****The paper has been accepted for publication by the American Journal of Psychology****

The intervention bias: people overpredict social problems upon which they believe society can intervene

Francesco Rigoli[1]

[1] *City, University of London, Northampton Square, London, EC1V 0HB, UK*

Correspondence: Francesco Rigoli

Department of Psychology

City, University of London

Northampton Square, London, UK EC1V 0HB

francesco.rigoli@city.ac.uk

Abstract

Evidence indicates that, when people forecast potential social risks, they are not only guided by facts, but often also by motivated reasoning. Here I apply a Bayesian decision framework to interpret the role of motivated reasoning during forecasting and assess some of the ensuing predictions. In two online studies, for each of a set of potential risky social events (such as economic crisis, rise of income inequality, and increase in violent crime), participants expressed judgments about (i) the probability that the event will occur, (ii) how negative occurrence of the event would be, and (iii) whether society is able to intervene upon the event. Supporting predictions of the Bayesian decision model, the analyses revealed that participants who deemed the events as more probable also assessed occurrence of the events as more negative and believed society to be more capable to intervene upon the events. Supporting the notion that a social threat is appraised as more probable when an intervention is deemed to be possible, these findings are compatible with a form of intervention bias. These observations are relevant for campaigns aimed at informing the population about potential social risks such as climate change, economic dislocations, and pandemics.

Keywords: forecasting; risk; motivated reasoning; Bayesian; intervention bias; decision;

Introduction

The factors that can threaten the prosperity of human communities are many and of different kind. Some, such as natural calamities, infectious diseases, and foreign invasions, are exogenous. Others are the product of dynamics internal to the community itself, and include, among others, exploitation of the weak, civil war, and crime. When comparing different human societies regarding the vulnerability to the above risks, a marked variability is evident. Yet, it is hard to point to any society safe from all. It is not surprising, thus, that a ubiquitous pursuit among experts and laypeople alike is attempting to forecast risks that can potentially endanger society. How are these forecasts constructed? An ideal forecaster, as for example envisaged by formal models employing Bayesian inference (Pole et al., 2018; West & Harrison, 2006), should look for the various signs available which are deemed to be indicators of a risky event. For example, extremely bad weather combined with the presence of parasites in the fields can be interpreted as anticipating a disastrous harvest and famine. Moreover, prior beliefs should play a role, too (Pole et al., 2018; West & Harrison, 2006): the bad weather and the parasites notwithstanding, famine might still be viewed as unlikely if it has never occurred before.

According to empirical evidence, the way people predict potential social risks is nonetheless rather different from the way an ideal forecaster does. An essential difference appears to be that, while an ideal forecaster works in a way that maximises accuracy, people are often driven by motives other than accuracy seeking; in other words, people's predictions appear often to be shaped by motivated reasoning (Dickerson & Ondercin, 2017; Kahan, 2016a; 2016b; Kahan et al., 2102; 2017; Kunda, 1990; Maguire, 2022). To illustrate how motivated reasoning works, consider the case of stereotyping. Empirical research suggests that, rather than being grounded on factual considerations, negative stereotypes often serve the interests of the dominant group within a society (Kunda & Spencer, 2003; Kundra & Sinclair, 1999). The processes underpinning the impact of motivated reasoning upon human forecasting are still poorly understood. Here, I examine whether any novel insight on these can be gained by considering a recent computational theory based on a Bayesian decision framework - I shall refer to this as the Bayesian Decision Model of Forecasting (BDMF). Originally, the theory was developed to explain how people make inference about something ongoing in the present or about something that occurred in the past (Rigoli, 2021a; Rigoli, 2021b; Rigoli, 2022a, Rigoli, 2022b; Rigoli et al., 2021). As discussed below, it is nonetheless straightforward to extend this framework to explain how people predict future events.

To identify the specific empirical predictions ensuing from the BDMF, it is instructive to compare it against an ideal forecaster model. Thus, before overviewing the BDMF in detail, ideal forecasting theory will be briefly considered in the next section.

Ideal forecasting

Let us consider an issue which, according to most scientists in the field (Change et al., 2006), is one of the most daunting currently faced by humanity: the issue of climate change. Consider an individual who has to arbitrate between one hypothesis claiming that “In the next decades, the earth’s temperature will increase dramatically” (a climate change hypothesis) versus the hypothesis that “In the next decades, the climate will remain more or less the same as it is today” (a climate hoax hypothesis). We can identify some of the key mental representations at play when an individual is considering the two hypotheses. A first form of representation is the belief about how likely the event in question is. For example: how likely is climate change to occur? I shall call this representation *Probability Belief*. Second, an individual represents how bad the occurrence of the event would be for society. For example: if climate change occurs, how bad would this be? I shall call this representation *Value Belief.* Third, an individual can have an idea of whether society’s intervention can affect the event.

For example, do people's choices have an impact upon climate change? This representation can be referred to as *Intervention Belief*.

What is the relationship between Probability, Value, and Intervention Beliefs? Let us examine how an ideal forecaster model, such as one based on Bayesian reasoning (Pole et al., 2018; West & Harrison, 2006), would answer this question. A first possibility is that an ideal forecaster estimates the likelihood of an event independent of its value, thus implying no link between Probability Beliefs and Value Beliefs. Yet, an alternative possibility can be envisaged by applying the principles of ideal forecasting theory. It has been documented that, in people's everyday life, very large rewards (e.g., winning a lottery) are relatively rare compared to smaller payoffs (e.g., receiving a salary payment) (Pleskac & Hertwig, 2014). Likewise, very large punishments (e.g., the death of a relative) are less common than smaller ones (e.g., the bus being delayed). This suggests that, at an ecological level, people typically experience a relationship between the probability of an event and its value. Following ideal forecasting principles, such prevalent experience may lead people to infer that more negative events are less probable and, vice versa, that less probable events are more negative. This predicts a relationship between Probability Beliefs and Value Beliefs whereby less probable events are usually attributed a more negative value[1]. As an example of how this logic can be applied,

[1] The same prediction can be derived if one hypothesises that people's forecasts are characterised by an optimism bias (Sharot, 2012). This hypothesis is consistent with evidence indicating that people often believe that they are less likely than others to experience certain negative events. For example, evidence indicates that people typically perceive themselves as being less at risk of being crime victims, that smokers believe they are less likely to contract lung cancer than other smokers, that first-time bungee jumpers believe they are less at risk of an injury than other jumpers, and that traders think they are less exposed to potential losses in the markets than other traders (Chapin & Coleman, 2009; Weinstein & Klein, 1996). This raises the hypothesis that optimism biases may be at play also when people are forecasting social risks. If this is the case, then the prediction is that Probability Beliefs and Value Beliefs should be related; specifically, that an event is deemed to be more likely when its value is perceived to be less negative. In the case of climate change, the prediction would be that people believing that climate change is more likely also believe that climate change, if it eventually happens, will not be too catastrophic after all.

people believing that climate change is less likely are predicted to believe that the potential consequences of climate change are gloomier.

What are the implications of ideal forecasting theory regarding the relationship between Probability Beliefs and Intervention Beliefs? Consider an individual who thinks that, independent of whether society intervenes to prevent climate change, the probability of climate change remains 0.9. Compare this with an individual believing that, while the probability of climate change is still 0.9 if society does nothing, this can be reduced to 0.5 if society intervenes. In this scenario, ideal forecasting theory predicts that, compared to the second individual, the first one deems climate change to be overall less probable. As this example illustrates, ideal forecasting theory implies a relationship between Probability Beliefs and Intervention Beliefs whereby a negative event is deemed to be more likely when society's intervention is deemed to be less effective.

In short, ideal forecasting theory predicts no relationship between Probability Beliefs and Value Beliefs or, alternatively, it predicts that people deem events that are more negative also to be less probable. Moreover, it predicts a relationship between Probability Beliefs and Intervention Beliefs whereby a negative event is viewed as less likely when society's intervention is expected to be more effective. Let us now turn our attention to the BDMF and assess how this diverges from ideal forecasting theory.

The Bayesian Decision Model of Forecasting

To illustrate the BDMF, let us focus once again on the climate change example. The theory posits that three factors are critical in establishing whether the climate change or the climate hoax hypothesis will be endorsed by an individual (for a formal overview of the theory, see the Appendix). The first factor is available evidence. Reading about an interview of a climate expert, or experiencing extreme weather events such as droughts or hurricanes, may be interpreted by the individual as evidence supporting the climate change hypothesis over the climate hoax hypothesis. The second factor is represented by prior beliefs, corresponding to assumptions of various kind available independent of any novel evidence. The general assumption that things never really change, compared to a general apocalyptic outlook about the future (Foust & O'Shannon Marphy, 2009), will render the climate hoax hypothesis more appealing vis-à-vis the climate change hypothesis.

The role of novel evidence and of prior beliefs is analogous to the role played by these factors in ideal forecaster models such as those grounded on Bayesian inference (Pole et al., 2018; West & Harrison, 2006). What differentiates Bayesian inference from the BDMF is the third factor, which is unique to the latter model: this is the utility[2] associated with accepting or rejecting the different hypotheses[3]. The proposal is that, when arbitrating between the climate change and the climate hoax hypothesis, the individual asks: what are the consequences of accepting the climate change hypothesis (and of supporting policies aimed at dealing with climate change) if this turns out to be true? And if it turns out to be false? What are the consequences of accepting the climate hoax hypothesis (and of opposing policies aimed at dealing with climate change) if this turns out to be true? And if it turns out to be false? The answers to these questions are, according to the BDMF, critical for arbitrating which hypothesis will be endorsed. To understand the implication of the utility component, compare two individuals: the chief executive officer (CEO) of a multinational oil company and a common person. What are the consequences of accepting the climate change hypothesis (and of

[2] The model presupposes a broad definition of utility, which comprises motives of different kinds (e.g., economic interest, status, pleasure, avoidance of pain, etc.) and encompasses purely egoistic motives as well as altruistic ones.

[3] This implies that, if one defines accuracy in terms of the number of correct predictions made, an agent who follows ideal forecasting is more accurate compared to an agent who follows the BDMF. This is because an ideal forecaster relies on an integration of prior beliefs and novel evidence, which, mathematically, is the optimal strategy in terms of accuracy. By adding the utility component, the BDMF adds a bias element which, compared to ideal forecasting, hampers accuracy to some degree.

supporting policies aimed at dealing with climate change) if this turns out to be false? While the consequences are not much costly for the common person, they are very costly for the CEO inasmuch as policies targeting climate change disrupt economic opportunities for the latter individual. Likewise, what are the consequences of accepting the climate hoax hypothesis (and of opposing policies aimed at dealing with climate change) if this turns out to be true? Once again, because of economic considerations, accepting this hypothesis appears to be more advantageous for the CEO compared to the common person. Thus, overall, the BDMF predicts that, other things being equal, the CEO will be more likely than the common person to accept the climate hoax hypothesis (Jaworska, 2018). The utility component is proposed to act subconsciously: the CEO might be staunchly convinced that the climate hoax hypothesis is true, without realising that this conclusion is not based on facts but on convenience.

To summarise, the BDMF proposes that people's forecasts are produced by the interaction among three factors: novel evidence, prior beliefs, and utility. While the first two factors encourage forecasts that are accurate, the utility component nudges people towards forecasts that support their motives, thus accounting for motivated reasoning.

Now that the BDMF has been illustrated, let us overview the specific empirical predictions ensuing from this theory, examining first the relationship between Value Beliefs and Probability Beliefs. Logically, when a potential event is more negative, the cost of rejecting the hypothesis claiming that the event will occur grows. For example, rejecting the climate change hypothesis if this is true is more costly for someone believing that climate change is more catastrophic. Since, as illustrated above, the BDMF postulates that people's judgments take the costs of rejecting the different hypotheses into account, the theory predicts a relationship between Value Beliefs and Probability Beliefs whereby events appraised as more negative are also deemed to be more likely. Note that this prediction is different from the one ensuing from ideal forecasting theory – remember that the latter implicates no relationship between Value Beliefs and Probability Beliefs or, alternatively, it implicates a relationship whereby more negative events are deemed to be less likely.

Let us now assess the BDMF with regard to the relationship between Probability Beliefs and Intervention Beliefs. As explained above, the theory posits that an individual asks the following questions: what are the consequences of accepting the climate change hypothesis (and of supporting policies aimed at dealing with climate change) if this turns out to be true? And if it turns out to be false? What are the consequences of accepting the climate hoax hypothesis (and of opposing policies aimed at dealing with climate change) if this turns out to be true? And if it turns out to be false? If a person believes that society can intervene effectively upon climate change, then the cost of neglecting the climate change hypothesis if this is true grows: failing to recognise the looming threat of climate change implies that society will not intervene, and therefore that society's intervention cannot reduce the risk. On this basis, the person will be more inclined to accept the climate change hypothesis compared to someone who believes that society can do nothing to deal with climate change. Thus, the BDMF predicts a correlation between Probability Beliefs and Intervention Beliefs whereby people believe that a social risk is more likely to occur when they believe that society can intervene to reduce the risk. Note that this prediction is opposite to the one ensuing from ideal forecasting theory as spelled out above (Pole et al., 2018; West & Harrison, 2006): the latter implies that a *lower* likelihood is attributed to risks upon which society is believed to be able to intervene, whereas the BDMF implies that a *higher* likelihood is attributed to such risks.

Overall, ideal forecasting and the BDMF make divergent predictions. The former implies no link between Probability Beliefs and Value Beliefs, or, alternatively, it implies an association between high probable and *less* negative events. The opposite, that is, an association between high probable and *more* negative events, is predicted by the BDMF. The two theories diverge also regarding the predicted link between Intervention Beliefs and Value

Beliefs, with ideal forecasting (the BDMF) postulating attribution of lower (higher) probability when society is judged as being able to intervene. Below, I present the findings of two studies where these predictions were tested empirically.

Study 1

Participants

Two-hundred participants resident in South Africa were recruited online from the Prolific website (age: mean = 29, SD = 9; 99 females) (no data were excluded from the analysis). This country was chosen because the number of South African participants potentially available from Prolific is vast, and yet not as many studies have been done there compared to other countries. The sample size was established a priori adopting G-Power based on a multiple regression analysis with effect size equal to $f^2 = .07$, statistical power equal to $\beta = 0.8$, and two-tailed type-I error probability equal to $\alpha = .05$. This requires a sample of 176 participants, which was rounded to 200. The study was approved by the research ethics committee of the University to which the author is affiliated.

Measures and procedure

I investigated people's reported beliefs about a set of potential risks facing society. I focused on three issues: economic crisis, income inequality, and violent crime. For each issue, I asked three questions aimed at measuring Probability Beliefs, Value Beliefs, and Intervention Beliefs, respectively. For example, for economic crisis I assessed Probability Beliefs by asking:

"In your opinion, how likely is that, in the next five years, there will be a severe economic crisis?".

Options available were: Very unlikely, Quite unlikely, Somewhat likely, Quite likely, Very likely. I assessed Value Beliefs by asking:

“Imagine that, five years from now, a severe economic crisis has indeed occurred. How bad would this be for the country?”

Options available were: Not bad, A little bad, Somewhat bad, Quite bad, Very bad, Extremely bad. And I assessed Intervention Beliefs by asking:

“In your opinion, do people's political decisions have an impact on whether severe economic crises will occur?”

Options available were: No impact, A little impact, Some impact, Substantial impact, Strong impact. For the other potential risks examined, while the content obviously changed, the same format of the questions was employed (the specific text for each issue is reported in table 1).

After being asked their age and gender, participants were presented with the questions about economic crisis, income inequality, and violent crime, in this order for all participants (for each issue, the question probing Probability Belief was presented first, followed by the question about Value Belief and by the one about Intervention Belief, in this order). Completing the survey took approximately two minutes and was rewarded with £.20.

Analysis

To address the issue of multiple comparisons, I calculated the total score for Probability, Value, and Intervention Beliefs, corresponding to $Probability_{TOTAL}$, $Value_{TOTAL}$, and $Intervention_{TOTAL}$, respectively. $Probability_{TOTAL}$ was equal to the sum of Probability Beliefs across the issues of economic crisis, income inequality, and violent crime; the same procedure was used to calculate $Value_{TOTAL}$, except that now Value Beliefs were summed, and for calculating $Intervention_{TOTAL}$, for which Intervention Beliefs were summed.

Next, I fitted a multiple regression model of $Probability_{TOTAL}$ including $Value_{TOTAL}$ and $Intervention_{TOTAL}$ as predictors. Testing the effect exerted by the two predictors allows one to compare the predictions ensuing from ideal forecasting theory and from the BDMF. The former predicts no effect exerted by $Value_{TOTAL}$, or alternatively, a negative effect exerted by $Value_{TOTAL}$, plus a negative effect exerted by $Intervention_{TOTAL}$. The BDMF predicts a positive effect of $Value_{TOTAL}$ combined with a positive effect of $Intervention_{TOTAL}$.

Note that, for the different issues considered here (economic crisis, income inequality, and violent crime), people may report on average very different Probability, Value, and Intervention Beliefs. For instance, people may overall report that increases in violent crime are not as negative and not as likely, while believing that economic crises are very negative and very likely. However, note that this is not a concern for the analyses performed here, because the analyses look at correlations across participants, and not across issues. In other words, for each issue taken individually, there will be some variability across people regarding Probability, Value, and Intervention Beliefs. The question is the following: are these variables correlated across participants? This question is independent of whether the issues at hand vary in terms of average among themselves.

Results

Descriptive statistics are reported in table 2. When assessing the regression model of $Probability_{TOTAL}$, a positive significant effect emerged for both $Value_{TOTAL}$ (b = .435, 95%CI [.330, .540], t(197) = 8.17, p < .001) and $Intervention_{TOTAL}$ (b = .077, 95%CI [.017, .136], t(197) = 2.55, p = .011)[4]. Table 3 reports separate tests for each issue, and shows that the effect of $Value_{TOTAL}$ was driven by all three issues, while the effect of $Intervention_{TOTAL}$ was driven by income inequality and violent crime, but not by economic crisis (given that, for the latter, the test was non-significant). Altogether, with the possible exception of the issue concerning economic crisis, the results support the BDMF insofar as they reveal that both $Value_{TOTAL}$ and $Intervention_{TOTAL}$ are positively related with $Probability_{TOTAL}$.

How general are these findings? To assess this, I planned a second study equivalent to the first one except for two features. First, in order to ascertain that the effects are at play in different contexts, the new study was conducted in a different country: the UK. Second, in order to examine whether the effects generalise across various social issues, I considered a broader array of potential risks. I focused on seven of such risks: economic crisis, income inequality, violent crime, climate change, immigration, pandemic, and war.

Study 2

Participants

Two-hundred participants resident in the UK were recruited online from the Prolific website (age: mean = 39, SD = 16; 97 females) (no data were excluded from the analysis). Like in study 1, the sample size was established a priori adopting G-Power based on a multiple regression analysis with effect size equal to $f^2 = .07$, statistical power equal to $\beta = 0.8$, and two-tailed type-I error probability equal to $\alpha = .05$. This requires a sample of 176 participants, which was rounded to 200. The study was approved by the research ethics committee of the University to which the author is affiliated.

[4] For completeness, I report also the correlation between the two predictors $Value_{TOTAL}$ and $Intervention_{TOTAL}$, which was significant (r(198) = .270, p < .001, 95% CI [.137, .394]).

Measures and procedure

The same measures and procedures adopted in Study 1 were employed here too, except that now seven issues were covered: economic crisis, income inequality, violent crime, climate change, immigration, pandemic, and war, presented in this order to all participants. The inclusion of a large set of issues is important. If the effect is restricted exclusively to a small number of issues, it may be due to accidental factors that apply only locally, and not to a general phenomenon as the one envisaged by the BDMF. Ensuring that the effect applies to most issues is essential to provide evidence against this possibility. Once again, for each issue participants were asked three questions aimed at measuring Probability Beliefs, Value Beliefs, and Intervention Beliefs, respectively (see above for an example of the questions concerning the issue of economic crisis; see table 1 for the questions employed for all other issues). Completing the survey took approximately three minutes and was rewarded with £.30.

Analysis

The same analysis approach employed in study one was employed here too, except that now I considered the seven issues listed above to derive $Probability_{TOTAL}$, $Value_{TOTAL}$, and $Intervention_{TOTAL}$. As before, I fitted a multiple regression model of $Probability_{TOTAL}$ including $Value_{TOTAL}$ and $Intervention_{TOTAL}$ as predictors, and I tested the effect exerted by the two predictors.

Results

Descriptive statistics are reported in table 4. Replicating study 1, the analyses showed that both $Value_{TOTAL}$ (b = .299, 95%CI [.182, .415], t(197) = 5.04, p < .001) and

$Intervention_{TOTAL}$ (b = .255, 95%CI [.165, .346], t(197) = 5.59, p < .001) exerted a positive effect upon $Probability_{TOTAL}$[5]. Table 5 reports separate tests for each issue, showing that the effect of $Value_{TOTAL}$ was driven by all issues with the exception of pandemic and war (given that, for these issues, the test was non-significant), while the effect of $Intervention_{TOTAL}$ was driven by all issues.

Discussion

The paper introduces the BDMF and, in two empirical studies, assesses its predictions vis-à-vis ideal forecasting theory. While failing to support predictions derived from the latter, the data appear to be broadly consistent with the BDMF. Specifically, the data indicate that people report risky events to be more probable when the events are appraised as more negative and when society is judged to be capable of intervening upon these events – we can refer to the latter effect as to an *intervention bias*.

These observations are broadly in line with the notion that people's forecasts, far from being grounded solely on epistemic considerations, are substantially shaped by motivated reasoning (Dickerson & Ondercin, 2017; Kahan, 2016a; 2016b; Kahan et al., 2102; 2017; Kunda, 1990; Maguire, 2022). Recently, scholars have argued that human reasoning has not evolved to develop veridical descriptions of reality, but rather as a tool for mastering social interactions (Butterworth et al., 2022; Mercier & Sperber, 2011; Trivers, 2011). The argument is that, when interpreting an event, the underlying psychological mechanisms have been shaped by the evolutionary imperative of persuading other group members in such a way that one can satisfy one's motives (note that these can be selfish as well as altruistic motives). This

[5] For completeness, I report also the correlation between the two predictors $Value_{TOTAL}$ and $Intervention_{TOTAL}$, which was significant (r(198) = .259, p < .001, 95%CI [.124, .384]).

evolutionary pressure, the argument goes, implies that accuracy is an important factor insofar as accurate beliefs are typically necessary to satisfy one's motives as well as to persuade others. Nonetheless, according to this perspective, accuracy seeking is not the whole story: one's motives can nudge reasoning away from the most accurate explanations. The BDMF is compatible with the picture just outlined. In essence, this model interprets forecasting as being akin to a decision-making process whereby, at a largely unconscious level, an agent selects the hypothesis which best integrates accuracy and other motives while, at a conscious level, *perceiving* the hypothesis as the one that is true.

The BDMF has analogies with a previous framework referred to as Error Management Theory (Haselton & Buss, 2000). The latter posits that, independent of what is true or false, endorsing certain beliefs and discarding others is beneficial in terms of evolutionary fitness, thus producing a bias towards some beliefs at the expense of others. A domain where this logic has been applied is mate selection, occurring when one has to establish whether a potential partner is sexually available or not. For males, the argument goes, there is a small fitness cost associated with judging the other as sexually available when the other actually is not; on the contrary, the fitness cost for this misjudgment is large for females. The ensuing prediction is that, compared to females, males are more likely to perceive a potential partner as being sexually available - a prediction corroborated empirically.

Both Error Management Theory and the BDMF assume that people's beliefs are shaped by considerations about the costs and benefits of rejecting and accepting the different hypotheses at hand. Yet, there is a fundamental difference between the two theories concerning how the costs and benefits of accepting/rejecting hypotheses are established, implying very different predictions. Error Management Theory presupposes that the cost/benefit calculation is not performed online by the brain, but, rather, that it is an implicit product of the species' evolutionary history. For example, the low costs of misjudging sexual availability for males means that, over the evolutionary history of the human species, males who were not equipped with a bias towards overpredicting sexual availability failed to reproduce. At the psychological level, the implication is that, according to Error Management Theory, the male's brain does not calculate the costs and benefits of accepting the hypothesis that the partner is sexually available; it simply manifests a fixed bias towards perceiving sexual availability. Put another way, according to Error Management Theory, the bias is rigid inasmuch as it neglects the specific costs and benefits associated with the ongoing context (e.g., the costs associated with one specific potential partner in a particular context).

By contrast, the BDMF implies that the cost/benefit assessment is performed online by the brain in a way that takes the ongoing context into account. In the case of mate selection, this implies that the male's brain is capable to assesses the specific costs and benefits associated with viewing one specific individual as sexually available in that specific context. In the case of forecasting, Error Management Theory predicts a bias towards overpredicting negative events compared to neutral ones, but it does not predict that the bias varies based on how negative the event is, nor based on whether society can intervene effectively to reduce the risk (Baumeister et al., 2001; Blaine & Boyer, 2018; Boyer & Bergstrom, 2011; Rozin & Royzman, 2001). Therefore, Error Management Theory struggles to interpret the empirical observations that emerged in the present paper.

The findings reported here have potential implications for the question of how to devise effective campaigns aimed at informing the population about potential social risks. In various domains, there is evidence that a substantial number of people is inclined towards denying looming social risks, even when exposed to compelling evidence about the urgency of such risks (Sinatra & Hofer, 2021). In contemporary society, a case in point is represented by climate change (Washington, 2013), though other important domains appear also to be characterised by alarming levels of denial. In these domains, it is particularly important for public campaigners to devise effective communication strategies. In a nutshell, the findings reported here encourage campaigners to develop messages that, in addition to stressing the negative nature of the risk, at the same time empower people by emphasising how society can intervene. In simple terms, the data suggest that an effective message may sound like: "If we ignore the issue, there will be very serious consequences, but, if we intervene decisively, we have the means to avoid these consequences" (of course, the details of the message need to be carefully crafted in light of the specific domain).

The argument just considered speaks to a classical debate in the study of persuasion, that is, the debate about whether, and to what extent, negative messages (i.e., those emphasising the negative consequences of social risks) are effective (Dillard & Shen, 2013). Empirical evidence on this is mixed, showing that negative messages work well in some contexts but not in others (Dillard & Shen, 2013). The findings presented here offer a possible explanation of this mixed evidence: negative messages might be effective when they also emphasise that negative consequences can be avoided, while being counterproductive when no possibility of avoidance is envisaged.

Our findings may be relevant not only for information campaigns on social risks, but also for campaigns targeting individual risks such as those linked with smoking or with other unhealthy behaviours. In this context, too, the suggestion is to employ messages where the negative consequences are stressed together with an emphasis on how appropriate behaviours can prevent these consequences. For example, rather than focusing only on the risk of lung cancer associated with smoking, a better strategy may be to stress also how the risk can be drastically reduced by quitting or reducing smoking.

One last point regarding information campaigns is worth stressing. Theories focusing on motivated reasoning such as the BDMF encourage campaigners to develop an accurate understanding of which specific motives are important for the group targeted by a message (Briñol & Petty, 2005; Leeper & Slothuus, 2014). Taking smoking as an example, different groups might be more or less motivated by avoiding aesthetic problems associated with smoking. Being able to recognise these specific motives is, according to motivated reasoning accounts, vital to design effective information campaigns.

Before concluding, two shortcomings of the research presented here need to be highlighted. First, this relies on self-report measures. A disadvantage of this approach is that the behavioural domain is ignored. For example, are participants attributing higher probability to climate change actually more likely to recycle? The link between the self-report measures employed here and their behavioural correlates remain conjectural. Second, the studies described here are correlational and not experimental. This means that, while they assess whether a relationship exists between two variables, these studies do not warrant the conclusion that one variable causes the other. More specifically, the BDMF presupposes that Value Beliefs and Intervention Beliefs exert a causal influence upon Probability Beliefs. Although this interpretation fits with the data presented here, other possibilities cannot be ruled out. Despite this limitation, we note at least one previous study supporting the notion that Intervention Beliefs affect Probability Beliefs (Kahan et al., 2015). This study has shown that people are more likely to believe in climate change when they are presented with a description of how geoengineering may help coping with this threat. The effect of offering geoengineering as a coping strategy can be interpreted as supporting the notion advocated here that Intervention Beliefs exert a causal influence upon Probability Beliefs.

To summarise, the paper proposes the BDMF, a novel perspective stressing the role of motivated reasoning at play when people forecast social risks. By showing that people judge that risky events are more likely when the events are appraised as more negative and when society is deemed to be able to intervene, the paper reports empirical evidence supporting the theory. Altogether, the research overviewed in the paper may be relevant for planning effective campaigns aiming at informing the population about potential risks such as climate change, pandemics, and economic turmoil.

Data availability

Data are available at https://osf.io/yf96n/.

Conflict of interest statement

The author has no competing interests to declare.

Appendix

The BDMF is based on a standard Bayesian decision framework (Bishop, 2006). This is implemented by relying on a probabilistic generative model defined by the following joint probability:

$$P(PBS, Hyp, HDec, EOut, Evi) = P(PBS)\, P(HDec)\, P(Hyp|PBS)\, P(Evi|Hyp)\, P(EOut|Hyp, HDec)$$

Let us spell out the variables in the generative model:

- Prior Belief System (PBS) is a categorical variable with number of categories equal to n_{PBS}, where each category is associated with a probability. Consider the example of climate change as illustrated above. Here PBS has two possible states, implying $n_{PBS} =$

2. These are *PBS = Stable*, indicating that the environment never really changes, and *PBS = Unstable*, indicating that the environment is unstable. The probability of the environment being stable is *P(PBS = Stable) = x* and the probability of the environment being unstable is *P(PBS = Unstable) = 1 - x* (where $0 \leq x \leq 1$).

- Hypothesis (Hyp) is a categorical variable with number of categories equal to n_{Hyp}. In our example, we can set $n_{Hyp} = 2$, *Hyp = Yes* for the climate change hypothesis, and *Hyp = No* for the climate hoax hypothesis. In the model, Hyp depends on PBS, according to the logic that whether the climate change hypothesis is true or not depends on whether, more generally, the environment is stable or not. The dependency between PBS and Hyp is described by the conditional probabilities of Hyp, which are *P(Hyp = Yes | PBS = Stable) = y*, *P(Hyp = No | PBS = Stable) = 1 – y*, *P(Hyp = Yes | PBS = Unstable) = z*, *P(Hyp = No | PBS = Unstable) = 1 – z* (where $0 \leq y \leq 1$ and $0 \leq z \leq 1$).

- HDec is a categorical variable with the number of categories being $n_{HDec} = n_{Hyp}$. In our example, *HDec = YesAcc* when the climate change hypothesis is accepted (or, equivalently, when the climate hoax hypothesis is rejected) and *HDec = NoAcc* when the climate change hypothesis is rejected (or, equivalently, when the climate hoax hypothesis is accepted). Probabilities for HDec are *P(HDec = YesAcc) = u* and *P(HDec = NoAcc) = 1 - u* (where $0 \leq u \leq 1$).

- Evidence (Evi) indicates novel available evidence. This is represented by a real number (with negative numbers supporting the climate change hypothesis and with positive numbers supporting the climate hoax hypothesis) drawn from a Gaussian distribution. Evi is conditioned on Hyp, the logic being that the evidence one experiences depends on whether the climate change hypothesis is true or not (e.g., experiencing extreme temperature episodes is more probable if the climate change hypothesis is true). The dependency between Hyp and Evi is described by the following conditional probability:

$$\mathrm{P(Evi \mid Hyp = k)} = \mathcal{N}(\mu_{Evi|k}, 1/\lambda_{Evi}{}^{2})$$

Here, every category of Hyp k has its own associated average $\mu_{Evi|k}$; for instance the model includes $\mu_{Evi|Yes}$ (conditional on the climate change hypothesis being true) and $\mu_{Evi|No}$ (conditional on the climate hoax hypothesis being true). The parameter $\lambda_{Evi}{}^{2}$ reflects the weight or precision of Evi and in our model it is equal for all levels of Hyp (in principle, a specific weight for each level of Hyp can be implemented).

- Expected Outcome (EOut) is a Gaussian variable conditioned on both Hyp and HDec. EOut indicates the benefits (when it is a positive number) or the costs (when it is a negative number) associated with accepting any hypothesis if it is true or false. Its conditional probability is:

$$\mathrm{P(EOut \mid Hyp = k, HDec = j)} = \mathcal{N}(\mu_{EOut|k,j}, \sigma^2_{Eout})$$

This indicates a specific average for each combination of Hyp and HDec. In our example, the model comprises $\mu_{EOut|Yes,YesAcc}$ (the expected outcome if the climate change is true and it is correctly accepted), $\mu_{EOut|Yes,NoAcc}$ (the expected outcome if the climate change hypothesis is true but it is wrongly rejected), $\mu_{EOut|No,NoAcc}$ (the expected outcome if the climate hoax hypothesis is true and it is correctly accepted), $\mu_{EOut|No,YesAcc}$ (the expected outcome if the climate hoax hypothesis is true but it is wrongly rejected). The parameter σ^2_{Eout} reflects the uncertainty about the outcome and it is equal for all combinations of Hyp and HDec (although, in principle, one can also implement a specific weight for each combination).

When a new Evi is observed, the generative model can be inverted to make a set of inferences. Each inference treats one distinct level of HDec j as observed and, on this basis, calculates the conditional probability of EOut given the observed value for Evi and given HDec = j. This corresponds to a posterior Gaussian distribution:

$$\mathrm{P(EOut \mid Evi, HDec = j)} = \mathcal{N}(\mu_{EOut|Evi,j}, \sigma^2_{POST})$$

Where $\mu_{EOut|Evi,j}$ is the posterior average for the expected outcome. In our example, two inferences are made. The first calculates $\mu_{EOut|Evi,YesAcc}$, that is, the posterior average if the climate change hypothesis is accepted. The second inference calculates $\mu_{EOut|Evi,NoAcc}$, that is, the posterior average if the climate hoax hypothesis is accepted. After these quantities have been estimated, the BDMF makes a decision simply by choosing the hypothesis associated with the highest posterior $\mu_{EOut|Evi,j}$. For instance, it either chooses to accept the climate change hypothesis or to reject it (or, equivalently, to reject the climate hoax hypothesis or to accept it, respectively).

References

Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Pychology*, *5*(4), 323-370.

Blaine, T., & Boyer, P. (2018). Origins of sinister rumors: A preference for threat-related material in the supply and demand of information. *Evolution and Human Behavior, 39*(1), 67-75.

Boyer, P., & Bergstrom, B. (2011). Threat-detection in child development: An evolutionary perspective. *Neuroscience & Biobehavioral Reviews, 35*(4), 1034-1041.

Briñol, P., & Petty, R. E. (2005). Individual differences in attitude change. In D. Albarracín, B. T. Johnson, & M. P. Zanna (Eds.), *The handbook of attitudes* (pp. 575–615). Lawrence Erlbaum Associates Publishers.

Butterworth, J., Trivers, R., & von Hippel, W. (2022). The better to fool you with: Deception and self-deception. *Current Opinion in Psychology*, 101385.

Change, A. D. C., Blair, T., Pachauri, R. K., & Pachauri, R. (2006). *Avoiding dangerous climate change*. Cambridge University Press.

Chapin, J., & Coleman, G. (2009). Optimistic bias: what you think, what you know, or whom you know?. *North American Journal of Psychology*, 11(1), 121-132.

Foust, C. R., & O'Shannon Murphy, W. (2009). Revealing and reframing apocalyptic tragedy in global warming discourse. *Environmental Communication*, *3*(2), 151-167.

Haselton, M. G., & Buss, D. M. (2000). Error management theory: a new perspective on biases in cross-sex mind reading. *Journal of Personality and Social Psychology, 78*(1), 81-91.

Jaworska, S. (2018). Change but no climate change: Discourses of climate change in corporate social responsibility reporting in the oil industry. *International Journal of Business Communication*, *55*(2), 194-219.

Kahan, D. M. (2016a). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. In R. A. Scott, & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*, https://doi.org/10.1002/9781118900772.etrds0417.

Kahan, D. M. (2016b). The politically motivated reasoning paradigm, Part 2: Unanswered questions. In R. A. Scott, & S. M. Kosslyn (Eds.), *Emerging trends in the social and behavioral sciences*, https://doi.org/10.1002/9781118900772.etrds0417.

Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, *2*(10), 732-735.

Kahan, D. M., Jenkins-Smith, H., Tarantola, T., Silva, C. L., & Braman, D. (2015). Geoengineering and climate change polarization: Testing a two-channel model of science communication. *The ANNALS of the American Academy of Political and Social Science*, *658*(1), 192-222.

Kahan, D. M., Peters, E., Dawson, E. C., & Slovic, P. (2017). Motivated numeracy and enlightened self-government. *Behavioural Public Policy*, *1*(1), 54-86.

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480-498.

Kunda, Z., & Spencer, S. J. (2003). When do stereotypes come to mind and when do they color judgment? A goal-based theoretical framework for stereotype activation and application. *Psychological Bulletin*, *129*(4), 522-544.

Kundra, Z., & Sinclair, L. (1999). Motivated reasoning with stereotypes: Activation, application, and inhibition. *Psychological Inquiry*, *10*(1), 12-22.

Leeper, T. J., & Slothuus, R. (2014). Political parties, motivated reasoning, and public opinion formation. *Political Psychology*, *35*, 129-156.

Maguire, A., Persson, E., Västfjäll, D., & Tinghög, G. (2022). COVID-19 and Politically Motivated Reasoning. *Medical Decision Making*, *42*(8), 1078-1086.

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57-74.

Pole, A., West, M., & Harrison, J. (2018). *Applied Bayesian forecasting and time series analysis*. Chapman and Hall/CRC.

Rigoli, F. (2021a). Masters of suspicion: A Bayesian decision model of motivated political reasoning. *Journal for the Theory of Social Behaviour*, *51*(3), 350-370.

Rigoli, F. (2021b). A computational perspective on faith: Religious reasoning and Bayesian decision. *Religion, Brain & Behavior*, *11*(2), 147-164.

Rigoli, F. (2022a). Deconstructing the conspiratorial mind: the computational logic behind conspiracy theories. *Review of Philosophy and Psychology*, 1-18.

Rigoli, F. (2022b). Belief formation in the social context: A Bayesian decision account. *Changing Societies & Personalities, 6*(4). doi:10.15826/csp.2022.6.4.201.

Rigoli, F., Martinelli, C., & Pezzulo, G. (2021). I want to believe: Delusion, motivated reasoning, and Bayesian decision theory. *Cognitive Neuropsychiatry*, *26*(6), 408-420.

Rozin, P., & Royzman, E. B. (2001). Negativity bias, negativity dominance, and contagion. *Personality and Social Psychology Review*, *5*(4), 296-320.

Sinatra, G. M., & Hofer, B. K. (2021). *Science denial: why it happens and what to do about it*. Oxford University Press.

Sharot, T. (2012). *The optimism bias: Why we're wired to look on the bright side*. Hachette UK.

Trivers, R. (2011). *Deceit and self-deception: Fooling yourself the better to fool others*. Penguin UK.

Washington, H. (2013). *Climate change denial: Heads in the sand*. Routledge.

Weinstein, N. D., & Klein, W. M. (1996). Unrealistic optimism: present and future. *Journal of Social and Clinical Psychology*, *15*(1), 1-8.

West, M., & Harrison, J. (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Tables

Issue	Belief	Text
economic crisis	Probability Belief	In your opinion, how likely is that, in the next five years, there will be a severe economic crisis?
	Value Belief	Imagine that, five years from now, a severe economic crisis has indeed occurred. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether severe economic crises will occur?
income inequality	Probability Belief	In your opinion, how likely is that, in the next five years, income inequality will grow substantially?
	Value Belief	Imagine that, five years from now, income inequality has indeed grown substantially. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether income inequality will grow?
violent crime	Probability Belief	In your opinion, how likely is that, in the next five years, episodes of violent crime will increase dramatically?
	Value Belief	Imagine that, five years from now, episodes of violent crime have indeed increased dramatically. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether episodes of violent crime will increase?
climate change	Probability Belief	In your opinion, how likely is that, in the near future, the earth's temperature will raise by two degrees or more?
	Value Belief	Imagine that, in the near future, the earth's temperature has indeed raised by two degrees or more. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether the earth's temperature will raise by two degrees or more?
immigration	Probability Belief	In your opinion, how likely is that, in the next five years, there will be a huge wave of immigration from foreign countries?
	Value Belief	Imagine that, five years from now, a huge wave of immigration from foreign countries has indeed occurred. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether there will be huge waves of immigration in the future?
pandemic	Probability Belief	In your opinion, how likely is that, in the next ten years, there will be a new pandemic as severe as the COVID-19 pandemic?
	Value Belief	Imagine that, ten years from now, a pandemic as severe as the COVID-19 pandemic has indeed occurred. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether there will be severe pandemics in the future?
War	Probability Belief	In your opinion, how likely is that, in the next two years, the country will be at war against another country?
	Value Belief	Imagine that, two years from now, the country is indeed at war against another country. How bad would this be for the country?
	Intervention Belief	In your opinion, do people's political decisions have an impact on whether, in the near future, the country will enter a war against another country?

Table 1. Questions employed in the survey. For study 1, the issues considered were only economic crisis, income inequality, and violent crime. For study 2, all issues in the table were included.

Issue	Variable	Mean	Std. Deviation	Skewness	Kurtosis
economic crisis	Probability	4.08	0.958	-0.854	0.249
	Value	6.23	1.111	-2.000	5.338
	Intervention	5.66	1.738	-1.336	0.775
income inequality	Probability	4.09	0.903	-0.716	-0.140
	Value	6.00	1.364	-1.778	3.225
	Intervention	5.43	1.853	-1.031	-0.190
violent crime	Probability	4.05	0.937	-0.757	0.042
	Value	6.42	0.841	-1.230	0.424
	Intervention	5.61	1.671	-1.121	0.154
total	Probability	12.2150	2.10498	-0.699	0.201
	Value	18.6500	2.41800	-1.038	0.655
	Intervention	16.6950	4.28530	-1.156	0.651

Table 2. Descriptive statistics of Study 1.

Issue	predictor	regression weight	95% CI regression weight	t (df = 197)	p
economic crisis	Value	.305	[.191, .419]	5.27	<.001
	intervention	.028	[-.044, .101]	.77	.442
income inequality	Value	.334	[.254, .414]	8.23	<.001
	Intervention	.064	[.005, .123]	2.15	.033
violent crime	Value	.363	[.215, .510]	4.85	<.001
	Intervention	.108	[.034, .182]	2.87	.034
total	Value	.435	[.330, .540]	8.17	<.001
	Intervention	.077	[.017, .136]	2.55	.011

Table 3. Results of the multiple regression analysis for Study 1.

Issue	Variable	Mean	Std. Deviation	Skewness	Kurtosis
economic crisis	Probability	4.03	0.910	-0.373	-1.040
	Value	5.3250	0.83237	-1.094	0.482
	Intervention	3.8700	1.08582	-0.785	-0.038
income inequality	Probability	3.81	0.989	-0.365	-0.783
	Value	4.5900	1.35316	-0.832	-0.051
	Intervention	3.9850	1.06322	-0.831	-0.140
violent crime	Probability	3.04	0.940	0.287	-0.183
	Value	4.8650	1.08288	-0.855	0.343
	Intervention	3.3750	1.04875	-0.166	-0.601
climate change	Probability	3.93	0.940	-0.656	-0.056
	Value	4.6100	1.18105	-0.664	-0.261
	Intervention	3.2150	1.23954	-0.256	-0.886
immigration	Probability	3.48	0.967	0.091	-0.807
	Value	2.9800	1.47665	0.517	-0.624
	Intervention	3.5150	1.14294	-0.394	-0.593
pandemic	Probability	2.92	1.140	0.179	-0.794
	Value	4.9650	1.00939	-0.877	0.110
	Intervention	2.5550	1.28265	0.432	-0.861
war	Probability	2.38	0.990	0.541	-0.152
	Value	5.3200	1.01625	-1.808	3.190
	Intervention	3.8600	1.06115	-0.685	-0.239
total	Probability	23.5750	3.88456	-0.016	-0.541
	Value	32.6550	4.10735	-0.446	-0.235
	Intervention	24.3750	5.31844	-0.364	0.046

Table 4. Descriptive statistics of Study 2.

Issue	predictor	regression weight	95% CI regression weight	t (df = 197)	p
economic crisis	Value	.382	[.245, .519]	5.50	<.001
	Intervention	.228	[.123, .333]	4.28	<.001
income inequality	Value	.303	[.208, .399]	6.27	<.001
	Intervention	.221	[.100, .347]	3.60	<.001
violent crime	Value	.210	[.102, .318]	3.83	<.001
	Intervention	.343	[.231, 454]	6.05	<.001
climate change	Value	.317	[.208, .426]	5.74	<.001
	Intervention	.111	[.007, .215]	2.11	.036
immigration	Value	.234	[.150, .319]	5.46	<.001
	Intervention	.133	[.023, .242]	2.39	.018
pandemic	Value	.093	[-.057, .244]	1.22	.223
	Intervention	.274	[.156, .392]	4.56	<.001
war	Value	-.059	[-.198, .081]	-.83	.409
	Intervention	.187	[.054, .321]	2.76	.006
total	Value	.299	[.182, .415]	5.04	<.001
	Intervention	.255	[.165, .346]	5.59	<.001

Table 5. Results of the multiple regression analysis for Study 2.