provide answers to help the research community avoid unravelling. Belief is defined, "…as a subjective probability that an entity (e.g., person, place, object, or behavior) and an attribute or outcome are linked.". Because we cannot assume beliefs have equal causal efficacy to bring about a behavior, Granados Samayoa and Albarracín provide a classification system of belief types (existence beliefs, descriptive beliefs, outcome beliefs). In combination with their organizing principle "a belief-to-behavior inference (i.e., the reasoning that connects beliefs to a behavior) is necessary for beliefs to exert causal impacts on behavior", predictions about behavioral outcomes from beliefs can be derived.

All this raises several practical implications and applications, three of which are discussed here using misinformed and disinformed beliefs as illustrators. First, the internal causal chain from beliefs, appraisals, motivations, to behavioral outcomes can be misinterpreted from an observer's perspective, and without careful use of measurements. Second, using causal analysis to predict group level behavior from beliefs requires situational factors to fill in the gaps. Third, causal analysis of belief-to-behaviors requires a concerted adherence to the Principle of Compatibility (Ajzen, 1988, 2020).

Misapprehensions influencing causal analysis of beliefs-to-behavior

Granados Samayoa and Albarracín make the case in their framework that of the three types of beliefs, outcome beliefs most closely impact behavioral outcomes, though this is conditional on other factors (i.e., internal appraisals, degree of motivation to act) that help oil the wheels. Our appraisal process, basically enriching outcome beliefs with belief appraisals (i.e., is this good and worth acting on, or bad and best avoided), is driven by our own value system. On this point, we see independent support of the same critical internal collection of processes. Transforming internal mental activity into action has been examined in value-based decision-making models (Rangel et al., 2008) found in moral decision-making (Osman & Wiegmann, 2017), effort-based decision-making (Ludwiczak et al., 2020; Rangel et al., 2008), and group decision-making (Adams et al., 2022; Ludwiczak et al., 2022). In short, value-based decision-making models and Granados Samayoa and Albarracín's belief-to-behavior framework describe how mental activity is transformed into measurable behavioral outcomes based on a stage-based process, where value assignment is instrumental (See Figure 1 Panel D). The critical point made here is that increasing measures across the different stages from beliefs-to-behavior helps form a more accurate causal analysis of the impact of beliefs on behavior. As well as this, when conducting a causal analysis of beliefs-to-behaviors experimenters should separate out the potential influence of value-based appraisals that they make from those of participants (e.g., Pasquetto et al., 2024; Williams, 2023; Yee, 2023, 2025).

Insert Figure 1 about here

Granados Samayoa and Albarracín's target article considers the influential role of content that people encounter that either interacts with their prior beliefs (e.g., existential beliefs, descriptive beliefs), or even helps to form new beliefs (see Figure 1 Panel C). The content that is of particular interest, and that has exercised many researchers, is misinformation and disinformation (for detailed discussion see Adams et al., 2023). The distillation of this work is a simple causal model of cognition (See Figure 1 Panel A). Researchers use the simple causal model to examine the assimilation of misinformation/disinformation into people's internal belief systems to then track how these beliefs in turn drive aberrant behaviors (e.g., antisocial acts towards individuals, groups, or public property). Several researchers have been cautioning the research community about this simple model, along with the erroneous assumptions made about where, how, and by what magnitude we succumb to misinformation and disinformation (Acerbi et al., 2022; Altay et al., 2023; Altay et al., 202; Krause, et al., 2022; Mercier, 2021; Scheufele, & Krause, 2019). For instance, assumptions that inform the hypotheses being tested to evidence causality (see Figure 1 Panel A) can lead to overinterpretation of evidence of causality when it is still largely correlational (Adams et al., 2023; Osman et al., 2022; Osman, 2024).

Cognitive psychology is replete with examples of motivated reasoning (e.g. belief polarization, biased assimilation, bias blind spot, conformist bias, false consensus effect, ideological bias, illusion of information adequacy, myside bias, naïve realism, objectivity illusion, opinion polarization, prestige bias, self-serving bias, selective scrutiny) (e.g., Banisch & Shamon, 2025; Greitemeyer et al., 2009; Hameleers & Brosius, 2022; Kobayashi, 2016; Oeberst & Imhoff, 2023; Pronin et al., 2002; Robinson et al., 2023; Ross, 2012, 2018; Ross & Ward, 1996; van Doorn, 2024; Williams, 2023). While it isn't a surprise, it is worth reminding the community that researchers are not immune from these forms of motivated thinking (e.g., Clark & Tetlock, 2023; Gulati & Palladini, 2023; Lamont, 2009; O'Connor & Weatherall, 2018; Podsakoff et al., 2012; Stanovich, 2023; Trafimow et al., 2024; Traunmüller, 2023; Watkins & Harvey, 2020).

As mentioned, motivated assumptions can have significant implications for the empirical investigation of the relationship (or lack thereof) between beliefs, supposedly influenced by misinformation and disinformation, and their impact on behavior. That is, from the outside, especially when researchers start to look at belief-to-behavior, their own appraisal process may be at odds with those participating in the research (Pasquetto et al, 2024, Yee, 2023, 2025). In turn this will influence the researcher's causal analysis (content [misinformation, disinformation] -> belief -> behavior) depending on their own assumptions and value assignments (see Figure 1 Panel B), succinctly referred to as the transmission heuristic (Osman, 2024). The transmission heuristic equates “good” (accurate/true) information to “good mental states” to “good” behavior, and “bad” (inaccurate/false) information to “bad mental states” to “bad” behavior (e.g., De Ridder, 2024; Levy, 2022).

Let’s imagine we apply the transmission heuristic in a consumer setting. In 2018 a multinational consumer goods company advertised a new plant-based range of products claimed to be sustainable, 50% of them are plant-based making them biodegradable, and manufactured from 100% renewable energy (Reilly, 2020). Being plant-based is one of several labels (e.g., eco-design, environmentally friendly, green, organic, sustainable) on products that we as consumers use as a reliable indicator of products minimizing harms to the environment (Iannuzzi, 2024). In 2021 (Almanzar, 2021) a class action against the multinational company alleges that the content claims are deliberately false and designed to tap into a growing market of eco-ethical consumers (Credence, 2025). If the outcome in court favors the defendant, then the content claims are true, and lead to positive beliefs that in turn lead to positive behaviors, aligned with the transmission heuristic. If the outcome of the class action is in favor of the plaintiff, from a researcher’s standpoint this context shows how disinformation duped people into holding good beliefs (e.g. ethical consumption) and good behavior (e.g. pro-environmental purchases). This clearly isn’t in line with a straightforward application of the transmission heuristic.

In fact, in the aforementioned example, as researchers, we wouldn’t be able to determine the veracity of the advertising claims of the product if we were studying their influence right now. Our value assignment needs to be carefully navigated when empirically investigating the relationships between content-to-belief-to-behavior. We see many examples of this kind where researchers too quickly classify claims as misinformation or disinformation to show links to specific behaviors when evidence for the veracity of the claims is evolving (Adams et al., 2023; Osman et al., 2022; Osman, 2024; Yee, 2023). As a corollary, when it comes to the construction of policies to address the suspected influence of misinformation and disinformation on aberrant behaviors, policy decision-makers are just as exposed to the same motivated reasoning as researchers (Trafimow et al., 2024; Trafimow & Osman, 2022; Yee, 2025).

To avoid this problem, three things are needed. The first is that researchers should suspend their own motivated reasoning or at least be more reflexive in acknowledging its presence (e.g. Pasquetto et al., 2024; Yee, 2023, 2025). Second, Granados Samayoa and Albarracín's framework offers a better alternative to over simplistic models (see Figure 1 Panel A and B). In fact, Granados Samayoa and Albarracín's framework is more psychologically informative (see Figure 1 Panel D) not least because it is supported by decade's worth of empirical findings (e.g., value-based decision-making research in cognitive psychology, neuroscience, comparative cognition, social psychology). Third, when conducting experiments, researchers need to measure participants' own belief appraisals and their own value system to avoid erroneously inferring there to be a weak or no causal chain when there is one, or a strong causal chain when there isn't one.

Going back to the consumer example, as researchers we can do several practical things by applying Granados Samayoa and Albarracín's framework. We can examine whether consumers are aware of greenwashing (i.e., the misrepresentation of sustainability practices or activities to promote a false image of responsibility) to examine their level of skepticism of content claims (e.g. de Freitas Netto et al., 2020). Rather than have the researcher determine the veracity of the content claims that are of interest (e.g., a product label, advert), the appraisal process should be based on the consumer's own evaluation, which avoids unnecessary complications from possible experimenter motivated reasoning. We can look at how much consumer choices are influenced by their own existential and descriptive beliefs to gauge overall propensity towards sustainable products (e.g., Akhtar et al., 2021; Iannuzzi, 2024). We can then look at what their outcome beliefs are and associated appraisal beliefs to determine the strength of those beliefs as a means of motivating their actions (e.g., action selection – which products consumers intend to buy, action execution – which products they buy) (e.g. Kamalanon et al., 2022). The valuation here reflects internal incentive structures that people will use to make actual purchasing decisions (e.g. Akhtar et al., 2021). These measures can be used in conjunction with their consumer choices in various settings (e.g., what goes into their shopping basket in bricks-and-mortal establishments, marketplaces online) to estimate content-to-beliefs-to-behavior in a specific consumer context (e.g., Bharani et al., 2024; Rausch et al., 2021).

In short, expanding the measurements collected in studies that are informed by the causal structure that Granados Samayoa and Albarracín's framework proposes will mean extra effort on the part of researchers. It is worth doing this because this level of diligence means researchers can more accurately estimate the level of impact of misinformation and disinformation on behaviors using multiple measures of internal belief mechanisms.

Aggregated beliefs-to-behaviors

Misinformation and disinformation are, by some organizations that currently rank them, the foremost global existential threat (for discussion see Osman, 2024). To understand why, and to pinpoint ways to address this has meant shifting from a micro to a macro unit of analysis of content[misinformation/disinformation]-to-beliefs-to-behaviors (e.g., whole populations, sub-sections of society, interactions on social media platforms).

We have seen concerted interest in using social media data to detect the presence of misinformation and disinformation, though of course this doesn't mean they aren't present in other contexts (Altay et al., 2023; Altay et al, 2024; Osman et al., 2023) which will be revisited later. Textual analysis of social media interactions is then used in modelling macro level behaviors such as civil unrest (e.g., violent protest, rioting) and acts of terrorism that have societal impact (e.g., Chinta et al., 2021; Grill, 2021; Hunt et al., 2022; Iyda & Geetha, 2023; Jamil et al., 2022; Lewis & Coaffee, 2024; Piazza, 2022; Singh & Jain, 2024; Rajendran et al., 2022; Tuke et al., 2020; Zadeh & Cicekli, 2023, Yee, 2025). Before going on to show why this has implications and applications for Granados Samayoa and Albarracín's framework, several aspects of this macro level unit of analysis needs to be spelled out. To start with, we need to differentiate explanatory modelling – using statistical inference to validate causal explanations of underlying relationships from models that are predictive – using data mining to inform machine learning to predict future events (e.g., Grossman et al., 20254; Hehman & Neel, 2024; Lassen et al., 2016). The former gives the rationale for the assumptions that go into the latter. We will revisit this point later in this section, but in short, the message is that prediction is not a substitute for explanation, and explanation should always precede prediction (Shmueli, 2010).

Predictive modelling work on civil unrest has outlined several limitations of models that use social media behavior to estimate future offline macro level behaviors of social unrest. There is considerable difficulty in setting appropriate baselines for social media datasets to then track new agitations that could predict spill over to offline violent behavior (Singh & Jain, 2024). Relatedly, it is difficult to systematize the time intervals across various predictive models on which social media analysis is carried out to then detect precursors online that predict offline behaviors (Jamil et al., 2022; Piazza, 2022; Rajendran et al., 2022). Challenges also arise from attempting to predict the dynamics of events, that is, predicting events as they mutate from peaceful protests to clusters of violent ones (Jamil et al., 2022; Singh & Jain, 2024). Because classifiers typically depend on key terms, they often struggle to classify dependencies within sentences which would provide key contextual information to make more informed predictions (Jamil et al., 2022; Philips et al., 2017; Rajendran et al., 2022; Tuke et al., 2020). Societies are often made up of multi-lingual communities that vary in their engagement on social media platforms, and that also vary in their reliance on alternative media channels (e.g. Danning, 2018; Stremlau et al., 2024). Even combining data sets from different countries of the same language can also lead to considerable inaccuracies in text-based veracity detection of news (Horne et al., 2020). In turn, predictive modelling is limited by relying on text analysis of a single language data set (e.g. from a popular social media platform) to accurately predict offline violent behaviors (Chinta et al., 2021; Philips et al., 2017).

Overall, many in the research field highlight that predictive models suffer from a lack of theoretical underpinning in their models. Where theory could offer protections against bias, in its absence modelers' own value-laden assumptions can lead to inaccuracies in prediction of content[misinformation]-to-beliefs-to-behavior (Yee, 2023, 2025). The models are highly domain-specific which limit their generalizability to other related event types (Philips et al., 2017; Singh & Jain, 2024; Yee, 2023, 2025). In addition, limits to their generalizability also come from focusing primarily on westernized countries, when there are many examples of social unrest occurring elsewhere (Stremlau et al., 2024).

How does this relate to Granados Samayoa and Albarracín's belief-to-behaviors framework? As a case study in diagnosing the problems in predicting aggregate belief-to-behavior, the long scene setting here is designed to lay out the various substantive practical limitations in predictive models. In short, we cannot advance as a research community if we don't know the full range of problems we face.

Causal analysis of content-beliefs-to-behavior

The problems in models predicting low-probability-high-impact events (e.g. violent protest, riots, and terrorist acts) through polluted social media content essentially come down to the following. First, they do not utilize explanatory models (e.g., Philips et al., 2017; Yee, 2023, 2025). Essentially, they, like many other predictive models (Grossman et al., 20254; Hehman & Neel, 2024), lack an underlying causal model of content-to-beliefs-to-behavior. Second, they do not have a principled approach to utilizing situational analysis (e.g., Grill, 2021; Lewis & Coaffee, 2024). Essentially, absent from predictive modelling is a systematic coordination of causal drivers. Temporal, spatial (e.g., regional, national, or global events), economic, political, and cultural factors need to precede quantitative analysis of misinformation or disinformation content used to predict adverse social events (e.g., Yee, 2023, 2025). Third, predictive models do not have a principled approach to classifying the events, including the full range of behaviors they are interested in predicting (e.g., Grossman et al., 2024; Jamil et al., 2022; Philips et al., 2017; Yee, 2023, 2025). Designers of predictive models have consistently called for high quality event-based datasets where common patterns in event types can be mined (Jamil et al., 2022; Philips et al., 2017; Rajendran et al., 2022; Singh & Jain, 2024).

The first section helped to show how Granados Samayoa and Albarracín's framework can support empirical investigations by increasing the range of empirical measures. Inputting data from them into a causal analysis (e.g. a Causal-Bayesian network) can improve estimates of the strength of content-to-belief-to-behaviors relationships. Granados Samayoa and Albarracín's framework is psychologically informed, based on independent empirical support from other literatures, that could support the design of explanatory models. Explanatory models can then inform predictive models (Grossman et al., 2024; Shmueli, 2010). The remainder of this section focuses on discussing the added value of situation analysis supported by Granados Samayoa and Albarracín's framework, and the necessary role of the Principle of Compatibility (Ajzen, 1988, 2020).

Viral social media content that experts judge as deleterious (e.g., disinformation, fake news) has been associated with violence offline (e.g. Njuguna et al., 2020; Scrivens, 2024), but its unpredictability (e.g. Mukherjee et al., 2022) means this association doesn't always occur (e.g., Fokou, et al., 2024; Stremlau et al., 2024). Not all trigger events (e.g., assassination of a political figure, election result, natural disaster, terrorist attack) agitate social media interactions enough to mobilize reactive social unrest (e.g., Allchorn, 2023; Innes et al., 2021; Rød et al., 2025; Wu & Gerber, 2017). Spill over into social unrest can come from outrage to news from legitimate news organizations (e.g., Henn & Posegga, 2023; Seutter et al., 2024; Stremlau et al., 2024; Tarafdar et al., 2021; Venneti, 2018) that disseminate misinformation and disinformation (e.g. Adams et al., 2023; Hunter & Biglaiser, 2022). Put starkly, where researchers may expect to find clear causal associations in line with the transmission heuristic - this time at a macro level, the evidence points in inconvenient directions. As explained before, accepting that motivated reasoning creates problems in the way evidence is sought, there are ways to avoid this by expanding the scope of data collected and how it is contextualized, which situational analysis enables.

For many researchers, forecasting destabilizing low-probability-high-impact events from exposure to online social media content demands detailed situation analysis (e.g., Danning, 2018; Grill, 2021; Hunter & Biglaiser, 2022; Jost et al., 2018; Lewis & Coaffee, 2024; Murphy, Sharpe, & Huang, 2024; Philips et al., 2017; Pinckney & RezaeeDaryakenari, 2022; Stremlau et al., 2024; von der Burg & Krasmann, 2024). Granados Samayoa and Albarracín's framework can be applied to help with this. Situational analysis means looking at general economic, political, cultural and social factors that inform, say, political discussions online, and then tracking changes in them over a specific time interval. These can be treated as explanatory factors to establish a baseline from which fluctuations in high level beliefs (existential beliefs, descriptive beliefs) expressed as political discourse online can be detected. When there are key events that are expected to increase contentious discussions with violent spillover potential local situation factors around that time can be recorded. In combination this level of detailed analysis can be used to detect specific outcome beliefs and appraisals of them as predictors of social unrest (e.g., Jost et al., 2018; Rød et al., 2025). This is not an impossible task. Data scientists and social scientists are recognizing each other's contributions in designing explanatory models that integrate situation analysis for latter predictive modelling (e.g., Grill, 2021; Jost et al., 2018; Philips et al., 2017; Shyalika et al., 2024, Yee, 2023).

Carefully cataloguing the behaviors themselves is the complement to understanding the background causal drivers that influence beliefs through social media interactions to predict behaviors. This also speaks to the point that many modelers have raised about needing high quality event databases to work with. For instance, the Global Database of Events, Language and Tone (GDELT) is shown to have significant drawbacks that limit the accuracy of predictive modelling (e.g., Halkia et al., 2020; Hoffman et al., 2022). Granados Samayoa and Albarracín anticipate this in their target article by referring to Ajzen's (1988, 2020) Principle of Compatibility; others have also proposed it's use in studies of misinformation and disinformation (Adams et al., 2023).

Applying the Principle of Compatibility means explicitly defining the behaviors (i.e. specific actions), the target the action is directed towards, the context in which this occurs, and the time of occurrence. This is used to assess the compatibility between beliefs and behavior (Ajzen, 1988, 1993, 2020); the greater the specificity of both, the better able researchers are in finding reliable associations (e.g. Davis et al., 2018). In fact, the same details that the Principle of Compatibility require can inform where improvements are needed in event databases (e.g., Armed Conflict Location and Event Data Project (ACLED), GDELT, Global Terrorism Database (GTD), Conflict Data Program Georeferenced Event Dataset (UCDP GED)). Crucially, the application of the Principle of Compatibility needs to be done prior to any predictive modelling of content-to-beliefs-to-behavior. Applying the principle after the fact starts to look like cherry picking behaviors that ought to have been caused by deleterious content shared on social media.

Taking a slightly different tack, the aim here is to show how Granados Samayoa and Albarracín framework and Ajzen's (1988, 2020) Principle of Compatibility can be applied to video games. Why might this be a useful context to examine? Researchers are beginning to focus on gamers exposure to deleterious content in video games or video-game adjacent applications to examine the impact on beliefs-to-behavior (e.g., Achmirowicz & Martin, 2023; Amarasingam & Kelley, 2024; Chew, 2022; Davey, 2024; Freed, 2024; Lakhani & Wiedlitzka, 2023; Schlegel et al., 2025; Stokes & Williams, 2018). The concern is that these online environments are a mechanism for spreading and reinforcing misinformation and disinformation. Moreover, because of the age of many gamers, and the length of time they typically spend gaming, videogames are an opportune way to shape impressionable cohorts into adopting beliefs that can lead to antisocial behaviors (e.g., Oksanen et al., 2024; Wells et al., 2024).

The type of immersive game played is a way to identify potential spillovers into real world social unrest, but also the type of discourse that takes place amongst communities of gamers. These contexts are an alternative way of examining the compatibility between beliefs and behaviors within virtual environments, but also the compatibility between them online and offline. Granados Samayoa and Albarracín framework can be applied to immersive gaming environments (Figure 1, Panel D). The content that gamers are exposed can be examined from the perspective of gamers prior beliefs (existential beliefs, descriptive beliefs) and the local formation of outcome beliefs while engaging with the game. When outcome beliefs are identified, gamers appraisals of them can be tracked in real time to show where the beliefs directly impact behaviors in the game. In conjunction, Ajzen's (1988, 1993, 2020) Principle of Compatibility can also be applied to systematically decompose behaviors by actions, target, and context in the game, and this can be used to predict when and where they might be prevalent offline. Having done this, broader level situational factors can be used to inform explanatory models of communities of gamers to then start to make predictions about how deleterious content in games could influence beliefs that then spill over to group level behaviors (e.g. violent protests) offline.

Conclusion

This commentary shows the value of Granados Samayoa and Albarracín's framework in applied research by using misinformation and disinformation as examples of influencers on beliefs that then impact behaviors. The overall take home from this commentary is that Granados Samayoa and Albarracín's target article helps systematizing the causal analysis of beliefs-to-behavior. The framework informs which measures can better inform the causal analysis of beliefs-to-behaviors. To complement this, there are ways of systematically analyzing behaviors that beliefs are expected to impact through the application of the Principle of Compatibility. Situational analysis helps to map out the context in which beliefs are formed or reinforced, and where behaviors are predicted to have profound impact on the individual and society.

References

Acerbi, A., Altay, S., & Mercier, H. (2022). Research note: Fighting misinformation or fighting for information? Harvard Kennedy School (HKS) Misinformation Review. https://doi. org/10.37016/mr-2020-87

Achmirowicz, J., & Martin, M. (2023). *Cognitive Warfare: Report* (Doctoral dissertation, Ecole de Guerre). https://hal.science/hal-04224034/

Adams, Z., Ludwiczak, A., Sharma, D., & Osman, M. (2022). Verbal interaction in a social dilemma. *Rationality and Society*, *34*(3), 334-367. https://doi.org/10.1177/1043463122109455

Adams, Z., Osman, M., Bechlivanidis, C., & Meder, B. (2023). (Why) is misinformation a problem?. *Perspectives on Psychological Science*, *18*(6), 1436-1463. https://journals.sagepub.com/doi/pdf/10.1177/17456916221141344

Ajzen, I. (1988). Attitudes, personality, and behavior. Homewood, IL: Dorsey Press.

Ajzen, I. (1993). Attitude theory and the attitude-behavior relation. In New Directions in Attitude Measurement, ed. D. Krebs and P. Schmidt, 41–57. Walter de Gruyter: Berlin, Germany

Ajzen, I. (2020). The theory of planned behavior: Frequently asked questions. *Human behavior and emerging technologies*, *2*(4), 314-324. https://doi.org/10.1002/hbe2.195

Akhtar, R., Sultana, S., Masud, M. M., Jafrin, N., & Al-Mamun, A. (2021). Consumers' environmental ethics, willingness, and green consumerism between lower and higher income groups. *Resources, Conservation and Recycling*, *168*, 105274. https://doi.org/10.1016/j.resconrec.2020.105274

Altay, S., Berriche, M., & Acerbi, A. (2023). Misinformation on misinformation: Conceptual and methodological challenges. *Social media+ society*, *9*(1), 20563051221150412. https://journals.sagepub.com/doi/pdf/10.1177/20563051221150412

Allchorn, W. (2023). Barriers to Violence Activism on the UK Far Right: The Case of the (Democratic) Football Lads Alliance. In *Routledge Handbook of Non-Violent Extremism* (pp. 235-245). Routledge.

Altay, S., Nielsen, R. K., & Fletcher, R. (2024). News can help! The impact of news media and digital platforms on awareness of and belief in misinformation. *The International Journal of Press/Politics*, *29*(2), 459-484. https://journals.sagepub.com/doi/pdf/10.1177/19401612221148981

Almanzar v. The Procter & Gamble Company - 1:21-cv-09196

Amarasingam, A., & Kelley, D. (2024). Hate and extremism on gaming platforms: Insights from surveys with the gaming community. In L. Schlegel & R. Kowert (Eds.), Gaming and Extremism: The Radicalization of Digital Playgrounds (pp. 110-129). https://doi.org/10.4324/9781003388371-7

Banisch, S., & Shamon, H. (2025). Biased processing and opinion polarization: experimental refinement of argument communication theory in the context of the energy debate. *Sociological Methods & Research*, *54*(1), 187-236. https://doi.org/10.1177/00491241231186658

Bharani, S., Roy, S., & Tawde, S. (2024). Green products wrapped and delivered: consumer's preference for sustainable e-commerce practices. *The International Review of Retail, Distribution and Consumer Research*, *34*(3), 331-360. https://doi.org/10.1080/09593969.2023.2263822

Chew, M. M. (2022). The significance and complexities of anti-corporate gamer activism: Struggles against the exploitation and control of game-worlds in 2000s China. *Games and Culture*, *17*(4), 487-508. https://doi.org/10.1177/15554120211042259

Chinta, A., Zhang, J., DeLucia, A., Dredze, M., & Buczak, A. L. (2021). Study of manifestation of civil unrest on Twitter. In *Proceedings of the Seventh Workshop on Noisy User-generated Text (W-NUT 2021)* (pp. 396-409). https://aclanthology.org/2021.wnut-1.44/

Clark, C. J., & Tetlock, P. E. (2023). Adversarial collaboration: The next science reform. In *Ideological and political bias in psychology: Nature, scope, and solutions* (pp. 905-927). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-29148-7_32

Danning, G. (2018). Did Radio RTLM Really Contribute Meaningfully to the Rwandan Genocide?: Using Qualitative Information to Improve Causal Inference from Measures of Media Availability. *Civil Wars*, *20*(4), 529-554. https://doi.org/10.1080/13698249.2018.1525677

Davey, J. (2024). Extremism on gaming (-adjacent) platforms. In L. Schlegel & R. Kowert (Eds.), Gaming and Extremism: The Radicalization of Digital Playgrounds (pp. 95-109). https://doi.org/10.4324/9781003388371-6

Davis, M. H., Schoenfeld, M. B., & Flores, E. J. (2018). Predicting conflict acts using behavior and style measures. *International Journal of Conflict Management*, *29*(1), 70-90. https://doi.org/10.1108/IJCMA-06-2016-0046

de Freitas Netto, S. V., Sobral, M. F. F., Ribeiro, A. R. B., & Soares, G. R. D. L. (2020). Concepts and forms of greenwashing: A systematic review. *Environmental Sciences*

Europe, *32*, 1-12. https://link.springer.com/content/pdf/10.1186/s12302-020-0300-3.pdf

De Ridder, J. (2024). What's so bad about misinformation?. *Inquiry*, *67*(9), 2956-2978. https://doi.org/10.1080/0020174X.2021.2002187

Fokou, G., Yamo, A., Kone, S., Koffi, A. J. D. A., & Davids, Y. D. (2024). Xenophobic violence in South Africa, online disinformation and offline consequences. *African Identities*, *22*(4), 943-962. https://doi.org/10.1080/14725843.2022.2157245

Freed, J. J. (2024). *The Interpenetration of Political Culture and Game Culture: Why We Need to Pay Attention to In-Game Communication* (Doctoral dissertation, The University of Utah). https://www.proquest.com/openview/a04c87e0579136814e6fa7e963371a84/1?pq-origsite=gscholar&cbl=18750&diss=y

Greitemeyer, T., Fischer, P., Frey, D., & Schulz-Hardt, S. (2009). Biased assimilation: The role of source position. *European Journal of Social Psychology*, *39*(1), 22-39. https://doi.org/10.1002/ejsp.497

Grill, G. (2021). Future protest made risky: Examining social media based civil unrest prediction research and products. *Computer Supported Cooperative Work (CSCW)*, *30*(5), 811-839. https://doi.org/10.1007/s10606-021-09409-0

Grossmann, I., Varnum, M. E., Hutcherson, C. A., & Mandel, D. R. (2024). When expert predictions fail. *Trends in cognitive sciences*, *28*(2), 113-123. 3. https://doi.org/10.1016/j.tics.2023.10.005

Gulati, K., & Palladini, L. (2023). How ‘Free’ is Free Speech in Academia? Effects on Researchers and their Research. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4440242

Halkia, M., Ferri, S., Papazoglou, M., Van Damme, M. S., & Thomakos, D. (2020). Conflict event modelling: research experiment and event data limitations. In *Proceedings of the Workshop on Automated Extraction of Socio-Political Events from News 2020* (pp. 42-48). https://aclanthology.org/2020.aespen-1.8.pdf

Hameleers, M., & Brosius, A. (2022). You are wrong because I am right! The perceived causes and ideological biases of misinformation beliefs. *International Journal of Public Opinion Research*, *34*(1), edab028. https://doi.org/10.1093/ijpor/edab028

Hehman, E., & Neel, R. (2024). Prejudice model 1.0: A predictive model of prejudice. *Psychological Review, 131*(5), 1235–1265. https://doi.org/10.1037/rev0000470

Henn, T., & Posegga, O. (2023). Attention-grabbing news coverage: Violent images of the Black Lives Matter movement and how they attract user attention on Reddit. *PLoS one*, *18*(8), e0288962. https://doi.org/10.1371/journal.pone.0288962

Hoffmann, M., Santos, F. G., Neumayer, C., & Mercea, D. (2022). Lifting the veil on the use of big data news repositories: A documentation and critical discussion of a protest event analysis. *Communication Methods and Measures*, *16*(4), 283-302. https://doi.org/10.1080/19312458.2022.2128099

Horne, B. D., Gruppi, M., & Adali, S. (2020). Do all good actors look the same? Exploring news veracity detection across the U.S. and the U.K. Association for the Advancement of Artificial Intelligence, 1-4. https://doi.org/10.48550/arXiv.2006.01211

Hunt, K., Agarwal, P., & Zhuang, J. (2022). Monitoring misinformation on Twitter during crisis events: a machine learning approach. *Risk analysis*, *42*(8), 1728-1748. https://doi.org/10.1111/risa.13634

Hunter, L. Y., & Biglaiser, G. (2022). The effects of social media, elites, and political polarization on civil conflict. *Studies in Conflict & Terrorism*, 1-28. https://doi.org/10.1080/1057610X.2022.2163461

Iannuzzi, A. (2024). *Greener products: The making and marketing of sustainable brands*. Boca Raton. CRC Press.

Innes, M., Dobreva, D., & Innes, H. (2021). Disinformation and digital influencing after terrorism: Spoofing, truthing and social proofing. *Contemporary Social Science*, *16*(2), 241-255. https://doi.org/10.1080/21582041.2019.1569714

Iyda, J. J., & Geetha, P. (2023). An improved deep belief neural network based civil unrest event forecasting in twitter. *Applied Intelligence*, *53*(5), 5714-5731. https://doi.org/10.1007/s10489-022-03746-3

Jamil, M.L., Pais, S. & Cordeiro, J. (2022). Detection of dangerous events on social media: a critical review. *Social Network Analysis and Mining,* 12, 154. https://doi.org/10.1007/s13278-022-00980-y

Jost, J. T., Barberá, P., Bonneau, R., Langer, M., Metzger, M., Nagler, J., ... & Tucker, J. A. (2018). How social media facilitates political protest: Information, motivation, and social networks. *Political psychology*, *39*, 85-118. https://doi.org/10.1111/pops.12478

Kamalanon, P., Chen, J. S., & Le, T. T. Y. (2022). “Why do we buy green products?” An extended theory of the planned behavior model for green product purchase behavior. *Sustainability*, *14*(2), 689. https://doi.org/10.3390/su14020689

Kobayashi, K. (2016). Relational Processing of Conflicting Arguments: Effects on Biased Assimilation. *Comprehensive Psychology* 5: 2165222816657801. https://doi.org/10.1177/2165222816657801

Krause, N. M., Freiling, I., & Scheufele, D. A. (2022). The “infodemic” infodemic: Toward a more nuanced understanding of truth-claims and the need for (not) combatting misinformation. The ANNALS of the American Academy of Political and Social Science, 700(1), 112–123. https://doi.org/10.1177/00027162221086263

Lakhani, S., & Wiedlitzka, S. (2023). “Press F to Pay Respects”: An empirical exploration of the mechanics of gamification in relation to the Christchurch attack. Terrorism and Political Violence, 35(7), 1586-1603. https://doi.org/10.1080/09546553.2022.2064746

Lamont, M. (2009). *How professors think: Inside the curious world of academic judgment*. Harvard University Press.

Lassen, N. B., la Cour, L., & Vatrapu, R. (2016). Predictive analytics with social media data. In Sloan, L., & Quan-Haase, A. The SAGE Handbook of social media research methods (pp. 328-341). 55 City Road, London: SAGE Publications Ltd. https://doi.org/10.4135/9781473983847

Levy, N. (2022). Bad beliefs: Why they happen to good people. Oxford University Press.

Lewis, T., & Coaffee, J. (2024). Atmosphere, imminence and the Manchester Arena Inquiry: on the affective modalities of becoming situationally aware to urban terrorism. *Critical Studies on Security*, 1-20. https://doi.org/10.1080/21624887.2024.2385185

Ludwiczak, A., Osman, M., & Jahanshahi, M. (2020). Redefining the relationship between effort and reward: Choice-execution model of effort-based decisions. *Behavioural Brain Research*, *383*, 112474. https://doi.org/10.1016/j.bbr.2020.112474

Ludwiczak, A., Adams, Z., & Osman, M. (2022). Actions do not always speak louder than words: The role of social incentives when choosing and executing cooperative effort. *Experimental Psychology, 69*(3), 155–162. https://doi.org/10.1027/1618-3169/a000555

Mercier, H. (2021). How good are we at evaluating communicated information?. *Royal Institute of Philosophy Supplements*, *89*, 257-272. https://doi.org/10.1017/S1358246121000096

Murphy, M., Sharpe, E., & Huang, K. (2024). The promise of machine learning in violent conflict forecasting. *Data & Policy*, *6*, e35. https://doi.org/10.1017/dap.2024.27

Mukherjee, P., Dutta, S., & De Bruyn, A. (2022). Did clickbait crack the code on virality?. *Journal of the Academy of Marketing Science*, *50*(3), 482-502. https://doi.org/10.1007/s11747-021-00830-x

Njuguna, C., Gikandi, J., Kathuri-Ogola, L., & Kabaria-Muriithi, J. (2020). Social Media as a Battleground for Electoral Violence among Netizens in Africa: The Case of Kenyan Youth. *East African Journal of Arts and Social Sciences*, *2*(1), 104-115. https://doi.org/10.37284/eajass.2.1.209

Oeberst, A., & Imhoff, R. (2023). Toward parsimony in bias research: A proposed common framework of belief-consistent information processing for a set of biases. *Perspectives on Psychological Science*, *18*(6), 1464-1487. https://journals.sagepub.com/doi/full/10.1177/17456916221148147

O'Connor, C., & Weatherall, J. O. (2018). Scientific polarization. *European Journal for Philosophy of Science*, *8*, 855-875. https://doi.org/10.1007/s13194-018-0213-9

Oksanen, A., Celuch, M., Oksa, R., & Savolainen, I. (2024). Online communities come with real-world consequences for individuals and societies. *Communications Psychology*, *2*(1), 71. https://doi.org/10.1038/s44271-024-00112-6

Osman, M. (2024). Public Evaluations of Misinformation and Motives for Sharing It. *Journalism and Media*, *5*(2), 766-786. https://www.mdpi.com/2673-5172/5/2/50/pdf

Osman, M., Adams, Z., Meder, B., Bechlivanidis, C., Verduga, O., & Strong, C. (2022). People's understanding of the concept of misinformation. *Journal of Risk Research*, *25*(10), 1239-1258. https://doi.org/10.1080/13669877.2022.2049623

Osman, M., Meder, B., Bechlivanidis, C., & Strong, C. (2023). To Share or Not to Share, Are There Differences Offline than Online? No. 8. Centre for Science and Policy, University of Cambridge Working Paper Series. Available online: https://www.csap.cam.ac.uk/media/uploads/files/1/offline-vs-online-sharing.pdf (accessed on 1 May 2024).

Osman, M., & Wiegmann, A. (2017). Explaining moral behavior: A minimal moral model. Experimental Psychology, 64, 68-81. https://doi.org/10.1027/1618-3169/a000336

Piazza, J. A. (2022). Fake news: The effects of social media disinformation on domestic terrorism. *Dynamics of Asymmetric Conflict*, *15*(1), 55-77. https://doi.org/10.1080/17467586.2021.1895263

Pasquetto, I. V., Lim, G., & Bradshaw, S. (2024). Misinformed about misinformation: On the polarizing discourse on misinformation and its consequences for the field. Harvard Kennedy School (HKS) Misinformation Review, 5(5). https://doi.org/10.37016/mr-2020-159

Pasquetto, I. V., Swire-Thompson, B., Amazeen, M. A., Benevenuto, F., Brashier, N. M., Bond, R. M., ... & Yang, K. C. (2020). Tackling misinformation: What researchers could do with social media data. *The Harvard Kennedy School Misinformation Review*. https://doi.org/10.37016/mr-2020-49

Phillips, L., Dowling, C., Shaffer, K., Hodas, N., & Volkova, S. (2017). Using social media to predict the future: a systematic literature review. arXiv preprint arXiv:1706.06134.

Pinckney, J., & RezaeeDaryakenari, B. (2022). When the levee breaks: A forecasting model of violent and nonviolent dissent. *International Interactions*, *48*(5), 997-1026. https://doi.org/10.1080/03050629.2022.2090933

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual review of psychology*, *63*(1), 539-569. https://doi.org/10.1146/annurev-psych-120710-100452

Pronin, E., Lin, D. Y., & Ross, L. (2002). The bias blind spot: Perceptions of bias in self versus others. Personality and Social Psychology Bulletin, 28(3), 369–381. https://doi .org/10.1177/0146167202286008

Rajendran, A., Sahithi, V. S., Gupta, C., Yadav, M., Ahirrao, S., Kotecha, K., ... & Alhammad, S. M. (2022). Detecting extremism on twitter during US capitol riot using deep learning techniques. *IEEE Access*, *10*, 133052-133077. https://ieeexplore.ieee.org/document/9978618

Rangel, A., Camerer, C. & Montague, P. A. (2008). Framework for studying the neurobiology of value-based decision making. *Nature Reviews: Neuroscience,* 9, 545–556. https://doi.org/10.1038/nrn2357

Rausch, T. M., Baier, D., & Wening, S. (2021). Does sustainability really matter to consumers? Assessing the importance of online shop and apparel product attributes. *Journal of Retailing and Consumer Services*, *63*, 102681. https://doi.org/10.1016/j.jretconser.2021.102681

Reilly, A. (2020). Social Media, Sustainability and Organizations. *Encyclopedia of the World's Biomes*, 5, 397-406. https://doi.org/10.1016/B978-0-12-409548-9.11972-4

Robinson, C. D., Gehlbach, H., & Fletcher, A. (2023). Naiver Realism II: A Key Refinement to the Theory of Naïve Realism. https://osf.io/cx5h2/resources

Rød, E. G., Hegre, H., & Leis, M. (2025). Predicting armed conflict using protest data. Journal of Peace Research, 62(1), 3-20. https://doi.org/10.1177/00223433231186452

Ross, L. (2012). Reflections on biased assimilation and belief polarization. *Critical Review*, *24*(2), 233-245. https://doi.org/10.1080/08913811.2012.711025

Ross, L. (2018). From the fundamental attribution error to the truly fundamental attribution error and beyond: My research journey. *Perspectives on Psychological Science*, *13*(6), 750-769. https://doi.org/10.1177/1745691618769855

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel, & T. Brown (Eds.), Values and knowledge (pp. 103–135). Psychology Press.

Scheufele, D. A., & Krause, N. M. (2019). Science audiences, misinformation, and fake news. *Proceedings of the National Academy of Sciences*, *116*(16), 7662-7669. https://doi.org/10.1073/pnas.1805871115

Schlegel, L., Wiegold, L., Winkler, C., & Junk, J. (2025). Exploring the digital extremist ecosystem: A preliminary analysis of hateful posts on Mod DB. *Frontiers in Psychology*, *15*, 1502098. https://doi.org/10.3389/fpsyg.2024.1502098

Scrivens, R. (2024). Examining Online Behaviors of Violent and Non-Violent Right-Wing Extremists During Peak Posting Days. *Studies in Conflict & Terrorism*, 1-27. https://doi.org/10.1080/1057610X.2024.2347860

Seutter, J., Müller, M., Müller, S. and Kundisch, D. (2024), Moment or movement – the heterogeneous impact of the Black Lives Matter movement on personal and societal charitable crowdfunding campaigns. *Internet Research*, *34*, 6, 2151-2174. https://doi.org/10.1108/INTR-11-2022-0904

Shmueli, G. (2010). To explain or to predict? Statistical Science, 25(3), 289-310. https://doi.org/10.1214/10-STS330

Shyalika, C., Wickramarachchi, R., & Sheth, A. P. (2024). A comprehensive survey on rare event prediction. *ACM Computing Surveys*, *57*(3), 1-39. https://doi.org/10.1145/3699955

Singh, P., & Jain, A. (2024). CUP_CDLSTM: Civil Unrest Event Prediction Using Convolutional Neural Network, DistilBERT, and Long Short-Term Memory. *IEEE Transactions on Computational Social Systems*. https://ieeexplore.ieee.org/document/10753286

Stanovich, K. E. (2023). Myside bias in individuals and institutions. In *The epistemology of democracy* (pp. 170-194). Routledge.

Stremlau, N., McGeer, C., & Straub, M. (2024). Deciphering Digital Hate: Assessing the Evidence between Online Speech and Offline Violence in Africa. *Global Media Journal - German Edition*, *13*(2). https://doi.org/10.60678/gmj-de.v13i2.280

Stokes, B., & Williams, D. (2018). Gamers who protest: Small-group play and social resources for civic action. *Games and Culture*, *13*(4), 327-348. https://doi.org/10.1177/1555412015615770

Tarafdar, M., & Kajal Ray, D. (2021). Role of social media in social protest cycles: A sociomaterial examination. *Information systems research*, *32*(3), 1066-1090. https://doi.org/10.1287/isre.2021.1013

Trafimow, D., Cosstick, N., & Osman. M (2024). Problem Testing (and beyond) in the Messy World of Public Policy. Scientific Journal of Research and Reviews. Scientific Journal of Research and Reviews, 4.3. https://irispublishers.com/sjrr/pdf/SJRR.MS.ID.000586.pdf

Trafimow, D., & Osman, M. (2022). Barriers to converting applied social psychology to bettering the human condition. *Basic and Applied Social Psychology*, *44*(1), 1-11. https://doi.org/10.1080/01973533.2022.2051327

Traunmüller, R. (2023). Testing the 'Campus Cancel Culture' Hypothesis. Available at SSRN: http://dx.doi.org/10.2139/ssrn.4392840

Tuke, J., Nguyen, A., Nasim, M., Mellor, D., Wickramasinghe, A., Bean, N., & Mitchell, L. (2020). Pachinko Prediction: A Bayesian method for event prediction from social media data. *Information Processing & Management*, *57*(2), 102147. https://doi.org/10.1016/j.ipm.2019.102147

van Doorn, M. (2024). The skeptical import of motivated reasoning: a closer look at the evidence. *Thinking & Reasoning*, *30*(4), 548-578. https://doi.org/10.1080/13546783.2023.2276975

Venneti, L. (2018). *Computational Study of the Politics of Emotions: An Analysis of Curiosity and Outrage in Online Platforms* (Doctoral dissertation, International Institute of Information Technology Hyderabad).

von der Burg, L., & Krasmann, S. (2024). Naming the city: on the governing forces of narratives in the formation of security dispositifs. *Critical Studies on Security*, *12*(1), 33-47. https://doi.org/10.1080/21624887.2023.2286770

Watkins, T. R., & Harvey, L. A. (2020). I am not biased. It is everyone else's problem. *Spinal Cord*, *58*(4), 389-390. https://doi.org/10.1038/s41393-020-0446-2

Wells, G., Romhanyi, A., Reitman, J. G., Gardner, R., Squire, K., & Steinkuehler, C. (2024). Right-wing extremism in mainstream games: A review of the literature. *Games and Culture*, *19*(4), 469-492. https://doi.org/10.1177/15554120231167214

Williams, D. (2023). Bad beliefs: Why they happen to highly intelligent, vigilant, devious, self-deceiving, coalitional apes. *Philosophical Psychology*, *36*(4), 819-833. https://doi.org/10.1080/09515089.2023.2186844

Wu, C., & Gerber, M. S. (2017). Forecasting civil unrest using social media and protest participation theory. *IEEE Transactions on Computational Social Systems*, *5*(1), 82-94. https://doi.org/10.1109/TCSS.2017.2763128

Yee, A. K. (2023). Machine learning, misinformation, and citizen science. *European Journal for Philosophy of Science*, *13*(4), 56. https://doi.org/10.1007/s13194-023-00558-1

Yee, A. K. (2025). Construct Validity in Automated Counterterrorism Analysis. *Philosophy of Science*, 1-18. https://doi.org/10.1017/psa.2024.65

Zadeh, M., & Cicekli, I. (2023). Protest Event Analysis: A New Method Based on Twitter's User Behaviors. *Information Technology and Control*, *52*(2). https://itc.ktu.lt/index.php/ITC/article/view/33077

Figure 1. Causal representations of content-to-belief-to-behavior

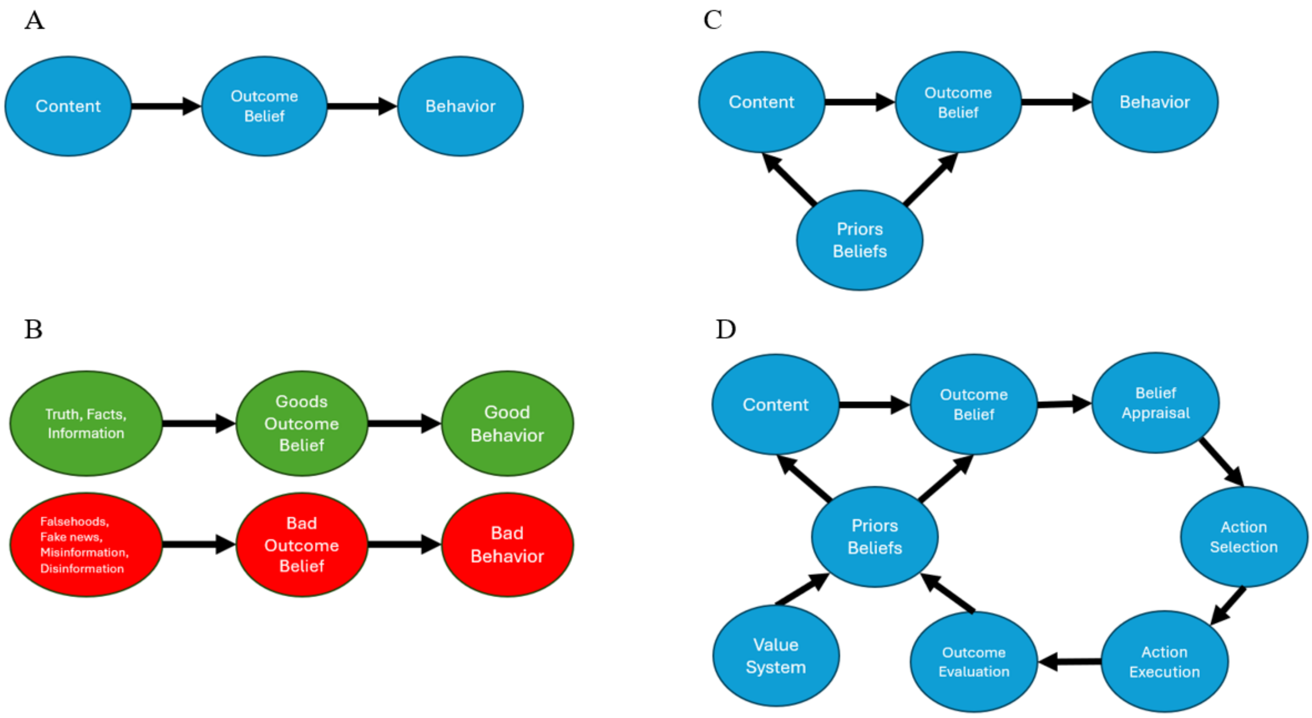

Figure 1. Causal representations of content-to-belief-to-behavior. In (A) the simple model captures what is typically assumed to be the causal relationship between the information we encounter (e.g. on social media) that influences outcome beliefs and that in turn raises the probability that we act on them (for discussion see, Adams et al., 2023). Granados Samayoa and Albarracín's (2025) Taxonomy of beliefs has been applied here, where outcome beliefs as per their definition most closely corresponds to the types of beliefs in the simple model. In (B) the same simple model presented in (A) is now recast as the transmission heuristic, where the content is valenced as well as the beliefs and behavioral outcomes along the causal chain. In (C) the simple chain model presented in (A) now includes a common cause structure where prior beliefs can influence what content is sought as well as what outcome beliefs are generated. If Granados Samayoa and Albarracín's (2025) Taxonomy of beliefs is applied here, prior beliefs could refer to existential beliefs or descriptive beliefs. In (D) the causal model integrates the Granados Samayoa and Albarracín's (2025) causal chain (C) and typical value-based decision-making models (e.g., Ludwiczak et al., 2020; Osman, & Wiegmann, 2017; Rangel et al., 2008).