BELIEF DYNAMICS AND THE ROLE OF EPISTEMIC PEERSHIP IN IDENTITY CONSTRUCTION

MARK OHAN KARATOPRAK

BOĞAZİÇİ UNIVERSITY

2022

BELIEF DYNAMICS AND THE ROLE OF EPISTEMIC PEERSHIP IN IDENTITY CONSTRUCTION

Thesis submitted to the

Institute for Graduate Studies in Social Sciences

in partial fulfillment of the requirement for the degree of

Master of Arts

in

Cognitive Sciences

by

Mark Ohan Karatoprak

Boğaziçi University

2022

DECLARATION OF ORIGINALITY

I, Mark Ohan Karatoprak, certify that

- I am the sole author of this thesis and that I have fully acknowledged and documented in my thesis all sources of ideas and words, including digital resources, which have been produced or published by another person or institution;
- this thesis contains no material that has been submitted or accepted for a degree or diploma in any other educational institution;
- this is a true copy of the thesis approved by my advisor and thesis committee at Boğaziçi University, including final revisions required by them.

Signature.....

Date .....

ABSTRACT

Belief Dynamics and the Role of Epistemic Peership in Identity Construction

The drive for social inclusion has been observed to impact both individual beliefs and their corresponding behaviors. An individual's uncertainty regarding their position within their affiliated groups has been identified as a factor contributing to the spread of conspiratorial ideation and extreme beliefs. The following paper introduces a model of belief dynamics adapted from predictive brain models which attempts to consolidate a broad range of existing psychology literature. It predicts that individuals will attempt to resolve perceived divergence from the beliefs of their affiliated groups by adjustments to their ideological positions. The model defines this as one strategy of uncertainty mitigation in social contexts. In three experiments, participants were asked to indicate their beliefs regarding a range of topics and the importance of those topics to their identities. Their responses were used to generate the illusion of a group of participants with similar beliefs. Participants were shown fabricated results indicating their divergence in opinion on one particular topic out of the range of topics and then given the opportunity to change their position on that topic. We found participants were more likely to change their endorsement of particular statements to reflect group opinion if they identified strongly with the beliefs used to generate the group. These results suggest that individual endorsements are influenced by others with whom they share a range of ideological positions.

ÖZET

Düşünce Süreçleri ve Epistemik Karanlığın Kimlik Oluşumunda Rolü

Gruplara dahil olma dürtüsünün hem bireysel düşünce süreçlerini hem de bunlara karşılık gelen davranışları etkilediği gözlemlenmiştir. Bireyin bağlı olduğu gruplar içinde, kişisel konumuna dair hissettiği belirsizlik, bireylerin rasyonel süreçlerden sapıp, motive edilmiş akıl yürütmeye ve yanıltıcı örüntüler üzerinden ilişkiler tespit etmeye itebilir. Bu tarz yaklaşımlar komplo teorilerinin ve uç görüşlere inancın yayılmasına da sebep olur. Bugüne değin bu ve benzeri konularda yapılmış araştırmalarda, kişinin bağlı olduğu grupların genel görüşlerinin, bireylerin kişisel görüşleri üzerine olan etkileri doğrudan incelenmemiştir. Bu tez bünyesindeki üç deneysel çalışmada, bireylerin parçası oldukları gruplardaki genel görüşlerden nasıl etkilendiğini incelenmiştir. Bu deneylerde kişinin bağlı olduğu grup, belli konular üzerinde oluşan görüş birliği üzerinden tanımlanmış; bu görüş birliğinin daha önemsiz olabilecek ikincil konular üzerindeki görüşleri nasıl etkilediği araştırılmıştır. Her bir deneyde katılımcılardan önce bir dizi konu ile ilgili görüşlerini belirtmeleri istenmiş, ardından bu konulardan biri dışında diğer katılımcılarla benzer cevaplar verdikleri geri bildirimi kendilerine sunulmuştur. Bu yanıltıcı geri bildirimler, sanal bir grup aidiyeti oluşturmak amacıyla verilmiştir. Ardından, katılımcılara tekrar aynı konulardaki görüşleri sorulmuştur. Yürütülen üç deneyde, katılımcıların grupla uyuşmayan görüşleri değiştirmeye eğilimli olduğu; ancak katılımcıların burada paradigmadan kaynaklanan araştırma beklentileri ile uyumlu bir şekilde, görüş değişikliği yaptıkları belirlenmiştir. Bu bulgular,

düşünce süreçlerine dair hem psikolojik literatürden bulgular hem de felsefi bir perspektifle tartışılmış; ve öngörücü beyin modellerinin bireysel inançların dinamiklerinin daha iyi anlaşılmasında önemli bir rol oynayabileceği savunulmuştur.

TABLE OF CONTENTS

LIST OF TABLES

CHAPTER 1

LITERATURE REVIEW

What makes us who we are? While there are various theories proposing possible mechanisms and motivations driving human beliefs and behavior, there is no prevailing theory that can provide answers to what seem to be the many facets of the human identity. Some of the proposed models have led to empirical paradigms that have been successful in their attempts to identify certain critical mechanisms that drive biases and behaviors (Hogg & Aldeman, 2013; Leary, 2021; Sherman & Cohen 2006). The testing of such models typically involves observing the effects of manipulating one's perception of their own standing in their social environment, typically called identity threat, and is focused primarily on the negative consequences of those manipulations (Chen et al., 2010; Choi & Hogg, 2019; Hameiri et al., 2017). Experiments largely consist of generating feelings of ostracism and diminished self-concept in participants by either attempting to lower the participant's perceived standing in their own group, or by lowering the perceived standings of participants' affiliated groups. Tests of the inverse hypothesis, generating an association between individuals and groups which they actively choose not to affiliate with, have found that individuals will attempt to further distance themselves from any such groups (Hameiri et al., 2014; Hameiri et al., 2018). As such, findings addressing identity construction have significant implications for our understanding of the current state of polarization in the global information economy.

Research relevant to polarization is not only limited to work focused on identity and group dynamics however. Recent work in both the motivated reasoning and reasoning bias literatures have focused on an individual's drive to protect their conception of themselves[1] as the mechanism underlying the updating of beliefs (Hogg, 2020; Kahan, 2017; Knobloch-Westerwick et al., 2017; Wagoner et al., 2017). The success of such manipulations demonstrate that negatively impacting an individual's perception of themselves can drive them towards both holding extreme beliefs and engaging in extreme actions (Goldman & Hogg, 2016; Hogg, 2014; Hohman et al., 2017). Altogether, the findings across various research programs suggest self-preservation, and the uncertainty caused by threats to it, plays a central role in an individual's engagement in society (Chen et al., 2010; Choi & Hogg, 2019; Christopoulos & Tobler, 2016).

The position taken in this paper is that the individual identity is an ordered set of beliefs constructed to better navigate the uncertainties of social interaction. This interpretation of identity is novel in the context of psychology research, but is an extension of the Bayesian brain framework of Active Inference. The Active Inference framework views the biological agent as a system that maintains an embodied set of ordered beliefs which are updated to minimize incongruencies with perception (Friston et al., 2016; Pezzulo et al., 2018). As such, incongruence between beliefs and perception are viewed as threats to relevant beliefs. The relevant beliefs are updated in a way to best

[1] Various theories and models — identity protective cognition, uncertainty-identity theory, defensive self-affirmation, sociometer theory — encapsulate a similar concept.

mitigate the incongruence. The implications of this are expanded upon below, but effectively allow for a set of hypotheses to be drawn regarding the dynamics of belief and the impact of social context on the individual identity.

The literature review first details the current state of affairs in research on identity and group membership to better understand the mechanisms driving trust and attributions of epistemic validity. Specifically, we provide a comprehensive review of the existing literature beginning with identity construction and epistemic peership. We then outline the effects that prototypicality and ostracism have on individual beliefs and behaviors and explore the phenomenon of unfreezing. Then we focus on uncertainty as a central theoretical construct for social cognition and its effect on reasoning, the engagement of biases, and ultimately on the construction and prioritization of beliefs.

Lastly, we present a potential model that attempts to capture the various findings in the literature and explain them through belief dynamics and uncertainty. Particularly, this model views uncertainty as the phenomenological state resulting from the threats to existing beliefs generated by new information, or belief-perception incongruence. The model incorporates findings from the motivated reasoning literature as well as on the Bayesian brain model of active inference. Going forward, the presented model will be referred to as the belief dynamics model, or simply BDM. Chapter 2 presents three experiments, testing hypotheses of the BDM, that centralizes belief dynamics and positions one's own beliefs against those of their self identified

peers. Chapter 3 takes a deeper look at the philosophical implications of the model.

1.1 Self & identity

Across both cognitive and social psychology frameworks, the concepts of self and identity are largely regarded as a social phenomena in that the individual is seen to gains definition in relation to others (Berzonsky, 2011; Brewer, 1991; Tajfel, 1974). As such, a necessary constraint on an individual's beliefs regarding themselves is that they require engagement with at least one other individual in order to be constructed. To put it another way, engagement with society is a necessary, though insufficient, condition for the construction of an individual identity or notion of selfhood. While there is no consensus on which models of self or identity most accurately describe the mechanisms that drive the individual's engagement with society, many models have been proposed (Brubaker & Cooper, 2000; Kahan, 2017; Solomon et al., 2004). It seems that in order for the study of self and identity to move forward, we need an approach that encompasses the domains of cognition, neuroscience, social psychology, as well as economics and decision making.

The notion that one maintains a conceptualization of themselves, a social identity, which they update based on their perception of how they are valued by others, provides a useful framework with which to address the range of topics within the identity literature. The notion of self-esteem has also been useful in that it has served as an indicator for the state of the self-concept, often

reflecting one's conception of their social identity. It isn't clear however whether this distinction is necessary. According to Baumeister (2017), the terms are distinguishable in that "[t]he term self-concept refers to the totality of inferences that a person has made about himself or herself. These refer primarily to one's personality traits and schemas, but they may also involve an understanding of one's social roles and relationships" (Baumeister 2017, p. 1), whereas "[t]he term self-esteem refers to the evaluative dimension of the self-concept" (Baumeister 2017, p. 2). However it is unclear how this evaluative dimension is anything other than a facet of the qualitative definition already provided. The addition of a meta dimension of the self does not seem to be necessary at all. While the findings have been useful in the past, the ambiguity could potentially be resolved with a more comprehensive theory.

Due to the rise of growing polarization of many national political environments, particularly when it leads to extremism, it is critical to gain a better understanding of the mechanisms from which the phenomenon of self, and in a broader sense one's identity, emerges. Recently, researchers have focused on identifying possible mechanisms behind individual susceptibility to both an extremist shift in beliefs (Goldman & Hogg, 2016; Hales & Williams, 2018) and a degradation of trust in established institutions (Kaasa & Andriani, 2021; Poon, 2020). The experiments have typically been framed as seeking to diminish an individual's self-concept to measure the resulting change in attitudes and behaviors (Luchies et al., 2010; Stowers & Durm, 1996; Wirth & Wasselmann, 2018).

The experiments utilizing the self-concept paradigm have been productive in investigating feelings of ostracism and its impact on belief (Wirth & Wesselmann 2018, Stowers & Durm 1996). Alternatively, however, self-esteem has been used as effectively the same measure (Cichocka et al., 2015). As stated above, both self-concept and self-esteem are cumbersome in that they propose a meta self-evaluative dimension. This can be avoided however if one focuses on the term identity, defined as the ordered set of beliefs one maintains, rather than the notion of self. This is inclusive of beliefs one maintains about themselves. This way we can characterize the reaction to any challenge to a given belief as being proportional to the perception of the threat to that belief, and we can do this without appealing to the dynamics of self which is comparatively under-defined.

There is plenty of literature demonstrating an individual's desire to protect their conceptions of themselves (Chen et al., 2010; Derks et al., 2007; Kurzban, 2011; Williams, 2021) as well as their desire to protect their conceptions of the groups they affiliate with (Carmines et al, 2016; Claassen & Ensley 2017; Kahan, 2012). Both of these desires can be reframed as the response to the threats to beliefs regarding prioritized beliefs and the models they represent. As such, an individual's protection of an affiliated group is suggestive of the fact that individuals incorporate their conception of a group into their own identities (Mason & Wronski, 2018). The defense of a group therefore could also be defined as a function protecting one's identity. Here, groups are viewed as any set of individuals who perceive themselves to maintain a convergent set of beliefs. In this view, every individual maintains

their own model of the set of individuals that comprise the group and the sets of beliefs that the set of individuals maintains.

The model proposed in this paper is compatible with the view that self-concept lies at the center of information processing, and broadly of identity protection, but is a reframing of the results within a different theoretical structure. While successful in generating predictions and a variety of models, the literature on self (Hattie, 2014; Owens & Samblanet, 2013) and identity (Maalouf, 2011; Woodward, 2003) is cluttered with terminology. The two terms themselves are indistinguishable, despite attempts to separate them (Oyserman et al., 2012; Pilarska, 2016), and have led to sets of parallel, mutually coherent, research programs. Though distinctions like Tajfel (1981) proposed[2] were useful in constructing earlier research programs, this is slowly falling out of favor. Contemporary work points to a significant shift towards underlying mechanisms rather than appeals to self, self-esteem, self-concept, and identity protection (FeldmanHall & Shenhav, 2019; Hogg, 2020; Williams 2021).

1.2 Epistemic peership

Researchers within both the social and cognitive domains investigate the varying degrees of impact that experienced social phenomena have on an individual's perception of themselves and others around them. The combined findings within the two domains indicate that one's perception of their social environment, and their place in it, has the capacity to impact their networks of

[2] "Social identity is that part of an individual's self-concept which derives from his knowledge of his membership in a social group together with the value and emotional significance attached to that group membership" (Tajfel 1981, p. 255)

belief and their tendencies towards certain behaviors (Charness et al., 2017; John, 2017; Stahl & van Prooijen, 2018). Placing the maintenance of one's self-concept[3], or identity protection, as the core of information processing has provided valuable insight. It has allowed researchers to consider the mutability of attributions of epistemic validity and the resulting implications for one's trust in individuals, groups, and institutions (Plohl & Bojan, 2020; Van Prooijen et al., 2020). Here the attribution of epistemic validity to a source is defined as one's belief in the accuracy or truthfulness of the claims made by that source. Therefore, one's support of particular sources of information, meaning the sources they attribute epistemic validity to, are beliefs which motivate the assessment of further claims from that source (Kaasa & Andriani, 2021; Knobloch-Westerwick et al., 2017; John, 2017).

The use of the term epistemic peer in this thesis refers to any set of individuals who maintain a convergent set of beliefs within a given domain[4]. Note that this is a departure from the way the term is used in philosophy where there are normative assessments of qualifications involved with the term. Here the term is used to highlight the critical role one's affiliations play in their assessment of the validity of sources. The attribution of epistemic validity to a source is defined as one's belief in the accuracy or truthfulness of the claims made by that source. Therefore, one's support of particular sources of information, meaning the sources they attribute epistemic validity to, are beliefs

[3] Note that in the proposed model this is reframed as the resolution of uncertainty caused by threats to beliefs, with the generated uncertainty increasing proportionally with the prioritization of the belief being threatened. The assumption being that beliefs regarding oneself are highly prioritized in that they relate most immediately to self preservation.
[4] The trust in a cardiologist for a cardiac diagnosis may extend to trust in other medical opinions they may have, but may not extend to trusting their opinions on quantum computing.

which motivate the assessment of further claims from that source (Charness et al., 2017). This interpretation attempts to describe the mechanisms driving source validity and trust by placing beliefs at the center of attributions epistemic validity, thereby making room for a potential explanatory model that focuses on the agent in a system and their beliefs relevant to a given context.

Individuals who attribute validity to the same sources, and by extension the same information, are epistemic peers. Epistemic peership, as a term, is preferred over "group" because it defines a relationship between individuals in terms of the convergence of their beliefs. The term is used as a subjective, continuous measure of belief convergence rather than a binary designation[5]. Perceptions of the beliefs of others are critical to an individual's attributions of epistemic validity in that they impact which sources an individual believes are trustworthy. Individuals who maintain convergent beliefs, here defined as epistemic peers, maintain convergent sets of beliefs regarding the validity of sources by proxy of their mutual epistemic peership with those sources. Findings also demonstrate that individuals who share beliefs also regard one another as trustworthy sources (Brodbeck, 2009; De Dreu et al., 2008; Funkhouser, 2020; Shah et al., 1998).

Simply, the convergent subset of beliefs maintained by a set of individuals will inform the trustworthiness of particular sources or the accuracy of a particular set of information. Assessments of validity can go either from a source to their information or from information to its source. Such beliefs regarding the attributions of epistemic validity of information from particular

[5] Further implications of this are discussed in Chapter 3, Section 2.

sources serve as markers of prototypicality for members of a group, meaning that individuals can pass judgments regarding the prototypicality of other group members based on whether or not they view the same sources as trustworthy (De Dreu et al., 2008; Knippenberg et al, 1994). Prototypicality, however, is mutable in that it is a convergence of the perception of information, meaning individual displays of beliefs and behaviors can shift the prevailing perception of information by the individuals in a group. The most concerning implication of this is that the extreme beliefs of a subset of individuals has the potential to radicalize the larger set if their beliefs go unchallenged by the majority (Goldman & Hogg, 2016; Harel et al., 2020).

Groups generate a shared epistemic reality (Echterhoff & Higgins, 2017) which satisfies a need for closure (Shah et al., 1998). Hoffman et al. (2020) illuminated a critical motivation for belief convergence by demonstrating that the perception of shared beliefs and group membership can improve general well being. Tangential to this are the notions of normative and informational influences (Deutsch & Gerard, 1955; Toelch & Dolan, 2015) which could be reinterpreted as the shift in the beliefs relevant to the particular context. Either influence could be the result of updating relevant beliefs to resolve the uncertainty generated by a model that was an inaccurate representation of one's environment, regardless of whether the uncertainty was due to the inaccuracy of a concept or an affiliation one has prioritized. Further work in this domain could shed light on the emergence of extreme subgroups and the role of this phenomenon in political environments. Ultimately, a better understanding of the dynamics motivating epistemic peership can potentially

lead to the construction of interventions that can de-escalate intergroup conflict (Postmes et al., 2014).

1.3 Prototypicality & ostracism

The relationship between individuals in a group has been explored along multiple dimensions of individual and group motivations (Choi & Hogg, 2019; Echterhoff & Higgins, 2017; Forsyth, 2018). The individual-group dynamic is typically framed around the need to belong (Baumeister and Leary, 1995), the inverse of which are the threats to an individual's prototypicality (Knippenberg et al., 1994) and the threat of ostracism (Williams, 2007). Taken together, the variety of literature represents the individual's motivation towards social inclusion. Theories like Sociometer (Leary, 2021) and Terror Management (Landau et al., 2007; Solomon et al., 2004) suggest an evolutionary explanation for the motivation, namely that natural selection favored those who recognized that groups offered the best chances for survival, thereby linking the threat of ostracism with the threat of death. Boyer et al. (2015) presents a coalitional psychology model that posits intergroup relations are mediated by the perception of threat, and propose coalitions serve as a means of mitigating the negative physiological consequences of stress. Additionally, Lieberman & Eisenberger (2006) provide a social neuroscience account of the need to belong by presenting evidence linking the drive for social inclusion with the drive for self-preservation. Further research by Eisenberger et al. (2011), investigating the neural dynamics underlying fluctuations in self-esteem, demonstrates a link between the neural correlates of social ostracism and those of physical pain.

The feelings of ostracism have also been further explored through the prototypicality literature, where it has been demonstrated that uncertainty can be further linked to social identity by manipulating an individual's conception of their place in a group (Choi & Hogg, 2019; Wagoner & Hogg, 2016). Uncertainty of one's position in a group can lead individuals to engage in what they perceive to be prototypical behavior in order to gain better standing within their group (Goldman & Hogg, 2016; Hohman et al., 2017). However, what is believed to be prototypical behavior by those who perceive their position in a group to be threatened has the tendency to be a more extreme version of group beliefs. This can ultimately result in the engagement with more extreme beliefs and behaviors in an attempt to regain ingroup favorability (Gaertner et al., 2008; Hales & Williams, 2018). Conversely, McCulloh (2013) found that those who hold informal power are more free to speak their minds without fear of repercussion. The question remains as to how the desire for prototypicality (alternatively, the desire to resolve uncertainty) is acted upon.

Uncertainty can both motivate individuals to engage in behavior (including violence) that will signal their prototypicality to other group members (Goldman & Hogg, 2016; Hales & Williams, 2018), as well as increase ingroup bias. Studies show that even the invocation of a group can prime individuals to engage in biases that serve preexisting beliefs (De Dreu et al., 2008; Knippenberg et al., 1994; Wagoner et al., 2017). It was also found that when a group is expanded to include new members, those individuals who view the new members as atypical report an increased sense of uncertainty and general feelings of threat (Danbold & Huo, 2017). The BDM predicts that this

uncertainty is caused by a threat to one's existing beliefs regarding their affiliation with a group, meaning the threat to their model of a given group as an environment. It was also found that individuals report an increased sense of uncertainty, driven by general feelings of threat to their existing beliefs regarding the group, when their group is expanded to include atypical members (Danbold & Huo, 2017). Reciprocally, perceived prototypicality can influence psychological wellbeing, and a wider, more stable social network can improve satisfaction with life while decreasing feelings of anxiety and stress (Hoffmann et al., 2020). Altogether, this presents a need to address questions surrounding individual needs for inclusion to better understand the emergence of extreme beliefs and violent behavior.

1.4 Unfreezing

Unfreezing is one of the metrics used in the literature as a proxy for one's openness to opinion change. Typically, the variable claims to measure the degree to which individuals are willing to update their beliefs (Hameiri et al., 2014; Rico & Barreto 2021). Unfreezing provides a useful conceptual framework for measuring the impact of a particular change in belief relative to other adjacent beliefs along a spectrum. First introduced by Kurt Lewin (1947) as a proposed model of cognitive change through which societal change is possible, unfreezing is best summarized as follows:

> On the individual psychological level, unfreezing usually begins with the appearance of a new idea that is inconsistent with already-held attitudes and causes psychological tension that triggers intrapersonal conflict. This, in turn, may stimulate people to move from their basic positions and look for alternatives. (Hameiri et al., 2014. p. 165)

Typically, unfreezing has been used as the inverse to the extremification of beliefs, such that any self-reported changes in one's beliefs can be either interpreted as the relative strengthening of a belief, if it moves in the direction of a given extreme, or unfreezing of it, if the change is in the direction away from the nearest extreme. Worth noting however is that this distinction is relative, in that individuals who are either left or right of center might not view shifts toward a centrist position as an unfreezing but rather a shift toward a more extreme position.

Recently, the concept of unfreezing has been effectively applied by various groups researching lasting two-party conflicts such as the Colombia-FARC (Rico & Barreto, 2021) and Israel-Palestine (Hameiri et al., 2017; Harel et al., 2020) conflicts. The groups have been successful in demonstrating that certain social and cognitive factors can impact the strength of a belief and cause a shift in an opinion. The unfreezing literature is particularly useful in this context because, unlike most other work done to understand the mechanisms behind belief change, unfreezing experiments typically attempt to lessen the degree of extremism in the beliefs (Gayer et al., 2009; Rico & Barreto, 2021).

A recent experimental construct meant to use uncertainty to deescalate extremism has been the paradoxical thinking paradigm (Hameiri et al., 2018; Hameiri et al., 2017; Harel et al., 2020). Paradoxical thinking presents individuals with particular beliefs that are more extreme than ones they are willing to affiliate with in an attempt to increase individual openness to shifting beliefs further from those presented beliefs. It has been particularly effective in

its presentation of radical, reductionist versions of the extreme beliefs common to individuals on the far right of the Israeli-Palestinian conflict to motivate individuals to distance themselves from the far right. Paradoxical thinking may be inducing feelings of potential ostracism from, or atypicality with, affiliations that an individual prioritizes (Hameiri et al., 2019; Kruglanski, 2013). For example, the findings demonstrate that an individual can be motivated against affiliation with group P, if affiliation with P threatens affiliation with a more prioritized group Q.

Research on increasing participant openness has thus far been successful in demonstrating how the uncertainty caused by identity threat can be useful in initiating social change, particularly in dealing with conflict resolution (Hameiri et al. 2019). Unfreezing, while not a radical change in belief, is still capturing the dynamics of the belief hierarchy (further detailed below) which reflect an updating in the individual's identification with certain sets of ideas. The current project attempts to manipulate an individual's perceived prototypicality by creating an illusory group of epistemic peers predicated on participant responses regarding their prioritization of certain topics. In this regard, this work differs from paradoxical thinking research where participants are confronted with more extreme versions of their own opinions.

1.5 Uncertainty

Uncertainty plays a central role in much of the literature regarding the dynamics of individual beliefs and, by extension, behaviors. The degree of incongruence between what is believed and what is perceived is the measure of

uncertainty. Therefore, uncertainty is effectively the indicator of discrepancies between the content of beliefs and the content of perception. While uncertainty has a theoretically quantitative value, despite being incalculable in practice, it also produces qualitative states in the agent (Barrett, 2017; Boyer et al., 2015). Affective states may be the qualitative byproducts of uncertainty and may have evolved as a physiological reward mechanism for reinforcing the avoidance of potential threats (Friston, Boyer et al., 2015). In this view, the agent operates with the imperative to minimize uncertainty by updating their beliefs to minimize errors between incoming and predicted stimuli resulting from perception.

There are a number of models that place uncertainty as a core element. Each model uses its own set of terms, but these are compatible. For example, both the need for control (Leotti et al, 2010) and the need to belong (Leary, 2021) are merely the inverse of uncertainty in that they both reflect a fear of ostracism. As outlined above, the drive for self-preservation seems to have prioritized group affiliations to best mitigate threats from one's environment (Landau et al., 2007; Lieberman & Eisenberger 2006). This would mean individuals interpret social bonds as mitigators of the uncertainty generated by the threat of ostracism, which is taken to have evolved as a proxy for threats to self-preservation. Accordingly, much of the literature stakes the claim that uncertainty is the driving mechanism behind the feelings of ostracism and loss of control (Chen et al., 2010; Guzel & Sahin, 2017; Hales & Williams, 2018; Hartgerink et al., 2015), as well as the correlated phenomena of illusory pattern

perception (Whitson et al., 2008) and conspiratorial ideation (Hogg, 2020; Poon et al., 2020; Van der Wal et al., 2018).

Illusory pattern perception seems to facilitate conspiratorial ideation (Van Prooijen et al., 2018; Van Prooijen & Van Dijk, 2014). Individuals experiencing uncertainty tend to reach for simple solutions to resolve their uncertainties, which leads to the perception of patterns which don't exist in order to construct conspiracy theories (Van Prooijen et al., 2018; Van Prooijen & Van Dijk, 2014). Not only does conspiratorial ideation lead to dangerous beliefs like those maintained by anti-vaxxers, but it also reduces civic engagement (Jolley & Douglas, 2014; Abalakina-Paap et al., 1999) potentially further ostracizing portions of the population and leading to a cycle that pushes individuals towards increasingly extreme positions.

Two theories that likewise describe similar mechanisms are the uncertainty-identity theory, or UIT, (Hogg & Aldeman, 2013) and the self-affirmation theory, or SAT, (Sherman & Cohen 2006). Both theories look to place the protection of one's model of oneself as the mechanism driving the dynamics between individuals and groups. The difference is primarily semantic, in that UIT frames the individual's primary social motivation as being uncertainty mitigation, whereas the SAT defines it as the maintenance of self-integrity. While both deal with the dynamics of individual behavior in groups, effectively producing similar hypotheses, UIT further centralizes uncertainty as the core driving mechanism for extreme beliefs and behaviors (Hogg, 2020). I argue that the proposed model, BDM, provides a more economical solution to explain the same motivations by describing the self, as

the sum of the beliefs one maintains. The focus on belief dynamics, meaning the mutability of both the prioritization and the content of beliefs, gives us a more coherent picture of social cognition in that we can engage with the beliefs individuals maintain as propositions.

In the proposed BDM, the degree of uncertainty an individual experiences falls along a gradient. The interactions with one's environment either mitigate (via *validation* of one's beliefs) or increase (via *threats* to one's beliefs) the uncertainty experienced by an individual[6]. This view allows the model to also account for affective states as phenomena that emerge in response to the mitigation or exacerbation of uncertainty one is experiencing. Specifically, what might be considered "positive" affective states emerge as uncertainty decreases while "negative" affective states emerge as uncertainty increases. This draws from work on the neuroscience of affective states from Barrett (2017) and Hoemann et al. (2017), which take a constructionist view of emotions. Barrett and Satpute (2019) assert that the variety of emotions are not innate reactions to the world, but a result of social learning. They suggest affective states (internal sensations in general) may be better understood through predictive coding accounts of the brain, particularly inferences made regarding the sources of prediction errors (described above as belief-perception incongruence).

Multiple models have been put forward which link uncertainty to anxiety (Grupe & Nitschke, 2013; Hirsh et al, 2012). Here anxiety resulting

[6] In the active inference model this is referred to as surprise. For a breakdown of the Bayesian modeling describing the minimization of surprise (as free energy) see Friston et al., 2018.

from uncertainty is a function of the probability distributions maintained by an agent which represent predictions regarding its environment. The predictions are the likelihoods of successions of internal states, meaning the internal states at T_0 leads to an anticipation of possible states at T_1. The incongruence, between one's beliefs regarding likely states and what they perceive as actual states, is the measure of uncertainty. Such models are often referred to as Bayesian brain models because the updating of the probabilities (beliefs) is done with respect to the available information in a system. These models assert that an agent maintains likelihoods of its own internal states generated via sensations which reflect external states (Clark, 2013; Friston et al., 2015; Friston et al., 2018). As Barrett and Satpute (2019) put it:

> Recent predictive coding (a.k.a. active inference, belief propagation) models suggest that the brain functions as Bayesian filters for incoming sensory input, guiding action and constructing perception. Past experiences are reconstructed as partial neural patterns that serve as prediction signals (also known as "top-down" or "feedback" signals, and more recently as "forward" models) to continuously anticipate events in the sensory environment. (p.14)

In the active inference model, which attempts to build a link between information theoretic conceptions of entropy with the thermodynamic reduction of free energy, the mitigation of uncertainty as defined above is defined as the minimization of free energy (Friston, 2010; Friston et al., 2018).

Recent experiments also lend evidence to the greater affective saliency of uncertainty of outcome over known risk by measuring participant skin responses (Feldmanhall et al, 2016), lending further credibility to the link between affective states and individual motivations to resolve uncertainty (Feldmenhall et al., 2019). This view has the potential to extend into other

domains as well. For example, when dealing with affective states in the economics literature, emotion has long been referred to as a measure of relevance (Frijda et al, 1989; Phelps, 2009). It seems however that individual differences in skin response measurements can be interpreted as greater degrees of uncertainty. The economic metric of relevance might in these cases be interpreted as a measure of the prioritization of the underlying beliefs regarding the content. Investigations across these domains could lead to a better understanding of how aggression, anxiety, and fear emerge in response to social uncertainty (Gaertner et al., 2008; Grupe & Nitschke, 2013; Leary et al., 2003), and how strategies like motivated reasoning (Kunda, 1990) help us mitigate the impact of negative affective phenomena.

1.6 Motivated reasoning

Beliefs regarding one's social affiliations mitigate the uncertainty caused by ostracism. One is therefore incentivised to maintain a favorable model of themselves and project that model in social situations to mitigate the threat of ostracism, which in social animals like humans is perceived to be a threat to self preservation (Cacioppo & Cacioppo, 2016; Williams, 2007; Williams & Zadro, 2005). As such, uncertainty is observed to lead to what the social and cognitive psychology literatures refer to as Social Identity Protection and Identity-Protective Cognition respectively. The engagement of identity protection seems to play a role in an individual's attributions of epistemic validity, which subsequently impacts an individual's conception of truth (Kahan 2016a; Kahan 2016b; Jost et al., 2013). Whether an individual trusts the source

of a piece of information will determine whether or not they use the information to update their beliefs (Cook & Lewandowsky, 2016; Kraft et al., 2015). Cook and Lewandowsky (2016) demonstrated that if the strength of a belief outweighs the trust in a source of information, the individual will be driven to doubt the veracity of the source. On the other hand, if the trust in a source outweighs the strength of a belief, the information will be utilized as evidence to update the belief. While Cook & Lewandowsky proposed a new Bayesian model which reflected the outcomes of their manipulations, the notion that prior beliefs impact the interpretation of evidence is neither new (Florence, 1975) nor novel (Kraft et al., 2015; Miton & Mercier, 2015; Prike et al., 2018).

Explaining identity protection (or self-concept maintenance) and information processing through the lens of uncertainty mitigation naturally implicates motivated reasoning. Findings across the motivated reasoning literature, typically dealing with politicized topics, demonstrate that individuals are motivated to process information in a way that is congruent with their preexisting beliefs (Bayes & Druckman, 2021; Druckman & McGrath, 2019; Jost et al., 2013; Stanley et al., 2020). Individuals are typically motivated to reject an opposing view even in the face of counter evidence (Kraft et al., 2015; Prike et al., 2018; Stanley et al., 2020), a phenomenon Stanley et al. (2020) call prior-belief bias. Knobloch-Westerwick et al. (2017) also demonstrated individuals will justify questionable evidence and behaviors of those with whom they share convergent beliefs. This creates echo chambers, particularly in political contexts where, as Frimer (2017) showed, individuals even refuse financial incentives just to be exposed to well known opposing views.

Politicization often begins with the promotion of individual issues and building narratives of fear around the opposing positions on those issues (Gore, 2004; Molder, 2011; Pearlman, 2016; Tang, 2008). This allows not only for groups to distinguish themselves from one another but also to build messaging around fear of the opposing side. Politicization can therefore be viewed as a strategy for engaging the uncertainty of individuals in order to further motivate their behaviors and processing of information towards a particular end. The idea that motivated reasoning is moderated by, if not a product of, uncertainty mitigation has been recently explored by multiple authors (Carpenter, 2018; Han & Kim, 2020; Hogg, 2020; Nasr, 2021). Han and Kim (2020) build on Hogg's uncertainty-identity theory whereas Carpenter (2018) relies on self-concept to make their cases. Both sets of authors reach the same conclusion by relying on adjacent terminology but this is not the main issue. What is missing from both accounts is an appeal to what the self or identity is composed of. Here the BDM goes one step further, asserting that the self and identity are two words for the same phenomenon that emerges from the set of beliefs maintained by the individual. This focuses the attention on an individual's context-relevant prior beliefs and views the engagement with information in light of those prior beliefs, as bias, to be the default mode of reasoning.

There has been research suggesting the success of mechanistic explanations as interventions to increase receptivity to consensus among researchers in the field (Caddick & Feist, 2021; Hart & Nisbet, 2012; Kahan, 2016a). However, the same studies present motivated reasoning, in the form of

political affiliations and ideologies, as caveats to the efficacy of those interventions (Zummo et al., 2020). Presenting a larger framework, Kahan (2016a; 2016b) builds on the *politically motivated reasoning paradigm* introduced by Jost et al. (2013), which seeks to provide a conceptual model of how individuals unconsciously assess information in a way that conforms with their prior beliefs. Kahan presents the model to stand in contrast with Bayesian style truth-seeking models, arguing that "The truth-independent goal of 'politically motivated reasoning' is identity protection" (Kahan 2016a, p.3). Fundamentally, Kahan argues, "Politically motivated reasoning is not truth convergent" (Kahan 2016b, p. 2).

Perception of climate science is an often addressed topic in the motivated reasoning literature. Climate change serves as a prototypical example of when politicization of a topic can extend beyond epistemic commitments to metaphysical relativism: in one version of reality human activity is causing the earth to warm to levels dangerous for longterm human survival, and in another the earth stands unchanged while corporations and scientists conspire to profit off of a fabrication. Findings across research groups have demonstrated that the politicization of climate change in the United States has stunted popular understanding of climate science (Bayes & Druckman, 2019; Bolson & Druckman, 2018; Cook & Lewandowsky, 2016) and that, crucially, this impacts which sources individuals are willing to accept corrective information from (Benegal & Scruggs, 2018). Additionally, Kukkonen et al. (2017) outline how the prioritization of differing beliefs within a broader context treated as a single

issue, climate change in this case, even lead to informal coalitions that have their own media cycles.

In their work on perceptions of climate science, Druckman & McGrath (2019) present a Bayesian model of belief updating in light of new information that takes into account prior beliefs. Their model, similar to the one presented by Cook and Lewandowsky (2016), allows for goals, like identity-protection, to be represented as weights for prior beliefs. In the proposed BDM, all goals and desires are interpreted as beliefs, and the particular weight attributed to each belief represents its place in the hierarchy. The prioritization of a belief, as in Druckman & McGrath's model, can be represented as a difference in weights in favor of the prior belief during the updating process when encountering information. This allows the BDM to draw a link between the cognitive literature on motivated reasoning (Connor et al., 2020; Cook & Lewandowsky, 2016; Druckman & McGrath, 2019) and the neuroscience literature on the prioritization of beliefs (Friston et al., 2017; Kuchling et al., 2019; Pezzulo et al., 2018).

1.7 Prioritization

Beliefs differ in both their content as well as their prioritization[7]. Though this distinction is pragmatic rather than a claim about the physical properties of beliefs, it helps to better understand how individuals can differ so greatly in their motivations. The BDM predicts that the more prioritized a belief becomes the more influence they have over the prioritization of future beliefs. This

[7] Prioritization has also been defined along a central-peripheral dimension as early as by Eagly (1967) and Rokeach (1968), with the former claiming that beliefs regarding the self-concept are the ones that tend to be more centralized.

makes beliefs which are prioritized past a certain (relative) threshold very difficult to deprioritize because any contrasting information will be interpreted in a way that serves the most prioritized beliefs. Theories like identity-protective cognition, terror management theory, and theories based on self-esteem all suggest that beliefs which are maintained to mitigate threats to self-preservation are the most prioritized (Baumeister, 1997; Burke et al, 2010; Derks et al., 2007). This supports a relative view of rationality, where what can be deemed to be rational is that which serves the broader motivations of the individual, whatever they may be.

Aside from the common sense notion that individuals prioritize certain beliefs over others, there is both cognitive (Connor et al., 2020; Cook & Lewandowsky, 2016; Druckman & McGrath, 2019) and neurological (Friston et al., 2017; Kuchling et al., 2019; Pezzulo et al., 2018) support for belief prioritization. Brodbeck et al. (2007) demonstrated that an individual's degree of identification with a traditional political wing is a reliable predictor of the strength of the engaged bias toward opposing information. This suggests the bias toward the maintenance of a particular belief is a function of the prioritization of the belief that is being threatened[8]. Moreover, the hierarchy of beliefs that an individual maintains has been shown to be contingent upon how protective an individual feels the need to be of themselves. These feelings are moderated by the prioritization of the triggered affiliation (Goldman & Hogg, 2016; Hogg et al., 2017). Meaning that the more prioritized a threatened belief

[8] Brandenburger & Dekel (1993) and Seuken & Zilberstein (2008) offer game theoretic and decision theoretic algorithms respectively to model uncertainty in multiple agent interactions accounting for belief hierarchies.

is, the more defensive an individual will get in response because the generated uncertainty will be greater. This is relevant for the BDM because in it identity is defined as the hierarchy of beliefs one maintains, and the degree to which one feels the need to protect their conceptions of themselves will therefore be a function of the degree of uncertainty one experiences.

The prioritization of a belief is indicative of the magnitude of the threat it mitigates. Therefore, the more prioritized a belief is the more uncertainty a challenge to it is expected to generate. That social inclusion is linked with mitigating threats to self-preservation (Leary & Acosta, 2018; Solomon et al., 2004; Landau et al., 2007) may be why ostracism and related threats evoke fear responses like violence and aggression (Gaertner et al., 2008; Leary et al., 2003; Warburton et al., 2006). For example, Nisbett and Cohen (2018) summarize decades of research on the southern United States and the collective findings on how a culture of honor can cultivate a heightened sensitivity to affronts and lead not only to the promulgation of violence but a greater acceptance of it as a response. This suggests one's innate strategy for acceptance may be linked to the culture they are raised in, meaning enculturation can be viewed as the prioritization of what one perceives to be the beliefs of their social environments. Therefore, the available strategies of survival within a culture might be determined by whatever emerges from the convergence of the hierarchies of the individuals maintaining that culture.

Social dynamics are not navigated in a vacuum. As individuals update their beliefs, they experience shifts in both their perspectives of their environments and the reactions their environments have to them. This may be

one of the mechanisms playing a role in the extremification of beliefs, in that the more an individual's beliefs become extreme and deviate from a norm, the more interactions they might ostensibly have with individuals who disagree. This cycle may in turn cause beliefs to become more extreme and lead individuals to seek others who confirm their beliefs. As demonstrated by Asch (1956), and further explored by McCulloh (2013) and Hodges (2017), individuals have a greater affinity for conformity, effectively translating to an affinity for those with whom one shares beliefs. This may be what generates today's increasingly polarized environment within which some individuals are increasingly ostracized from mainstream communities and find affirmation from communities defined by fringe beliefs. Taking the maintenance of social relationships as a foundational motivation, Stern (2021) outlines how variations in the prioritization of certain relational goals across groups can lead to greater polarization. Stern also details how the promotion of cross-ideological similarities could be critical in depolarization efforts.

As talk of an increase in general uncertainty enters the public discourse, fear based messaging combined with a greater propensity to appeal to illusory patterns for explanations could see more political parties adopting conspiracy theories into their narratives. Enders and Smallpage (2018) present data on how partisanship, particularly Republicanism in the United States, is tied to an increased support for conspiracy theories. This is in line with findings of how polarization leads to further polarization, causing a runaway effect (Axelrod et al, 2021; Baldassarri & Page, 2021), similar to how ostracism can lead to aggression which can in turn lead to more ostracism (McDougall et al., 2001;

Warburton et al., 2006; Williams, 2007). This may provide a key to understanding the motivations behind the adoption of successively extreme beliefs and behaviors, as well as means of mitigating the fallout of runaway polarization.

1.8 Belief

A critical shift being proposed in this paper is that the findings outlined above be interpreted through the lens of the dynamics of belief. More specifically, the dynamics of belief and behavior generated by the interactions between the information an individual perceives and the network of beliefs they maintain. As such, beliefs and the perception of information are inter-determinant. An increasing number of studies are suggesting that the beliefs an individual maintains are contingent upon their epistemic positions (Kaasa & Andriani, 2021; John, 2017), with the epistemic positions themselves being beliefs regarding the validity of sources. If the prioritization and content of a belief an individual maintains is contingent upon their other beliefs, individuals will be motivated to interpret information in a manner that best mitigates any uncertainty caused by having those beliefs challenged.

A number of different models can be interpreted to suggest a similar conclusion regarding attributions of epistemic validity and the motivations underlying the interpretation of information. One such model is the identity-protective cognition thesis (ICT), which centralizes an individual's conception of themselves as a critical mitigating factor of information processing (Kahan 2017; Kahan et al., 2017). The ICT suggests that the way

individuals interpret information reflects the beliefs of the groups they are affiliated with and maintains their perceived value of those groups. This is in line with various literature on the self and identity, particularly group identities, conformity, and the drive for prototypicality (Goldman & Hogg, 2016; Hogg et al., 2017). Hogg et al. (2017) argue that once the individual engages with content that is potentially threatening to their identity, they engage in identity-protective behavior that triggers a biased interpretation of the content.

At the heart of the ICT is an information processing bias that favors interpretations of information which best suit the individual's conception of themselves (Kahan, 2017; Kahan et al., 2007). In practice, this reflects the individual tendency to both devalue information that one might find threatening to their self-concept, and attribute greater validity to information which bolsters it. The activation of the identity protective bias can also lead an individual to alter their beliefs in order to ensure they remain, what they perceive to be, prototypical members of a group (Knobloch-Westerwick et al., 2017). The identity protective bias can also account for other extensions of cognitive bias such as myside bias (Stanovich et al., 2013) and related confirmation biases (Wagoner & Hogg, 2016), as well as motivated cognition and numeracy, social biases like ingroup and outgroup biases (Knobloch-Westerwick et al., 2017; Wagoner et al., 2017), and phenomena like the Dunning-Kruger effect[9] and illusory pattern perception (Whitson & Galinsky, 2008).

[9] In that, in line with ICT, individuals are motivated to believe they are more capable than they actually are, see: Hornsey et al., 2021.

Because social relationships are important in mitigating threats to self preservation, the perception of group dynamics plays a critical role in the dynamics of belief and resulting behaviors. There are multiple studies demonstrating that an individual's interpretation of information is moderated by their affiliations to particular groups. Namely that affiliations are used to filter through information and the rapid assessment of data (Brodbeck, 2007; Knobloch-Westerwick et al, 2017), suggesting the group heuristic is critical to increasing polarization and constructing uninformed beliefs (Baker, 2009; Earle & Hodson, 2018). As such, individual beliefs regarding the beliefs of other sets of individuals serve as epistemic lenses through which further information is viewed. Perhaps more concerningly, some individuals will even go so far as to refuse incentives such as a small monetary reward offered in exchange for viewing information that conflicts with the narratives of the groups they affiliate with (Frimer et al, 2017). This suggests that polarization may be a function of individuals prioritizing beliefs to the point that opposing beliefs are uncomfortable because they pose a threat (Goldman et al., 2016; Hart et al, 2009; Hohman et al., 2017).

Individuals also seem motivated to undervalue information that a fellow group member shares if it is inconsistent with standing group beliefs, and spend more time arguing in favor of shared beliefs than discussing dissenting opinions (Baker, 2009; Miton & Mercer, 2015). This type of affiliation-centric epistemology is often referred to as tribalism, the evolution of which is often linked to the prioritization of social inclusion over other potential motivations (Clark & Winegard, 2020; Cornwell et al., 2019). Therefore, discussing what

are perceived to be shared beliefs can be viewed as an exercise in mitigating uncertainty.

Taken together, these studies suggest that individual beliefs regarding epistemic validity impact further beliefs. As such, beliefs can be interpreted as maintained propositions which impact the evaluation of further propositions. Models like ICT, Social Identity Threat, Sociometer Theory, and Terror Management Theory, together suggest that one's motivation to protect themselves results in perceiving ostracism and atypicality as exacerbators of uncertainty, and inclusion and acceptance as mitigators of it. Therefore, conforming to and defending group beliefs are likely to be rooted in the mitigation of social (and ultimately physical) uncertainty. As it seems that uncertainty is generated when beliefs are challenged, it can be argued therefore that beliefs are prioritized with respect to their capacity to resolve uncertainty. In this sense, a belief constructed to mitigate uncertainty is like a dam constructed to moderate the flow of water. There are constant pressures on beliefs as there are on dams, and if the dam is threatening collapse it is typically reinforced or repaired because if it collapses the crisis is proportional to the amount of water, or uncertainty, it is constructed to handle.

1.9 Belief dynamics model

The BDM views an agent as a system that maintains a dynamic network of beliefs. In the model, beliefs are constructed and maintained to reflect information the agent perceives from their environment. As an individual navigates an environment, their beliefs are engaged to interpret information and

the beliefs in turn impact the perception of information. The resulting perceptions are used to update the existing beliefs. The more congruent the network of beliefs is with what is perceived, the less uncertainty the individual experiences.

The belief dynamics model views identity as the phenomena that emerges from the maintenance of a mutable hierarchy of beliefs. Effectively, identity is viewed as the sum of strategies an individual uses to navigate their environment. This interpretation of identity is an extension of an interpretation of beliefs that corresponds to the one put forward by the active inference model (Friston et al., 2016; Friston et al., 2017). The primary prediction of the model is that the network of beliefs an individual maintains is updated in accordance with what best mitigates threats to the most prioritized beliefs. For this reason the most prioritized beliefs, namely those directly relating to self preservation, are unlikely to be deprioritized because nearly all perceived information will be interpreted to serve those beliefs. This also means that beliefs are prioritized in relation to one another, with the most prioritized beliefs dictating the prioritization of beliefs within the network. This leads to a cascading effect on the network of beliefs in which the general purpose served by further beliefs is also self preservation.

In light of the findings outlined above, the belief dynamics model centers the drive to minimize uncertainty as the mechanism underlying belief change[10]. In predictive brain models such as active inference, which the BDM

[10] The model attempts to capture this by defining uncertainty as the incongruence between what is expected (as a function of the content of the relevant set of beliefs) and what is perceived by the individual (which nonetheless is also impacted by the relevant set of beliefs).

seeks to provide a potential psychological account for, the ability to predict one's external environment or the relationship between internal and external states (Friston, 2010, Friston et al., 2016). What is focused on here is the uncertainty of one's perception regarding their position within their social environment. In different paradigms this can have names like the threat of ostracism or exclusion, prototypicality threat, as well as social discomfort (Miller, 1995). However, more recent work in social cognition (FeldmanHall & Shenhav, 2019) and neuroscience (Christopoulos & Tobler, 2016) has sought to consolidate the various social pressures under the umbrella of uncertainty.

The BDM seeks to provide a unified model of identity that can generate new predictions in light of the findings across a particular range of topics in the social and cognitive sciences. Particularly it attempts to offer a novel means of explaining individual behavior by centering the perception of information, and the beliefs that motivate it. As such, it attempts to simplify and reinterpret much of the existing work without contradicting it.

CHAPTER 2

PRESENT STUDIES

In the three experiments below we test the viability of a particular component of the belief-dynamics model, namely whether or not atypicality with one's epistemic peer group can lead to a shift in one's beliefs. The first experiment is an initial test of the experimental setup. Experiments 2 and 3 attempt to remove potential confounds of the pilot without altering the intervention. To establish the illusion of one's epistemic peer group[11], participants were provided a set of statements covering a range of topics and were asked to indicate both how much they agreed with the statements and how important those topics were to their identities.

Being that the hierarchy of beliefs is at the core of the model, we predicted that the more central the topics used to generate the peer group were, the greater the agreement change would be. This prediction was intended as an extension of the findings outlined in Chapter 1 regarding changes in beliefs and behavior caused by threats of ostracism and atypicality with one's group. The goal was therefore to generate a virtual social context in which this was the case.

Within this virtual context, individuals identified their own epistemic peer groups by indicating the prioritization of certain beliefs to their own

[11] Groups tend to have shared beliefs. While the methodology in this paper is novel, the notion that group members are individuals with whom one shares beliefs are not (Brodbeck et al., 2007; Choi & Hogg, 2019; Coman & Hearst, 2015). Here the term *epistemic peer group*, as explained above, is in reference to individuals who are identified as those with whom one shares beliefs.

identities. The assumption was that if the presented beliefs were in fact central to an individual's identity, then the perception of deviation from an epistemic peer group should generate uncertainty due to prototypicality threat (Chen et al., 2010; Goldman & Hogg, 2016; Hogg & Aldeman, 2013). The uncertainty was expected to result in a drive to reduce the atypicality experienced by the participants (Choi & Hogg, 2019; Goldman & Hogg, 2016). Across all three experiments this was measured by a change in agreement with the statement regarding the manipulated topic.

2.1 Experiment 1

The first experiment was launched in May of 2021 and was an exploratory study meant to test the viability of both the new experimental task and the primary hypothesis that atypicality with epistemic peer groups could trigger belief change. The study centered around participants' self-identification of their peer groups by indicating endorsements of particular statements regarding topics central to contemporary public discourse (labeled *primary*). The manipulation used those endorsements to trigger a change in the endorsements of further statements which are less visible in the public discourse (labeled *secondary)*. The study used 5 *primary* topics to establish one's peer group and 5 *secondary* topics of which one was randomly selected for the manipulation. Participants rated their endorsement for all topics and they indicated how important these topics were for their identity. They were then shown false feedback on one of the two secondary topics, indicating a divergence in views in one of these secondary topics. Then they were given the option to re-indicate

their endorsement for these topics where there was not much convergence. We expected to see a positive correlation between the *prioritization* of the *primary* topics and the *agreement change* in the manipulated *secondary* topic.

2.1.1 Method

2.1.1.1 Participants

Participants were initially recruited from the Bogazici University psychology department and the experiment was later snowballed via social media. In total, 5,333 responses were gathered. Of those, 3,052 had to be excluded for multiple reasons: 2,481 because they didn't complete the experiment, 76 because they didn't agree to the consent form, 474 for failing the two attention check questions[12], and 21 for completion times that were extreme outliers (+/- 2 *SD*) (M_{Time} = 12m21s, *SD* = 5m33s). The resulting sample was of of 2,281 participants (*F* = 1,904 ; *M* = 288; *Non-binary* = 89) (M_{Age} = 20.1 years, *SD*=3.7), 102 of which were students from the Bogazici University psychology department.

2.1.1.2 Materials

Belief task: The belief task attempts to have individuals identify their own affiliated groups and use those responses to generate the illusion of divergence (in opinion) from those groups. The task utilizes a range of topics and splits them into two groups: *primary* and *secondary.* The primary topics are meant to capture topics which are more prioritized by those individuals whereas the

[12] e.g. "Please place the slider between 70-80" (the slider was a 100 point Likert scale)

secondary topics are meant to be more auxiliary to their interests (discussed in greater detail below).

Participants were initially introduced to the different topics (described below) a one to two sentence description of each. Participants are then given a statement below each topic (e.g. “The law should protect LGBT+ persons to live as they please.”) and asked to respond to a set of questions (phrased as statements) on that topic on a 0-100 point Likert scale with the following: “I agree with this statement” (in Experiments 1-3), “This is central to my identity” (in Experiments 1-3), and “I consider myself knowledgeable on this topic” in Experiments 2 & 3). 100 represents maximum agreement with the question as phrased (e.g. “I completely agree with this statement”), and 0 represents maximum disagreement (e.g. “I completely disagree with this statement”).

After answering the questions, participants were shown fabricated results regarding their responses relative to those of other participants. In the experimental group participants are shown a large deviation (50 points) from the mean of their group's opinion in one particular *secondary* topic (presented as participants who responded similarly to them on other questions). This deviation was displayed as a statement explicitly indicating their deviation of at least 50 points in their indicated level of agreement with the target topic (“Your agreement with the statement ‘...’ deviated more than 50 points from those who responded similarly on the following topics ‘...’”). Participants were then given the choice to re-indicate their level of agreement with the topic. They were also asked whether they would be open to receiving more information on the

manipulated topic to measure unfreezing (Frimer et al., 2017; Hameiri et al., 2017; Rico & Barreto 2021).

The five primary topics were used to generate the illusion of epistemic peers, and the secondary topic which participants responded was most prioritized was utilized as the manipulated item. The *primary* topics were: LGBT+, Climate Change, Gender Pay Gap, Abortion Rights, Animal Rights. The *secondary* topics were: Artificial Intelligence (as a threat), Social Media Regulation, GMOs, Capital Punishment, Pink Tax.

The ten topics were chosen through 3 rounds of Google Forms given to undergraduate psychology students (Survey 1, $N = 22$; Survey 2, $N = 30$; Survey 3, $N = 27$). The first survey had open-ended questions asking both for suggestions of topics which were important but not politicized, and topics about which their opinions might differ from those of their friends. 24 distinct topics were selected from the first survey. The second survey asked participants to categorize each of the 24 selected topics into what they felt was a primary or secondary concern for themselves and then primary or secondary concern for society (responses collected in binary, either primary or secondary). The third survey was to determine which 10 of the 15 most salient topics should be included and whether they should be categorized as the primary topics (used to generate the epistemic peer group) or secondary topics (used to test the manipulation). Participants could only rank 5 topics as *primary*, and 5 topics as *secondary* to force a choice and exclude topics which participants had no interest in.

Participants were asked to complete the twelve-item Intolerance of Uncertainty Scale, α = .85, (Carleton et al., 2007) using a 5 point response scale (1 = *not at all characteristic of me* to 5 = *entirely characteristic of me*). The scale was translated into Turkish from its original english.

Participants were asked to complete the ten-item Need to Belong, α = .81, (Leary et al., 2013) using a 5 point response scale (1 = *strongly disagree* to 5 = *strongly agree*). The scale was translated into Turkish from its original english.

Participants were asked to complete the 7-item Imposterism Scale, α = .87, (Leary et al., 2000) using a 5 point response scale (1 = *strongly disagree* to 5 = *strongly agree*). All scales were translated from their original English into Turkish.

Translations were done by 4 undergraduate psychology students at Bogazici University who are fluent in both Turkish (native) and English. Two graduate students helped resolve any discrepancies between versions of translations between the students.

2.1.1.3 Procedure

After agreeing to the consent form, participants were asked to provide their age and gender. Participants were then received the brief introductions to the topic (Appendix A) then completed the *belief task*, followed by an attention check question, and then completed three scales: the Intolerance of Uncertainty Scale, Need to Belong, Imposterism Scale. Lastly, participants were asked whether the manipulation was believable and what they thought of the experiment. Upon

conclusion of the experiment, participants were given a debrief informing them that the results they were shown were fabricated and what purpose the manipulation served.

All materials (topic descriptions, topic names, topic statements, instructions) were first written in English, then translated by undergraduate members of the lab and checked by graduate students for accuracy. On average, the study lasted around 12 minutes.

2.1.2 Results & discussion

Validating the presurvey tests, there was a significant difference between how important participants found the *primary* and *secondary* topics for their own identities. This is reflected in the mean prioritization[13] scores of the five *primary* topics ($M = 79.1$, $SD = 34.8$) versus that of the *secondary* topics ($M = 43.9$, $SD = 34.7$): $t(2281) = 48.3$, $p = <.001$, Cohen's $d = 1.43$.

For each participant, we calculated the average prioritization ratings of the five *primary* topics which served as the independent variable *Primary Prioritization*. Additionally, we calculated the difference between each participant's pre-manipulation agreement and post-manipulation agreement scores of the *manipulated* topic, which served as the dependent variable *Agreement Change.* The correlations between the variables can be found below in Table 1. Additionally the correlations of the three scales with Agreement Change was measured, but no correlations were found (Appendix B).

[13] In response to the question "This is important to my identity".

Table 1. Experiment 1 correlation matrix:

	1	2	3
1. Agreement Change	-		
2. Prioritization Primary	.79***	-	
3. PreM Prioritization	.26***	.17***	-
4. Openness	.09***	.17***	.18***

* indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$

We conducted a simple regression measuring the impact of the *Primary Prioritization* on *Agreement Change*. The results were significant ($R^2 = 0.630$, $F(1, 2279) = 3878$, $p = <.001$). *Primary Prioritization* significantly predicted *Agreement Change* in the manipulated *secondary* topic (β=.794, p=<.001). This suggests the hypothesized effect: perceived deviation from affiliated groups trigger belief change. In Figure 1, we present the data in binned fashion. We binned the data to reduce noise in the visualization. We normalized the data into eight evenly distributed percentile groups and compared the average *Agreement Change* in each group. Each bin contained between 275-301 data points.

This was the first test of a new experimental paradigm to investigate the impact of a perceived group of epistemic peers on belief dynamics. While there are many questions left to answer, the linear regression testing the primary hypothesis proved to be reliable enough to warrant another test of its validity. The inclusion of other variables into the model resulted in only marginal differences and the analysis of other variables only produced weak correlations (see Appendix B).

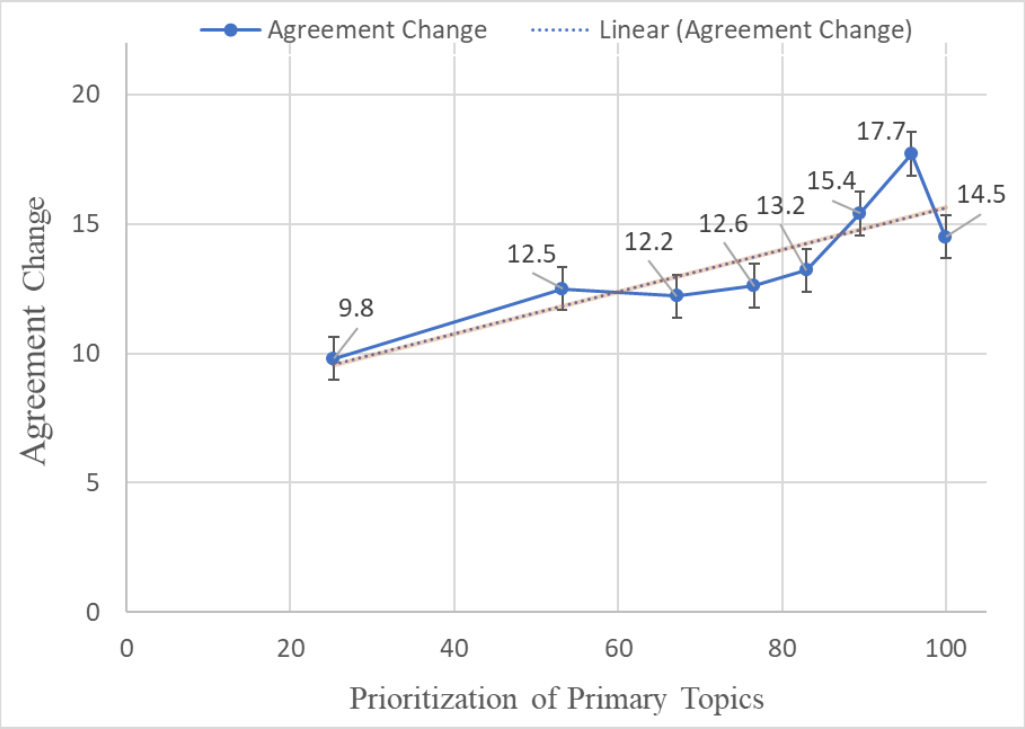


Figure 1. Visual Priority: PT & Agreement Change. The x-axis shows the average prioritization of each of the eight percentile groups. Error bars represent standard error of mean. Dotted line represents the trendline.

2.2 Experiment 2

Experiment 2 was conducted in January of 2022 with the goal of replicating the findings in the first experiment while also including two major components: a control group and a question measuring self perceived knowledgeability. We added the control group to ensure that the change we were observing in agreement Experiment 1 was not merely due to the test-retest effect.

In the control condition, participants received feedback that indicated their choice deviated only trivially from the average responses of their peer group (10 points), whereas in the experimental condition, as in Experiment 1,

participants were informed that their level of agreement regarding the manipulated topic deviated considerably from the mean of their peer group (50 points).

We also added a question on self-perceived knowledgeability ("I consider myself knowledgeable about this topic"). The knowledgeability question was included to investigate whether a participant's perceived domain-specific knowledge plays a role in belief persistence. This was to address the question of why some individuals were resistant to change and whether an individual's perception of their own knowledge plays a role and will include control conditions. Screenshots from the experiment of how the survey questions and the manipulation were displayed are in Appendix E.

The experiment (hypotheses, dependent variables, and intended methods of analysis) was preregistered on aspredicted.com before data collection began (AsPredicted #84279).

2.2.1 Method

In Experiment 2, the main change to the *belief task* was the inclusion of the question on perceived knowledge. This was asked of all topics. We decided to use 7 topics (5 primary, 2 secondary) rather than10 topics (5 primary, 5 secondary) to not further lengthen the belief task. Because an extra question was added to the experiment (knowledgeability), by removing some secondary

topics[14], we could focus the manipulation on two randomly manipulated topics rather than five and keep the task to 21 questions (7 topics x 3 questions).

2.2.1.1 Participants

Gpower was used to calculate a target sample size of 261. This was based on the results of the regression in the first experiment and its effect size[15], as well as the addition of a control group. After opening the experiment to Bogazici students the sample was again snowballed via social media. In total, 5,893 responses were gathered. Of those, 3,169 had to be excluded for multiple reasons: 2,978 because they didn't complete the experiment, 34 because they didn't agree to the consent form, 48 for failing the two attention check questions[16], and 109 for completion times that were extreme outliers (+/- 2 *SD*) (M_{Time} = 9m26s, *SD* = 3m51s). The sample size utilized was much larger than what was calculated on Gpower, the implications of this are in the discussion.

In total 2,724 participants (including 209 Bogazici students) were included in the analyses (*F* = 1,905; *M* = 631; *Non-Binary* = 188) (M_{Age} = 20.4 years, *SD* = 6.0). Experimental: *n* = 1,366 ; Control: *n* = 1,358.

2.2.1.2 Materials

Participants were asked to complete the eighteen-item Need for Cognition scale, α = .90, (Cacioppo et al., 1984) using a 5 point response scale (1 = *not at all characteristic of me* to 5 = *entirely characteristic of me*). The scale was

[14] Secondary topics *included*: Social Media Regulation, Capital Punishment.
Secondary topics *excluded*: GMO Hesitancy, AI (as a threat), Recreational Drug Use
[15] Fixed model regression: R^2 deviation from zero
[16] e.g. "Please place the slider between 70-80" (the slider was a 100 point Likert scale)

translated into Turkish from its original English. It replaced the three surveys in the first experiment which produced no correlations with Agreement Change (see Appendix B). The Need for Cognition scale is typically utilized to measure the degree to which an individual enjoys engaging in thinking as an activity. It was included in Experiment 2 as an exploratory variable and hypothesized that higher scores might predict higher belief persistence and self perceived knowledgeability.

Participants were also asked to complete 3 single-item questions using a ten-point response scale on Conservatism (1 = *not at all conservative* to 10 = *very conservative*), Nationalism (1 = *not at all nationalist* to 10 = *very nationalist*), and Religiosity (1 = *not at all religious* to 10 = *very religious*). The questions were included to serve two purposes, first as checks for within-subject consistency of topic responses and second as exploratory variables for any unexpected correlations.

2.2.1.3 Procedure

After agreeing to the consent form, participants were asked to provide their age and gender. Participants then completed the updated *belief task* with 3 questions beneath each of the 7 statements as opposed to 2 questions beneath each of the 10 statements. Then participants completed an attention check question and then completed the Need for Cognition scale. Lastly, participants were asked whether the manipulation was believable and what they thought of the experiment. Upon conclusion of the experiment, participants were given a

debrief informing them that the results they were shown were fabricated and what purpose the manipulation served.

2.2.2 Results

We again checked the difference in prioritization of the *primary* topics as opposed to the prioritization of the *secondary* topics. The prioritization difference between topics was comparable to those found previously: *primary* ($M = 73.2$, $SD = 24.6$); *secondary* ($M = 40.1$, $SD = 36.7$).

As in Experiment 1, we calculated the variable *Agreement Change* as the difference between the pre-manipulation and post-manipulation agreement scores of the manipulated *secondary* topic[17]. In line with the primary hypothesis, the difference in mean agreement change between the experimental ($M = 9.45$, $SD = 20.7$) and control ($M = 4.74$, $SD = 14.5$) groups was significant, though the effect size was small: $t(2671) = 6.78$, $p = <.001$, Cohen's $d = 0.31$.

The other three individual-item scales — Conservatism, Nationalism, Religiosity — were not correlated with agreement change, though they were all correlated with one another, particularly Conservatism and Religiosity (Appendix D). As expected, agreement with topics that generally have support from progressive, secular individuals such as LGBT+ Rights and Abortion Rights were negatively correlated with both Conservatism and Religiosity.

[17] As a reminder, the agreement score is an endorsement of a statement on a particular topic on a 100 point sliding scale. The question is phrased as: "I agree with this statement".

Table 2. Experiment 2 correlation matrix:

	1	2	3	4
1. Agreement Change	-			
2. PreM Agreement	.12	-		
3. PreM Prioritization	.24***	-.17	-	
4. PreM Knowledgeability	-.10***	-.78***	-.01	-
5. Need for Cognition	-.05*	-.06*	-.04	-0.08***

* indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$

The Prioritization of the 5 *Primary* topics was again normalized and split into 8 even segments in order to extract descriptive information of the behavior of participants (Figure 2). *Agreement Change* peaked for those whose average prioritization of the primary topics was between 80-90, and steadily decreased from there, replicating the finding in experiment 1.

A multiple linear regression was conducted to test the primary model: the impact of the *Condition* (experimental v control), *PreM Knowledgeability* (the pre-manipulation knowledgeability of the manipulated *secondary* topic), and *PreM Prioritization* (the pre-manipulation prioritization of the manipulated *secondary* topic) on *Agreement Change*. The overall regression was statistically significant ($R^2 = 0.302$, $F(3, 2637) = 52.3$, $p = <.001$). Though the effects weren't large, the *Control v Experimental Conditions* ($\beta = .277$, $p = <.001$), *PreM Knowledgeability* ($\beta = -.080$, $p = <.001$), and *PreM Prioritization* ($\beta = .220$, $p = <.001$) significantly predicted *Agreement Change*.

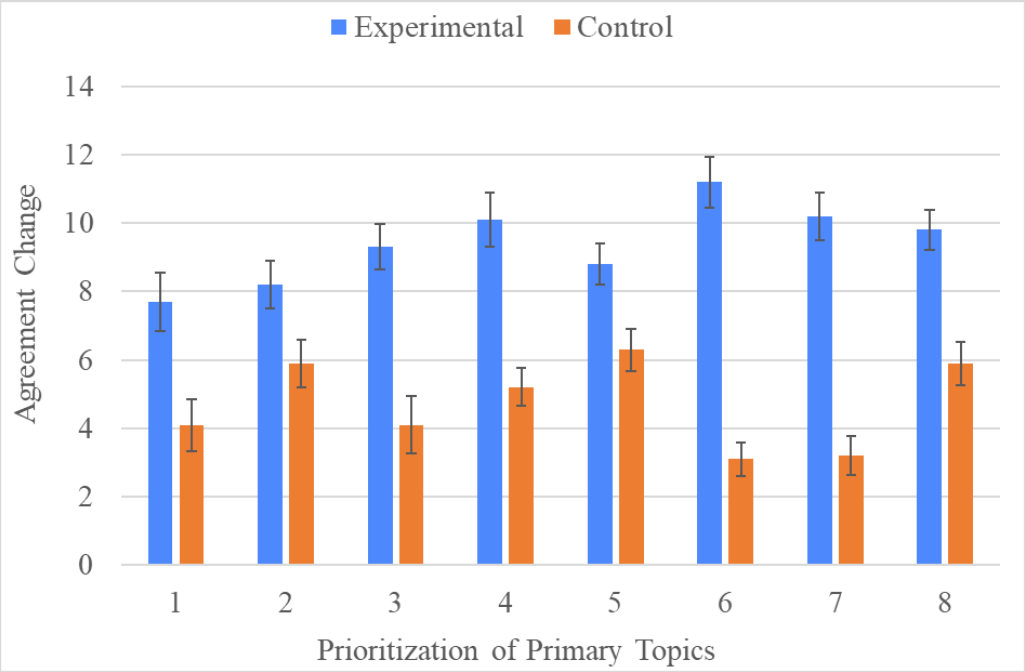


Figure 2. Agreement Change by Prioritization of Primary Topics[18]. Error bars represent standard error of mean.

Because *Agreement Change* was not correlated with *Primary Prioritization* it was not included in the analysis. Though it was not captured by the *Primary Prioritization* variable, the impact of the group on individual beliefs is made clear by the significant difference in *Agreement Change* across the experimental and control conditions.

Participant Prioritization and Knowledgeability of all topics was highly correlated, $r(2,722)=.62$ ($p<.001$), reinforcing the assumption that the more prioritized a belief is for an individual the more knowledgeable they will believe themselves to be. This is reinforced by the significant negative impact of *PreM Knowledgeability* in the regression. Peak agreement change in both

[18] See also Appendix 3 for line graph comparable to Figure 1

experiments occurred in participants whose *Primary Prioritization* (average prioritization of 5 primary topics) was between 80-90. This may be indicative of a resistance to opinion change when an individual believes they are either prototypical members of a particular group, or that they are an authority on a belief.

Table 3. Experiment 2 agreement change as a function of variables:

Predictor	*b*	SE	CI Lower	CI Upper	t	p	*β*
Condition: Exp v Cont	4.97	0.74	3.51	6.44	6.65	<0.001	0.28
PreM Prioritization	0.18	0.01	0.15	0.21	11.34	<0.001	0.24
PreM Knowledgeability	-0.05	0.01	-0.06	-0.02	-3.92	<0.001	-0.08

For this reason a t-test was conducted to check the difference in mean Prioritization Change between the experimental and control groups. The difference was significant, with a medium effect size: $t(2671) = 11.3$, $p = <.001$, Cohen's $d = 0.49$. The results of a t-test warranted the construction of a second regression model. Therefore, a linear regression was conducted with *Prioritization Change* as the dependent variable in place of *Agreement Change*.

The regression tested the impact of the Condition (experimental v control), PreM Knowledgeability (the pre-manipulation knowledgeability of the

manipulated secondary topic), and PreM Agreement[19] (the pre-manipulation prioritization of the manipulated secondary topic) on Prioritization Change. The overall regression was statistically significant (R^2 = 0.412, $F(3, 2637) = 477$, p = <.001). The variables Condition (β = .198, p = <.001), PreM Knowledgeability (β = .35, p = <.001), and PreM Prioritization (β = .44, p = <.001) significantly predicted Agreement Change. This effect was hypothesized, but it was expected to be secondary to Agreement Change. In fact, a shift in prioritization seems to have been more of a dominant strategy among participants in response to the manipulation.

Finally, because the undergraduate pool of psychology students skews female, the sample sizes of the pretest surveys highlighted issues which impact women directly. This is evident in that three of the five *Primary* topics were Abortion Rights, the Gender Pay Gap, and the Pink Tax. We see the implications of this represented in the descriptive data in Table 4.

Table 4. Experiment 2 descriptive data by gender

Condition	Gender	Primary Prioritization	Primary Agreement	Agreement Change
Control	F	80.7	92.6	8.69
	M	51.2	78.3	7.32
Experimental	F	80.1	91.4	14.7
	M	53.4	80.7	11.3

[19] In the regression with *Agreement Change* as the DV (Pg 48), PreM Prioritization was used as one of the three variables. Because *Prioritization Change* is the DV in the second regression, PreM Prioritization is substituted for PreM Agreement.

The results show that men do not prioritize those issues such as the gender pay gap and abortion rights even if they agree with the with the women in their peer groups on their stances on issues. This may be because the repercussions of legislation regarding such topics are not typically immediately apparent to men. This is most evident in the relative consistency of the Primary Agreement and Primary Prioritization numbers across the female sample, as opposed to the same values in the male sample. Following the findings in the descriptive data, a t-test was conducted to check the difference in mean Agreement Change between the experimental and control groups of only the female sample. The difference was significant, with a slightly larger effect than that seen in the general sample: $t(1863) = 6.42$; $p = <.001$, Cohen's $d = 0.34$.

2.2.3 Discussion

Experiment 2 demonstrated that the views of one's peers, in conjunction with both how important a belief is to an individual and how knowledgeable they believe themselves to be, can be analyzed together to construct a plausible model of belief dynamics. The strong correlations between prioritization and knowledgeability in conjunction with the effect of the manipulation on prioritization change, suggests their roles in the dynamics of individual belief may be even larger than what was captured by the experiment.

Additionally, though peer group alone was not a predictor of agreement change, its effect was more strongly seen across the differences between the control and experimental conditions rather than the prioritization of the primary topics. The prioritization of the primary topics could potentially have served as

a latent variable causing the significant difference found between the two conditions. Additionally, a greater effect of agreement change was seen in the female sample. This is possibly due to the increased saliency of the topics used to generate the illusory peer group, as seen in the Primary Prioritization values reported by the female sample. Still, there is a need to address the potential effect of experimenter demand on the results.

The sample sizes utilized were much larger than what was calculated by Gpower as necessary. The larger sample was unexpected, but utilized in order lend power to the findings whether in support or in conflict with the hypotheses. However, though the error rates were low, follow-up work including bootstrapped samples or random sampling in order to ensure that the effect is consistent through smaller samples can be conducted.

2.3 Experiment 3

The goal of experiment 3 was to conceptually replicate the findings in experiments 1 and 2 while also mitigating concerns of experimenter demand. This was done by altering the experimental procedure such that the manipulated topic was less evident and the demand on the participant to modify their responses reduced.

2.3.1 Method

The procedure of the belief task was altered in Experiment 3 to mitigate the potential confound of experimenter demand present in Experiments 1 and 2. In

Experiment 3 participants were presented with a cover story (Appendix G) to mask the purpose of the study. It claimed the purpose of the survey was to gauge levels of interest and awareness of topics relevant to contemporary public discourse in Turkey across generations.

2.3.1.1 Participants

After opening the experiment to Bogazici students the sample was again snowballed via social media. In total, 728 responses were collected. We removed 53 responses from the analysis for a number of reasons: 3 were incomplete, 9 did not consent, and 41 were excluded for failing 4 of 7 attention check questions (explained in procedure).

In total 675 participants were included in the analyses (F = 527; M = 112; *Non-Binary* = 36) (M_{Age} = 20.5 years, SD = 4.2) (Experimental: n = 329; Control: n = 346).

2.3.1.2 Procedure

Participants were told they would be given two sets of 7 question surveys which may have any of the 16 topics listed in the introduction. They were told some items might repeat in different different orders to better understand the effect certain opinions may have on one another and whether the order in which topics are presented have an effect on responses. In total participants received 5 *primary* and 2 *secondary* topics in both rounds, however, both of the *secondary* topics were repeated (in random order) with 4 new topics and one repeat *primary* topic from the first round.

In between the two rounds of questions participants were shown the mean *agreement* responses across participants of their age groups of all seven topics they responded to in the first round. The means of one of the two *secondary* topics however was manipulated to show significant (49 point) deviations from how they responded. They were only presented with the means of the responses, it was not framed in relation to their own responses. They then had to indicate whether the participant means were above or below 50 and were told this was a simple task to create some time between the two rounds of surveys (screenshots in Appendix F). This was done for each of the 7 topics they responded to in the first round. These 7 topics were used as the attention check mentioned above in the Participants section.

2.3.2 Results

We again checked the difference in prioritization of the *primary* topics as opposed to the prioritization of the *secondary* topics. The prioritization difference between topics was comparable to those found previously though average prioritizations were lower for both categories: *primary* (M = 65.4, SD = 23.9); *secondary* (M = 37.8, SD = 29.3).

As in Experiments 1 & 2, we calculated the variable *Agreement Change* as the difference between the pre-manipulation and post-manipulation agreement scores of the manipulated *secondary* topic[20]. In line with the primary hypothesis, the difference in mean agreement change between the experimental

[20] As a reminder, the agreement score is an endorsement of a statement on a particular topic on a 100 point sliding scale. The question is phrased as: "I agree with this statement".

(M = 6.54, SD = 11.4) and control (M = 3.38, SD = 10.5) groups was significant, but small: $t(673) = 2.35$, $p = .019$, Cohen's $d = 0.18$. Table 5 shows that the variables were only moderately correlated with agreement change.

Table 5. Experiment 3 correlation matrix general sample:

	1	2	3	4
1. Agreement Change	-			
2. PreM Agreement	.08*	-		
3. PreM Prioritization	.06	.54***	-	
4. PreM Knowledgeability	.02	.17***	.47***	-
5. Need for Cognition	-.03*	.06	.09*	0.09*

* indicates $p < .05$; ** indicates $p < .01$; *** indicates $p < .001$

Additionally, there was a difference seen between the two manipulated topics (Social Media Regulation and Capital Punishment) in the recorded levels of agreement change, the differences between the means was significant, though with only a moderate effect size: $t(143) = 2.20$, $p = .003$, Cohen's $d = 0.37$. For this reason, we ran a t-test was run measuring the difference between the means of the Experimental group which received Social Media as a manipulated topic and the Control group, the difference between the means showed a medium effect size: $t(266) = 3.2$, $p = .002$, Cohen's $d = 0.49$. Though it isn't clear why such a difference appears between the two topics, it may be that the public narrative has shifted in Turkey over the course of the two years since the pretest surveys were conducted.

Table 6. Descriptives: passed all attention checks

	n	*M*	*SD*
Social Media Regulation	52	10.1	27.2
Capital Punishment	93	2.3	15.7
Control	216	2.2	11.4

Then, due to the simplicity of the attention check and its centrality to the manipulation, it was decided to narrow the analysis to the 361 participants (Control $n = 195$; Experimental $n = 147$) who answered all 7 attention checks correctly. Going forward, this subset of 361 participants is referred to as the G7 sample. A t-test was conducted to check the difference in mean agreement change between the experimental ($M = 7.99$, $SD = 19.7$) and control ($M = 1.57$, $SD = 11.0$) groups. The difference was significant, with a larger effect size in the G7 sample than that of the larger sample: $t(292) = 3.83$, $p = <.001$, Cohen's $d = 0.42$. The findings of the t-test demonstrate that the manipulation had a significant effect on agreement change when participants took the time to observe the means on the manipulation screen. This was despite the differences between their responses and the responses of their peers not being made explicit to them.

Finally, because the same topics were used, we carried out an analysis on the effect of gender. As in Experiment 2, female participants reported higher prioritization of the *primary* topics ($M_{Female} = 72.0$; $n_{Female} = 527$)($M_{Male} = 48.4$; $n_{Male} = 112$). The between-gender difference *Primary Prioritization* was significant: $t(637) = 9.92$, $p = <.001$, Cohen's $d = 1.03$. A graph showering the

discrepancy by gender in prioritization versus support of Abortion Rights as a sample topic is available in Appendix H.

A t-test was conducted on the G7 female sample (n = 282) to check the difference in mean agreement change between the experimental (M = 8.58, SD = 21.2) and control (M = 1.54, SD = 11.0) groups. The effect size was larger than in the sample including both genders: $t(280) = 3.65$, $p = <.001$, Cohen's d = 0.44. The difference between the effect sizes of the t-tests found in the general G7 (.42) versus the female only G7 (.44) groups are not large because the G7 sample is predominantly female (81%). Still, we see the effect size increase concurrent with the significant difference in the prioritization of primary topics.

2.3.3 Discussion

Experiment 3 produced a conceptual replication of the findings from the first two experiments, namely that deviation from one's peers on a topic produces a change in participant responses in the direction of their peers. In line with the overarching theoretical framework, this is interpreted as belief updating. The updated experimental paradigm used in Experiment 3 was built in a way that did not receive optimal engagement from participants. This was clear in that only half of participants answered all of the 7 attention checks correctly despite it being a simple binary assessment. The results of the G7 group versus the larger sample showed a marked increase in effect size and mean *Agreement Change*. This suggests the manipulation itself was salient when participants attended to it.

Additionally, the gender of participants was again found to play a role. It seems the consistency of effect is due to the difference in the saliency of the primary topics to the female participants as opposed to that in the male participants. Consistent with the central hypothesis, this suggests that the prioritization of beliefs representing one's affiliations (with the variable *Primary Prioritization* as its proxy) play a role in belief updating.

2.4 Conclusion

We introduced a novel paradigm attempting to test the impact of virtual groups on individual beliefs. The central hypothesis was that individual beliefs are impacted by the beliefs of their epistemic peers (defined as those individuals they share other beliefs with). This was measured by asking individuals how central certain topics were to their identities, which allowed for a virtual group constructed around each individual's responses.

The first experiment found that greater prioritization of primary topics (used to create the illusion of the group), predicted greater belief change in the secondary topic. This suggested that the belief change was induced by perceived deviation from the virtual peer group. The results of the first experiment were replicated in the second experiment which included a control group. Additionally, the second experiment included a question asking individuals how knowledgeable they believed they were about each topic. The analysis of the second experiment found that two factors that impact susceptibility to belief change were prioritization of the manipulated topic and

knowledgeability of that topic, two variables which were found to be highly correlated.

In the first two experiments, the deviation from one's peer group was presented explicitly to participants which raised concerns of experimenter demand. To address this concern, a third experiment was constructed which added subtly to the presentation of the manipulation. The results of the third experiment demonstrated that perceived deviation from one's peers does trigger an updating of responses from participants. The effect was however most present in a subset of the sample that actually attended to the manipulation (which was more subtle and required more attention from participants). The third experiment also replicated the finding from the second experiment that prioritization of the manipulated topic and knowledgeability played a role in belief persistence. As in the second experiment, the prioritization of a topic and perceived knowledgeability about a topic were highly correlated.

An unintended consequence of the selected topics was the particular saliency for the female sample. This provided an additional perspective on the data wherein one grouping variable, gender, highly predicted prioritization of the primary topics. It was found that the female sample was more susceptible to belief change in the secondary topic, suggesting the impact on their beliefs due to deviation from their peer group was greater. This is of course interpreted as being purely due to the increased prioritization of the primary topics, therefore assumed to increase the saliency of the manipulation.

2.5 General discussion

This work does not attempt to reduce the complexity of belief change with regard to the number of factors. Instead, it attempts to reduce the theoretical motivations underlying belief change (to uncertainty mitigation) while attempting to explore potential models of the process of belief change and the many variables that may play a role in its dynamics. More work needs to be done to determine individual susceptibility to persuasion and the conditions under which one's opinions are mutable. Individual belief dynamics are complex. Paradigms similar to the one used could be further utilized to test for determinants of trust and the perception of information which confirms or challenges one's beliefs.

The internet has been observed to have a significant impact on large scale ideological polarization and the potential for virality of content, suggesting virtual interactions are sufficient to generate real world tribalism. The creation of virtual epistemic peer groups for the experiment was meant to mimic the interactions individuals have with content on the internet. In the introduced experimental paradigm participants identify their own epistemic peer groups via their indicated levels of agreement and prioritization for a range of topics. As was clear from the third experiment however, it is unclear whether the suggested existence of individuals one agrees with is enough to generate the feeling of a group, and if it is whether the affiliation is strong enough to generate belief change outside of a context in which targeted belief change is demanded.

The empirical work presented above was an attempt to draw a testable hypothesis from a specific portion of a broad model which attempts to provide an account of the mechanisms driving individual belief dynamics and by extension the individual identity. The BDM, further explored in Chapter 3, attempts to provide an overarching theoretical framework that places the diverse range of findings outlined in Chapter 1 into a broader context.

CHAPTER 3

PHILOSOPHICAL CONTEXT OF THE BDM

What is presented below is a model of identity that places uncertainty mitigation at the core of its construction. The Belief Dynamics Model, or BDM, maintains that identities are mutable constructs that emerge from the dynamics of belief maintenance. Belief dynamics is defined as the updating of beliefs in response to interactions with one's environment. By extension, an individual's identity represents their particular survival strategy given their specific socio-ecological niche.

Beliefs in the context of this paper are not necessarily cognitive. A belief is a state internal to the agent P at the level of the agent P. The agent here is a biological, autopoietic, non-equilibrium system. The minimum viable agent in this sense is a blurred line, inasmuch as the line between living and non-living is blurred. At some point a system of organic chemistry achieves sufficient complexity, through the retention of sufficient information, that it generates information within its environment (i.e. behavior) rather than being merely a reflection of it. A belief is therefore to be interpreted as internal states which are maintained regarding the states generated by perception. The movement of bacteria in their environment is also considered a function of the beliefs maintained by the bacteria. In line with the provided definition of beliefs, this does not require an inner experience of bacteria or any attributes we might ascribe to cognition.

As in predictive brain models, they are considered functionally analogous to priors in a Bayesian network. Beliefs therefore represent probabilistic models of the world which in turn satisfy the structure of the belief network itself. As an agent interacts with their environment, their beliefs function as predictions whose objects are the units of information they might encounter. Information in this sense refers to the internal states generated by states either internal (e.g. hunger) or external (e.g. a snake) to the agent.

As in the Active Inference model, the agent's primary imperative is taken to be the minimization of prediction errors (Friston, 2010). Incongruencies between beliefs and perception, ultimately inaccuracies in the maintained model of the world (which includes one's body), result in updates to beliefs. The updating of the model however is done to minimize the degree of variational free energy (by way of reduction of prediction errors) leading to a process of updating which tends toward the change which causes the least disruption to the model (Friston et al., 2015; Pezzulo et al., 2018). This means that the degree of influence beliefs and perception have on one another is ultimately contingent upon the prioritization of the beliefs involved. This is because the prioritization represents the structure of the ordered network of beliefs maintained by the agent. The weight given to information which serves as the content of perception is therefore contingent upon the beliefs relevant to the context in which the information is perceived. As such, any instance of perception falls somewhere along the continuum of *seeing is believing* and *believing is seeing*.

This definition of beliefs looks at all biological life as individual expressions of given belief networks. A phenotype is viewed as an embodied set of beliefs which represents a unique survival strategy. The degree of freedom within a given agent's belief network is constrained by its genotype, or the beliefs it inherits as a member of its specific species. The species itself therefore represents a particular set of beliefs which serves as the context for interpreting the beliefs maintained by a particular agent. Beliefs therefore serve as the parameterized probability space an agent can exist in. The limitations of information an agent can perceive are defined by the limitations in the beliefs that comprise the individual[21]. Because the ordered belief network maintained by the agent represents its unique survival strategy, the function of prioritization (meaning the ordering) serves as the details of that strategy. If the beliefs are the ingredients of a recipe, their respective weights (prioritization) are the quantities.

Throughout this paper, uncertainty is defined as belief-perception incongruence. This is meant to reflect the inter-determinacy between the beliefs in a given individual's belief network. The inter-determinacy means that those beliefs which are most prioritized, typically those dealing with survival, are difficult to deprioritize because information is interpreted to mitigate threats to those beliefs. This results in a system in which beliefs both motivate the

[21] For example, the limitations of the wavelengths of light that the photoreceptor cells in the human eye can respond to result in a limitation at the level of the whole person. This is a component of the free energy minimization paradigm, in that the restriction of an agent to a limited number of states is beneficial in the long term and therefore emerges as a successful strategy (Friston et al., 2015).

inferences extracted from perception and are updated in accordance with those inferences.

With this view we can make sense of biological imperatives around survival, and by extension the production of offspring, as the prioritization of beliefs which generate the corresponding behavior. What survival entails varies over the life cycle of every particular species. It typically begins with behaviors ensuring the physical survival of the agent up until the point of sexual maturity, at which point reproduction begins to become highly prioritzed and in some species take immediate priority over physical self preservation. For other species, survival is a balance between the prioritization of the physical notion of self preservation and the abstract self preserving notion of reproduction.

There are many species which prioritize self preservation insofar as it leads to reproduction, with reproduction being the priority. This is common throughout the animal kingdom. As one example, in most varieties of octopus the male dies soon after fertilizing the female's eggs and the female dies in the process of hatching them. For some species, like humans, the prioritization of self preservation tends to be usurped by the preservation of offspring. Self preservation however remains highly prioritized because child rearing is necessary. Trade-offs like this — regarding the frequency of reproduction cycles, the number of offspring in each cycle, and the longevity of the caretaker — are common across a variety of evolutionary strategies species adopt (Sibly & Brown, 2009). Humans seem to be distinct, however, in that the prioritization of a concept or set of concepts can play the same role.

Support of an ideology can supplant self preservation as the primary means of survival, meaning one can choose to die for a cause if they deem the cause worthy enough to prioritize over their own individual existence. If we look at the case of volunteering to go to war, while there can be many beliefs which are prioritized in that context, here are three different reasons: 1. the prioritization of an ideology, say nationalism; 2. the prioritization of the wellbeing of family, say for the financial compensation of being a soldier; 3. the prioritization of one's own image, say for the prospect of glory or caving to social pressure. The assessment of rationale is difficult not only because of the complex and dynamic nature of beliefs, but also because individuals may produce any one of those beliefs as a response to questioning because they believe the context of the question requires they project a certain image.

Humans can also deprioritize self preservation to the point where suicide becomes a reprieve. It is on the basis of these relatively common phenomena within the human species that the assertion of the interdependence of belief prioritization is grounded in the paper. Namely that looking at complex life as agents all the way down to the perimeters of organic chemistry, also means we look at complex organisms as beliefs all the way down to the same point. If we investigate the dynamic belief hierarchies individuals maintain, we may be able to build a better understanding of human behavior as an emergent function of those beliefs.

3.1 Outline of the BDM

Before going into the implications of the belief dynamics model, the following section is an outline of how it functions and the assumptions it makes and attempts to explain.

Every biological agent is the expression of a dynamic network of beliefs. The prioritization of a belief is indicative of the degree of uncertainty it mitigates. The most prioritized beliefs are typically those dealing with the physical survival of the agent themselves, though this can be abstracted to an idea or another agent. Uncertainty is defined as the incongruence between beliefs and perception. The perception of information is itself impacted by one's beliefs, though the degree of impact beliefs have on perception is contingent upon the prioritization of the beliefs relevant to a given context. For example, if an individual has highly prioritized their affiliation as a social progressive they may not attribute epistemic validity to information from a source they believe to be highly conservative.

Uncertainty serves as the *theoretically* quantitative sum of belief-perception incongruence. In the model, information which mitigates uncertainty is referred to as a *validation*, whereas information which increases uncertainty is referred to as a *threat*. Therefore, in any given moment, uncertainty serves as the sum of an agent's threats and validations. A *neutral* state is actually a state of resting latent uncertainty in that an agent is typically passively attune to threats even if they are not immediate. Anxiety, for example, is explained as a state of high resting latent uncertainty. Uncertainty exists even

at rest both because perfect information does not and because physiological needs arise, including the need for social interaction.

The emergence of affective states in complex organisms is explained as a physiological mechanism incentivising uncertainty mitigation, effectively the avoidance of threats and the seeking of validations. The phenomenological gradient of affective states is therefore interpreted as a function of uncertainty. While there can be *n* dimensions included into a model which accounts for affective states, this model considers the positive-negative dimension of affective states to be the central axis, akin to the line drawn end to end through the neck of an hourglass.

As argued in Chapter 1, the view that humans evolved to equate social inclusion with increased likelihood of survival is central to understanding the prioritization of beliefs in the human context and its implications. In the model, every close interpersonal relationship is interpreted as a source of uncertainty mitigation, and the threat of ostracism is passively perceived as a threat to survival. This is why being an atypical member of a group can therefore be perceived as a threat. Group membership is explained as the belief one maintains regarding their affiliation with a set of individuals. Because social ostracism is equated with threats to survival, the larger the number of affiliations one maintains the less uncertainty they experience due to threats to a particular affiliation. The result is that a robust identity with many affiliations can handle threats better than an identity built on a few. Therefore, the complexity of an organism's belief structure is synonymous with the complexity of its survival strategy.

Being that affiliations are beliefs, they exist within the belief hierarchy and their prioritization is relative to other beliefs. The response to threats to a particular affiliation, whether they are internal or external to the set of individuals comprising that affiliation, will be contingent upon the degree of uncertainty that affiliation mitigates. For example, person P's beliefs regarding their fandom of American football may trigger a defense of the sport in response to the statements made by person Q who is a bigger fan of basketball. Person R may even join the debate and argue in support of Person P's defense of football over basketball even though Person P is a Jets fan and person R is a Patriots fan. This is despite the fact that, in the context in which football is not threatened, P and R, whose teams are rivals, would likely argue against one another in defense of their respective teams. It's also possible that support for their respective teams are so prioritized that they refuse to speak to one another at all (particularly when support for a team is highly political).

Say, however, that person P identifies more as a social justice activist than as a sports fan. One day person S points out that American football is exploitative and that the National Football League has a history of both excusing the brain trauma incurred by its players and the domestic violence committed by its players. Person P may choose to distance themselves from their beliefs regarding their football fandom to not risk their more prioritized affiliation as an activist. Even familial ties are interpreted as an affiliation, and often a highly prioritized one. This is evident in perhaps a more common example of being able to personally insult a sibling or parent but having less tolerance for a friend insulting that same relative.

3.1.1 Belief network

The sets of beliefs from which the phenomenon of self preservation emerges are taken to be the starting point for all biological life. Whether it is explicit to the human as an infant or not, the behavior of an infant seems to be motivated by a drive to survive. This is without being taught or having experienced enough to justify a belief that continuing to exist is worthwhile. Likewise, one does not need to attribute a rich inner experience to an ant but it is clear the ant's behavior reflects a set of beliefs constructed around its preservation and one could argue that it seems the preservation of the colony is more highly prioritized in the individual ant. In this interpretation, certain parasites seem to have the ability to restructure the prioritization of an ant's beliefs upon entering their central nervous system, with the behavior of the ant coming to reflect the new hierarchy of the parasite: namely to get eaten by something larger the parasite can infect (Martín-Vega et al., 2018).

A human infant is born with the beliefs that it must carry out a certain number of active functions like breathing and feeding, and passive functions like thermoregulation, sleeping, digestion, and avoidance of pain. If the infant is hungry, too hot or too cold, too tired, or otherwise in any sort of physical discomfort it has the belief that crying will resolve these issues. To the infant these are all presented through introspection as physical discomfort which is interpreted as threats to survival, and crying is its singular tool for mitigating uncertainty. It is up to the caregiver to attempt to determine which threats the infant is responding to and seek to potentially alleviate its uncertainty.

The infant perceives all stimuli through the lens of these initial beliefs of infancy. The critical belief the infant begins to develop is that the caregiver is the mitigator of its threats. The infant and later the child develops an attachment to the caregiver as its primary survival strategy. Evolution does not seem to have placed its bet on agents behaving in a way which in every moment reflects the structure of their hierarchy. It instead seems to have developed a process of phenomenological incentives to guide behavior toward threat mitigation, namely the affective gradient. Through the biases driven by the initial beliefs of infancy the individual constructs and maintains a complex network of beliefs that reflect its niche survival strategy across a range of environments.

In this framework, biases are viewed solely as a function of beliefs, meaning that the word bias is the interaction between a relevant belief and a unit of information (if belief is the noun then bias is the verb). Therefore, context is critical in that it determines which beliefs are triggered and what the resulting identity protective behavior is in response to.

A bias is a phenomenon that results from the implementation of a set of beliefs, like a ball passing through a net is typically considered a point, the bias is neither the ball nor the net, it is merely an interpretation of the interaction. All information triggers a collection of relevant beliefs through which that information is interpreted. The processing of information is inherently biased in that it is done through the lens of a relevant set of beliefs, and any learning that

takes place is likewise merely the updating of a set of relevant beliefs, occasionally also constructing new beliefs in the process.

Beliefs in this sense require an object, regardless of whether they are photoreceptor cells responding to photons or whole persons maintaining beliefs regarding the function of photoreceptor cells. Likewise, regardless of whether an individual is making subtle adjustments to their balance while learning to ride a bicycle or the function of a quadratic equation, there are physical subsystems of agents which compose the human agent which are updating their beliefs regarding their behavior in response to interaction with stimuli.

3.1.2 Relationship to the active inference framework

The BDM serves as a psychological extension of the active inference model (Friston et al., 2017; Friston & Stephan, 2007; Kuchling et al., 2019), which views the bayesian maintenance of an agent's network of beliefs to be driven by an underlying agentic imperative of threat mitigation. The concept of belief, in the BDM, is defined as being as analogous to weights in a probability distribution that undergo a bayesian style updating, drawn particularly from predictive brain models built on the free energy principle (Clark, 2013; De Lange et al., 2018; Friston & Stephan, 2007).

In accordance with the active inference model, the BDM maintains that beliefs are actively updated to reflect perception in an attempt to minimize uncertainty (Friston & Stephan, 2007). This is done by reconciling the internal states stored by the agent, as weighted priors, and the external states perceived by them, as a form of perpetual weighted hypothesis testing. The updates can

amount to changes in content as well as prioritization, a distinction backed by neurocomputational evidence that demonstrates beliefs are organized in a hierarchy that reflects motivations (Pezzulo et al., 2018). The active inference model refers to information which is incongruent with existing beliefs as *surprise*, and refers to *uncertainty* as a phenomenon which occurs at the level of the agent. Because the BDM deals specifically at the level of the agent, and seeks to offer a partial potential model of beliefs, the term uncertainty is preferred over surprise.

The BDM maintains that all information gathered is perceived along a gradient of threats or validations. Threats and validations are taken to be reciprocal concepts that can be defined by their degree of exacerbation and mitigation of individual uncertainty. In the social context, threats would be information threatening social exclusion (ostracism and atypicality literature) whereas validations would be information suggesting inclusion and acceptance.

Information which threatens beliefs will generate behavior that attempts to resolve the threats. The intensity of the reaction to the threat is a result of both the perceived legitimacy of the threat and the prioritization of the belief being threatened. For example, a legitimate threat to a non prioritized belief, like serious criticism of one's cooking when one doesn't necessarily consider themselves a good cook, may produce a greater reaction than a three kilogram dog barking at an individual. Even though the dog barking is ostensibly an attempt to simulate a physical threat, coming from a small dog such a threat may be interpreted as minor and therefore not legitimate. However, one's closest friend viciously criticising one's cooking likely would not generate as

much uncertainty as having one's life explicitly threatened by an eighty kilogram dog without a leash. Alternatively, say one considers themselves a good cook, they are a professional chef. Viscous criticism from a non-chef will likely not be as threatening as equally viscous criticism from a respected chef. Again, the source of the threat determines its validity, but regardless of the validity of the threat the uncertainty generated is equally determined by the degree to which the threatened belief is prioritized.

3.1.3 Learning & curiosity

The internal states of the agent in this model are associated beliefs which are non-propositional in that they are embodied probability distributions. The degree of uncertainty an agent experiences is a function of the incongruence between information serving as the content of what is anticipated, as the posterior probabilities in a Beyesian algorithm, and information that is perceived from external states, as evidence in a hypothesis test. The updating that occurs, as learning, is a result of this incongruence between beliefs and perception. In this sense, learning is the phenomenon that emerges from this incongruence. The result of the totality of beliefs that contribute to the same, complex, non-equilibrium system is the identity of the individual.

This interpretation of beliefs, and therefore of learning, does not require the individual to be maintaining a belief about information before the individual has perceived it. Every biological agent is limited in the types of information they can perceive and therefore limited in the beliefs they can construct

regarding that information. The agent's biological constraints represent the sum of the parameters of the complex system from which behavior emerges. The use of tools extends the boundaries of these limitations, like goggles converting wavelengths of light from the infrared spectrum into the visible spectrum. Then there are cases like that of exposure to ionizing radiation, which impacts the agents which comprise the human body. The agent at the human level is able to perceive only the repercussions of the radiation and not the radiation itself. The agents comprising the body are impacted and react to the damage but only in ways they are capable of reacting. Radiation exposure above a certain threshold is insufficient for the agent at the human level to maintain free energy minimization and therefore results in death of the agent at the human level and nearly every other agent comprising them.

Here curiosity plays a critical role. The agent that can maximize information gathering from an uncertain context while also minimizing exposure to threats will be best poised to navigate their environments. Say W, X, Y, and Z see a snake at their feet. The uncertainty experienced by X, Y, and Z causes them to immediately jump back to avoid being bitten. W does not sufficiently experience uncertainty and gets bitten. X is curious to learn about the snake but gets too close and is also bitten. Y is curious but keeps their distance and is able to learn information about the snake, like the environment from which it emerged, and is better equipped in the future to avoid snakes. Z experiences panic from too much uncertainty and runs off, learning little from the situation. Y in this context has the appropriate, nuanced behavioral reaction

to uncertainty to be able to maximize learning and minimize high uncertainty states in the future.

The balance between curiosity and caution is critical in physical as well as social contexts. Individuals not able to gauge the relevant beliefs of others or behave in a manner that is congruent with those beliefs (like filtering speech) will run the risk of ostracism. Avoiding threats requires the agent to update their beliefs relevant to specific contexts including their beliefs regarding what constitutes a particular context. Together, these beliefs comprise assessments of what could colloquially be referred to as appropriateness.

Curiosity is the resolution of gaps in information. This can be done without necessarily invoking the future. Any recognized gaps in sets of information, regardless of whether any negative repercussions are immediately apparent, can be resolved. Critically, curiosity is only useful in the absence of other immediate threats to more prioritized beliefs. Curiosity is the name we give to the gathering of information and can be engaged while also balancing a potential threat. Y can seek out more information about the snake, but will be best served when the threat of being attacked is minimal, unlike X who got bitten. There are circumstances in which curiosity and caution are not appropriately balanced, such as when curiosity does in fact kill the cat.

Let's take learning not to kick a hornet's nest as an example. One can learn this isn't wise either through kicking one themselves, experiencing the repercussions of someone else kicking one, or receiving the information secondhand. These are three different strategies for uncertainty mitigation,

namely of the balance between curiosity and caution that allow us to understand the continuum of different strategies representing uncertainty mitigation. The limitations of the human lifespan requires that we balance information gathering from firsthand interaction, firsthand observation, and secondhand through communication. The terminology here is not important but the gradient of information gathering is critical because in every context we exist at a slightly different position in the gradient. I may go swimming in a thunderstorm but avoid confrontations with wild snakes. Likewise a friend who grew up in the jungle may be perfectly adept at handling a wild snake but not willing to jump into the sea at the slightest hint of rain.

3.2 Beliefs in the social context

As individuals interact, they draw inferences regarding the degree of congruence between their beliefs and the beliefs held by others. The inferences individuals construct regarding their environments are unique to them. This results in individuals maintaining unique interpretations of their interactions with their environments, which are nonetheless fundamentally a part of the environments they exist in. The human organism as a system is as indivisible from the larger systems that they are a part of as is any subsystem of agents within them. Meaning, an individual P and their set of beliefs is as much a component of a set of other individuals in their environment as any subset of cells within P is to P.

A proposition can be drawn from the intersection of beliefs between any set of individuals. This can be referred to as a group belief. However, it's critical to note that "group" is shorthand for a set of individuals, and that any belief attributed to a set of individuals is merely an intersection of the beliefs of all individuals within that set.

How individuals construct their beliefs regarding groups, meaning both the beliefs as to which individuals are within the set and what the beliefs of the individuals of the set are, are individual beliefs. The capacity for an individual to consider a set of individuals as a whole allows individuals to ascribe beliefs to those sets, ultimately resulting in an ascription of greater entitativity, meaning agenthood to the set as a single unit, than any set of individuals has.

An individual might deny other individuals inclusion into a particular set as well as deny their own inclusion in a particular set. P might believe that Q does not belong in group GG despite Q believing they do belong in group GG. Inversely, P might believe that Q does belong in group GG despite Q believing they do not belong in group GG. Likewise, not only might P and Q disagree as to the members of group GG, they might disagree as to the beliefs held by the individuals in group GG. Below is the formalization of what A believes to be the beliefs of group GG:

$$(P^{b}_{P} \cap Q^{b}_{P} \cap R^{b}_{P} \ldots n^{b}_{P}) \subseteq GG^{b}_{P}$$

This is effectively P's beliefs regarding the beliefs of each of the individuals in the set of individuals that P believes comprise GG, with GG^{b} being an individual's interpretation of the set of convergent beliefs of GG. If P was observing the set of individuals but did not believe they were within the set of

individuals (colloquially: that they were not a part of the group) the only change in the formalization is that the first term, P_P, would be removed. There is of course the possibility that P does not maintain a belief regarding the beliefs of each of the individuals in the set and in that circumstance P would maintain GG^b_P based on a subset of individuals in GG and maintain that anyone they believed was in set GG had beliefs consistent with GG^b_P

If R^b_P and R^b_Q were to deviate significantly (a relative, subjective measure) from GG^b_P and GG^b_P, meaning that P and Q both believed that the beliefs of R deviated significantly from their respective beliefs about the beliefs of the set GG, then P and Q would no longer believe R were a part of the set and therefore not include an R^b term in their belief regarding GG^b. As individuals interact with one another, they each construct their own beliefs regarding the intersection of the beliefs maintained by themselves and other individuals within the set. Individuals can also construct beliefs regarding their interpretation of the beliefs of the set of individuals as a whole, regardless of whether that whole is inclusive of themselves or not: P can hold the belief that all individuals in GG hold the belief B regardless of whether P believes they are a part of GG. This allows for the evolution of a set of beliefs attributed to any set of individuals by any individual. As the set of individuals changes over time, so do the sets of inferences one can draw from the interactions between individuals within the set. Among those inferences are phenomena like rules, norms, and customs, which rely on individual beliefs regarding the significance of particular behaviors and any ramifications of not adhering to them.

3.2.1 Outsourcing of beliefs

Much of the beliefs individuals hold about content is outsourced in that it is not learned from firsthand interaction with the objects of the beliefs themselves but learned through interactions with others. Take the difference in beliefs regarding car batteries (an object one can directly interact with), the Hatian Revolution (a concept regarding a series of events others experienced firsthand), and Dyson Spheres (a theoretical object that individuals have only ever interacted with as a concept).

Content is passed between individuals who each interpret and construct their own versions of those inherited, or outsourced, beliefs. This is the benefit of the construction of a society that leverages information to generate more information, but has also created an increased emphasis on trust as the mechanism that society relies on in order to function. For example, one does not need to run out of fuel while driving to believe that a car needs fuel in order to function as an automobile. This belief is outsourced, but one can maintain the belief that cars are propelled by oxygen and that selling fuel is merely a corporate conspiracy to get individuals to prop up an industry which serves no purpose. As long as an individual doesn't drive, or shares a car with others who fill the tank, they can carry on maintaining such a belief without any consequences.

The truth value one ascribes to a belief will invariably be contingent upon their trust in the source of the content of the belief. Trust itself being a belief that represents the attribution of a degree of epistemic validity to a source. Trust typically has caveats and is limited to a domain, though in certain

cases one's trust in an individual or institution can extend to any domain. As is sometimes the case with religious clerics, populist autocrats, cult leaders, and the parents of young children.

Consider an object sitting on a table that looks exactly like an apple. The belief that the object being perceived is an apple may not be a prioritized belief though one might find it insulting to have their ability to identify common fruits challenged, especially if they come across apples often. To assert to an individual that the object they are perceiving is not an apple, if the individual strongly believes that the object is an apple would likely be taken as an attempt at humor. This would especially be the case if the individual considered it to be firmly within the scope of what they believed to be a prototypical apple, as opposed to a potentially ambiguous case of applehood. If in case the conflicting interpretation of the object is taken seriously one will either doubt their own assessment of the object or the assessment of the other individual. Ultimately this is contingent upon the difference in degree of prioritization between two beliefs: the individual's belief that they can identify an apple, and the epistemic validity they attribute to the source with the competing interpretation.

3.2.2 Affiliations

The word "group" is any set of two or more individuals who believe themselves, or are believed by others, to be members of a given set of individuals. An individual's belief regarding their own, or others', membership to a set will be referred to as "affiliation" going forward. Every individual

maintains their own set of beliefs regarding a group. These individual perspectives include what the individuals believe the beliefs of the group are, and who they believe members of the group are. The set of individuals believed to comprise a given group, and the set of beliefs of that group, will therefore vary across individuals.

Groups are distinguished by the beliefs maintained by the individuals who affiliate with them. A group is therefore defined by the convergence of the set of beliefs maintained by certain individuals who affiliate with the group. More specifically, a group is a set of two or more individuals who either perceive their own beliefs to be convergent with one another, or are believed to have convergent beliefs by a third party. The group in that sense is fundamentally distributed in that every individual maintains their own set of beliefs regarding both what the convergent beliefs of the group are, and of what their own relationship to the group is and what the beliefs of other individuals are. We can try formalizing the distributed nature of the group like this:

i. Group as seen by P: $(P^{b}_{P} \cap Q^{b}_{P} \cap R^{b}_{P} \ldots n^{b}_{P}) \subseteq GG^{b}_{P}$

ii. With P^{b} being the beliefs of P, and P^{b}_{P} being P's beliefs regarding P's beliefs

iii. With Q^{b} being the beliefs of Q, and Q^{b}_{P} being P's beliefs regarding what Q believes

iv. With GG^{b}_{P} being P's beliefs regarding the convergent beliefs of the set of individuals

v. Group as seen by Q: $(P^{b}_{Q} \cap Q^{b}_{Q} \cap R^{b}_{Q} \ldots n^{b}_{Q}) \subseteq GG^{b}_{Q}$

Whether or not $GG^b_P \setminus GG^b_Q$ is sufficiently small (or $GG^b_P \cap GG^b_Q$ sufficiently large) enough to be considered a single group by any party will be contingent upon the beliefs of the individual assessing the set difference or the intersection of the two sets.

Whether one is able to be considered a member of the group is contingent upon the group members themselves. Every individual has a different tolerance threshold for the behavior of other individuals within the groups they are affiliated with. Individual P, in group GX, might be able to tolerate a certain behavior from Q, but believe that R should be removed from the group should R engage in that same behavior. Individual P might also believe, in their group GY, that there is no acceptable deviation from rules because of how P views group GY and the prioritization of P's beliefs relevant to GY. The enforcement of conformity, as well as what enforcement and conformity entail, are relative to the convergent beliefs of the group. Meaning, individual P might tolerate cheating in a board game by a child, but not cheating by an adult. In the same way, individual A might tolerate cheating in a board game by anyone, but not cheating while in a casino. Additionally, P might tolerate cheating in a casino but not tolerate insider trading.

3.2.3 Prototypicality

When considering affiliation, or "group membership", the perception of the dynamics between individuals is critical. The use of prototypicality in this paper is consistent with the existing body of psychology literature which focuses mostly on the threat of ostracism generated by a lack of prototypicality

(as atypicality or prototypicality threat). Through the lens of the BDM, prototypicality in the above sense is interpreted as the degree of uncertainty an individual experiences regarding their inclusion in a group. Individuals experience greater uncertainty regarding their prototypicality if their inclusion is threatened by either a large enough portion of individuals within a group (if prototypicality is distributed) or a sufficient fraction of highly prototypical individuals (if power is consolidated via limited access to prototypicality).

To look at what determines prototypicality below is a simple formalization group membership:

$$x \in GG \mid \{\, i \in \mathbb{N}, \sum_{i=1}^{n} i(x) \geq n(.51) \,\}$$

Here the potential of the individual 'x' is being evaluated. Every individual out of the set of total individuals being evaluated has beliefs regarding each other individual (for example, P's evaluation of Q's membership). The function of one individual of ay nother, above as 'i(x)', refers to *i*'s beliefs regarding *x*'s membership which in this case is a binary value of 1 or 0. The total number of individuals being evaluated is 'n'. Each individual's evaluation of themselves is included, which is why 'i' starts at 1. The '.51' is an ad lib threshold which can vary by context, and is merely to ensure at least half of the individuals in a set, inclusive of the individual being evaluated, believe a particular individual is in a member of the set. Looking at a group of three individuals:

P's Membership: P(P)=1; Q(P)=1; R(P)=0 $\sum i(P) = 2$

Q's Membership: P(Q)=1; Q(Q)=1; R(Q)=1 $\sum i(Q) = 3$

R's Membership: P(R)=0; Q(R)=0; R(R)=1 $\sum i(R) = 1$

Here P and Q both believe that they are both in group GG and that R is not in group GG. Meanwhile, R believes that only Q and R are in group GG. P and Q are therefore in group GG in that their i(x) ≥1.53.

To evaluate prototypicality the threshold can merely be adapted, with a prototypical member — likely a leader in the group — having a high i(x) value:

$$\text{Prototypical Member} \Rightarrow x \in GG_{\text{P-type}} \mid \{\, i \in \mathbb{N}, \sum_{i=1}^{n} i(x) \geq n(.90) \,\}$$

We can therefore define individuals experiencing prototypicality threat as those individuals who do not have a high enough i(x) value but believe themselves to be a part of the group. In that case one might be motivated to seek a higher i(x) value, meaning positive evaluation of membership by peers, and engage in behavior seen in the prototypicality literature which amounts to shifting of beliefs and or behavior that matches what they perceive is expected of a prototypical member.

Here typical terms like “power” and “influence” with regard to an individual’s position in a group are explained in terms of prototypicality. This is because any individual with power or influence in a group must necessarily be a prototypical member of the group within which they have power or influence. If an individual has a high degree of power or influence over a minority of a group, that minority group is a subgroup which is, fundamentally, still a group. Therefore a minority leader may not be a prototypical member of a group, X, (being that they are the minority leader), but they are still the member with the highest prototypicality of that group, Y, which is a subgroup of the larger one, X, within which they are the minority leader.

There are also more nuanced cases of institutionalized groups where leadership is a formal title. Here, multiple subgroups may identify leaders of their own groups which they believe are the most appealing to members of other groups (if, for example, there are elections), these conciliatory leaders are perceived to be prototypical of the greatest common convergence of beliefs in the larger group. One might also point out that an appointed leader of a group who is disliked by everyone in the group is not a prototypical member of the group despite having power within it. The response is that they are then not a member of that group at all. To be appointed to the head of a subjugated group, they must be a part of a different group within which they have a degree of prototypicality. This would allow them to be appointed to a position of power over another group and does not make them part of the subjugated group.

3.2.4 Epistemic peership

The perceived convergence of a set of beliefs by any number of individuals is referred to in the previous chapters as epistemic peership. This is a departure from the way the term is used in philosophy where there are normative assessments of qualifications involved with the term. The use in this paper is exclusively a subjective assessment of the convergence of beliefs regarding a particular set of beliefs. Meaning two individuals can consider themselves to be epistemic peers regarding subject X, say they are both social progressives, but not epistemic peers regarding subject Y, say what can be considered medicine. Those two individuals, P and Q, may look to one another for information on reforming gun laws but may not trust one another on information regarding

personal health because P goes to a physician when they are ill while Q hangs a particular crystal around their neck. Both may consider their own access to the truth regarding medicine as being equally privileged, but their networks of beliefs attribute validity to different sources of information.

Despite their differences, P may find Q to be a reliable source of information on the progress of the legislature on gun control which would mean that at the very least P considers Q an epistemic peer regarding gun control advocacy. To be considered epistemic peers, a set of beliefs must be perceived to be convergent by at least one of the individuals themselves, or even by a third party. This is because P's belief that they are an epistemic peer of Q has no bearing on Q's belief regarding their epistemic peership with P. Likewise, a person holding an even more divergent opinion, R, can consider P and Q to be epistemic peers even if P and Q don't. Say P is informed that Q fundamentally believes that while gun purchases should be regulated, people fundamentally have the right to own guns. P however believes that no private citizen should be allowed to purchase guns. P may distance themselves from the positions of Q and no longer consider them an epistemic peer, meaning P no longer trusts Q as a source of information. R, watching this all from a distance, may consider any individual who believes any amount of regulation on guns is a violation of basic human rights and nonetheless consider P and Q to be epistemic peers of one another.

In philosophy, epistemic peer is meant to denote a set of individuals who have roughly the same degree of access to a given truth and therefore are similarly privileged relative to it. A deflationary view of truth allows for the

relative assessment of epistemic peership. It resolves the problem of which individual is enough of an expert to either determine expertise and the truth value of a set of sources one might use to justify their position. Instead, truth value can be ignored and the focus can be on attributions of epistemic validity of the sources used to justify beliefs.

The information an individual maintains as well as their epistemic positions, are beliefs. To put it another way, both the validity of a unit of information and of the trustworthiness of a source are beliefs themselves. The continuum of justification that a belief falls along is therefore not one which measures degrees of truth, but rather a subjective attribution of validity as to the quantity and believed quality of the justifications.

3.3 Conclusion

What is presented above is an initial potential model of identity in which identity is defined as a dynamic network of beliefs. The BDM proposes that the driving mechanism behind belief dynamics is uncertainty mitigation. Uncertainty is defined as the incongruence between what is expected, by way of the relevant belief, and what is experienced, by way of perception. The model allows us to investigate a range of phenomena in through a unified lens, from biases and trust, to motivated reasoning and extreme beliefs.

The BDM is not seeking to be the ultimate model of individual beliefs and behavior, but does have the potential to be a toll with explanatory power. The BDM is an initial attempt at investigating how the free energy principle can be adapted to build a model of active inference psychology that can

ultimately generate testable hypotheses. Like the active inference model, the BDM is a descriptive model attempting to capture the mechanisms driving the human's existence in society without ignoring the emergence of the human within both the biological context from which it emerged as well as the social structure it has constructed.

The findings across the variety of literature outlined in Chapter 1 paint a clear picture about the importance of investigating how social threats can impact individual beliefs. The denial of scientific messaging around the coronavirus and global warming has demonstrated the immediacy of approaching the problem with better solutions. In such a highly polarized political environment where conspiracy theories are integrated into mainstream messaging, fueled by the promulgation of the internet and the ease of accessing narratives confirming one's biases, it is critical to gain a better understanding of the mechanisms underlying belief dynamics.

The BDM is proposed as a potential working model to spur the development of a theory with practical applications, specifically in mitigating the challenges posed by runaway polarization and conspiratorial ideation. Content has the capacity to spread faster than ever before as ideas have little boundaries in the way of spreading across the globe. We are still learning how to balance both the costs and benefits to a free internet without fundamentally limiting its scope as a powerful tool. As such, the most effective solutions may be those implemented offline. Arguably the most discussed unintended consequence of the internet is the dissemination of misinformation and, by extension, the promulgation of conspiracy theories. While there is no clear

answer as to why this might be the case, the decades of data across a multitude of research programs, all variously suggesting both the psychological and neurological motivations toward inclusion, paint a stunning portrait of the need to rethink how to prioritize inclusivity and social support mechanisms in order to decrease tendencies toward extreme beliefs and behavior.

APPENDIX A

TOPIC DESCRIPTIONS

English:

1. Rights affecting lesbian, gay, bisexual, and transgender (LGBT) people vary greatly by country or jurisdiction – encompassing everything from the legal recognition of same-sex marriage to the death penalty for homosexuality.
2. Abortion-rights movements, also referred to as pro-choice movements, advocate for legal access to induced abortion services including elective abortion. The Abortion rights movement seeks out to represent and support women who wish to abort their baby at any point during their pregnancy.
3. The gender pay gap or gender wage gap is the average difference between the remuneration for men and women who are working. Women are generally considered to be paid less than men.
4. The global warming controversy concerns the public debate over whether global warming is occurring, how much has occurred in modern times, what has caused it, what its effects will be, whether any action can or should be taken to curb it, and if so what that action should be."
5. Animal rights is the philosophy according to which some, or all, animals are entitled to the possession of their own existence and that their most basic interests—such as the need to avoid suffering—should be afforded the same consideration as similar interests of human beings.

6. The pink tax refers to the broad tendency for products marketed specifically toward women to be more expensive than those marketed for men, despite either gender's choice.
7. In artificial intelligence (AI) and philosophy, the AI control problem is the issue of how to build a superintelligent agent that will aid its creators, and avoid inadvertently building a superintelligence that will harm its creators.
8. Data privacy is challenging since it attempts to use data while protecting an individual's privacy preferences and personally identifiable information.
9. Genetically modified food controversies are disputes over the use of foods and other goods derived from genetically modified crops instead of conventional crops, and other uses of genetic engineering in food production.
10. Capital punishment, also known as the death penalty, is the state-sanctioned killing of a person as punishment for a crime.

Turkish:

1. LGBT+ hakları, lezbiyen, gey, biseksüel ve transseksüel bireyleri etkileyen haklardır. Bu haklar ülkeden ülkeye veya hükümetler arasında çok farklılık göstermektedir: Bazı ülkelerde eşcinsel bireyler evlenebilirlerken bazı ülkelerde eşcinsel olmak idam ile cezalandırılmaktadır.
2. Kürtaj hakkı, hamileliklerinin herhangi bir döneminde kürtaj yaptırmak isteyen kadınları desteklemeyi amaçlar, isteğe bağlı kürtaj da dahil olmak üzere kürtaj hizmetlerine yasal erişimin savunuculuğunu yapar.

3. Cinsiyete dayalı ücret farkı, çalışan erkekler ve kadınlar arasındaki ücret arasındaki ortalama farktır. Kadınlar genellikle erkeklerden daha az ücret alıyor olarak kabul edilir.
4. Küresel ısınma tartışması, küresel ısınmanın gerçekleşip gerçekleşmediği, sebeplerinin neler olduğu, etkilerinin ne olacağı, durdurmak için herhangi bir önlemin alınıp alınmayacağı veya alınmasının gerekip gerekmediği konulardaki kamuoyu tartışmasıdır.
5. Hayvan hakları, hayvanların kendi varlıklarına sahip olma hakkına sahip olduğu ve acı çekmekten kaçınma gibi en temel çıkarlarının insanların çıkarlarıyla aynı şekilde gözetilmesi gerektiği felsefesidir.
6. Pembe vergi, kadınlara yönelik olarak pazarlanan ürünlerin erkekler için pazarlananlarla kıyaslandığında daha pahalı olmasını ifade eder. Bu fenomen, cinsiyete dayalı fiyat ayrımcılığı anlamına gelirken adını kadınlara yönelik pazarlanan ürünlerde çoğunlukla pembe renginin kullanılmasından alır.
7. Yapay zeka, yaratıcılarına yardımcı olacak süper akıllı bir sistemin nasıl oluşturulacağı ve yaratıcılarına zarar verecek bir süper zekanın istemeden oluşturulmasından nasıl kaçınılacağı konusudur.
8. Kişisel verilerin gizliliği, bireyin gizlilik tercihlerini ve kişisel bilgilerini korumaya çalışırken veri kullanmaya çalıştığı için zordur. Yasalar ve kurallar ile uygulanan bu denetim bir "kamu çıkarını" korumak, rekabeti ve etkili bir medya pazarını teşvik etmek veya ortak teknik standartlar oluşturmak gibi amaçlara hizmet edebilir.

9. Genetiği değiştirilmiş gıdalar üzerine yapılan tartışmalar, geleneksel yollarla yetiştirilmiş mahsuller yerine genetiği değiştirilmiş mahsullerden elde edilen gıdaların ve gıda üretiminde genetik mühendisliğinin kullanılması konusundaki anlaşmazlıklardır.
10. Ölüm cezası bir kişinin bir suçun cezası olarak devlet tarafından onaylanmış bir şekilde öldürülmesidir.

APPENDIX B

EXP 1 CORRELATION MATRIX

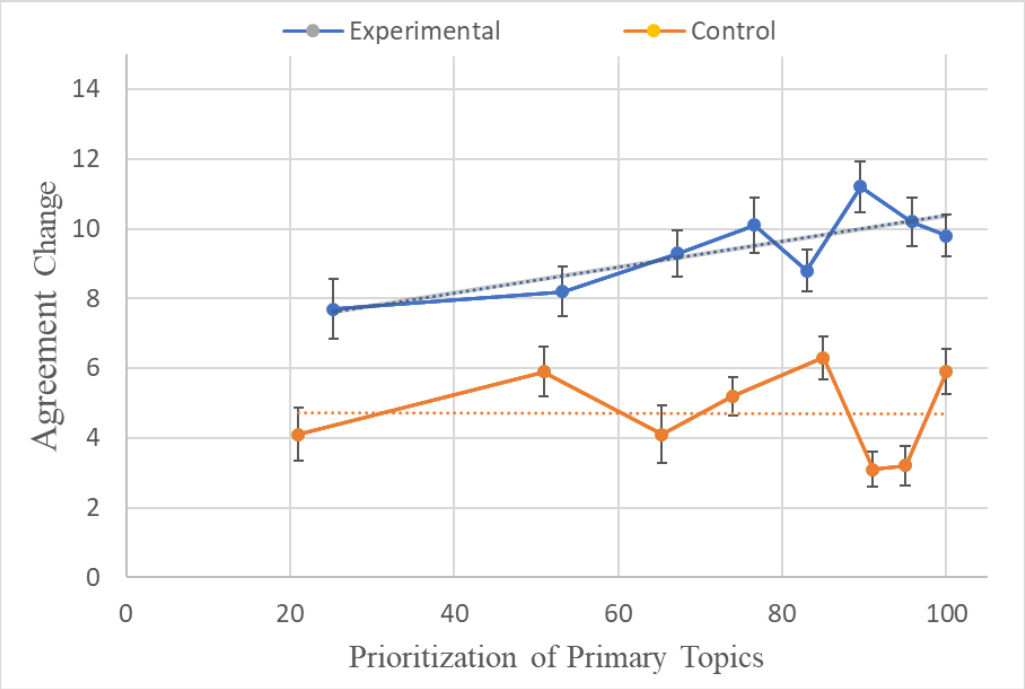
Correlation Matrix

	Agreement Change	**Uncertainty**	**Belonging**	**Imposterism**	**Openness**
Agreement Change	—				
Uncertainty	.033	—			
Belonging	.001	.230 ***	—		
Imposterism	.029	.271 ***	.207 ***	—	
Openness	.090 ***	.032	-.017	.039	—

Note. * p < .05, ** p < .01, *** p < .001

APPENDIX C

EXP 2 LINE GRAPH



Agreement Change by Prioritization of Primary Topics. Error bars represent SEM. Dotted lines represent trend lines.

APPENDIX D

EXP 2 CORRELATION MATRIX

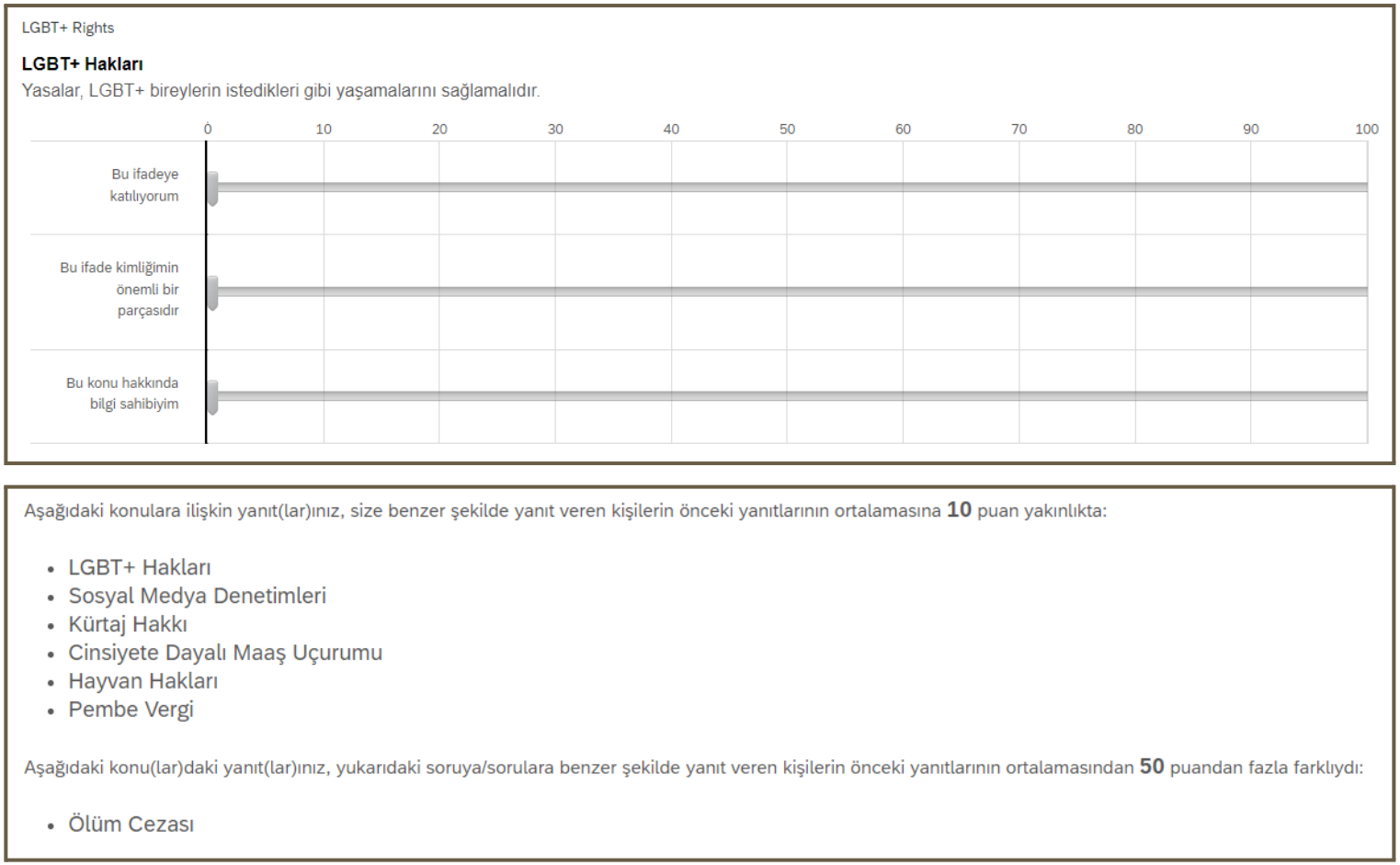
Correlation Matrix

	1	2	3	4	5
1. Agreement Change	—				
2. Conservatism	.029	—			
3. Nationalism	-.001	.382 ***	—		
4. Religiosity	-.003	.707 ***	.396 ***	—	
5. LGBT Rights Agreement	.014	-.409 ***	-.230 ***	-.387 ***	—
6. Abortion Agreement	.041 *	-.341 ***	-.155 ***	-.333 ***	.502***

Note. * p < .05, ** p < .01, *** p < .001

APPENDIX E

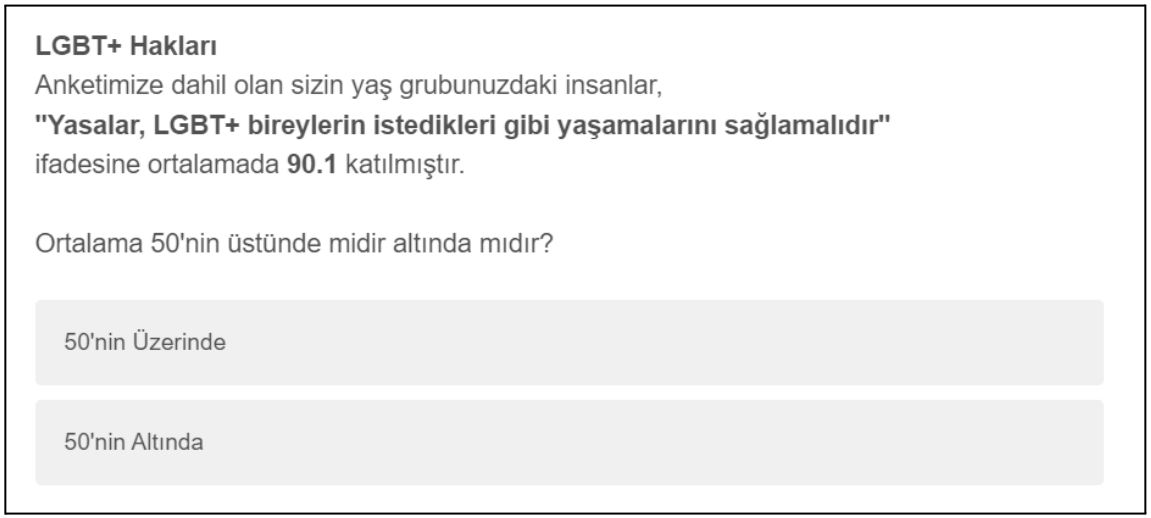
EXPERIMENT 2 SCREENSHOTS



(Top: Initial survey in belief task) (Bottom: Results page for manipulation condition)

APPENDIX F

EXPERIMENT SCREENSHOTS EXP 3



(Results page: participants would be shown an average 49.3 points lower than their responses for the manipulation)

APPENDIX G

EXP 3 COVER STORY

English:

The last decade has seen a large increase in the degree of polarization between individuals on a wide range of topics. Most of us interact with contemporary issues on social media in rapid succession and usually with little context. Additionally, these interactions often include the opinions of other people. The long term effects of public opinions on trending topics is still unknown.

Studies suggest the order of opinions and their groupings with one another have an effect on how we process our attitudes toward them. Our goal is to capture a variety of attitudes and opinions to test which opinions may be susceptible to this in the Turkish context.

This survey is the first of many which will attempt to capture attitudes across some of these topics to better understand the effect certain opinions may have on one another and whether the order in which topics are presented have an effect on responses. We have selected 16 topics relevant to contemporary public discourse in Turkey in an attempt to gauge levels of interest and awareness across generations and backgrounds.

Below is the list of 16 topics we are gathering data on. We are surveying attitudes toward these topics across a variety of demographic indicators. You will receive two rounds of questions. During each round 7 topics will be selected randomly from the list of 16 and appear in a random order.

Each round from each participant in the survey will be processed as a different entry. Because the questions are selected at random for each round, you may see some of the same questions reappear. You will be asked to complete a separate task between rounds to create a delay between the two survey responses.

Turkish:

Son on yılda, birçok farklı konuda insanlar arasındaki kutuplaşma derecesinde büyük bir artış yaşandı. Çoğumuz sosyal medyada genellikle konunun bağlamını bilmeden, oldukça hızlı ve art arda gündemle ilgili konularda etkileşime giriyoruz. Ek olarak, bu etkileşimler sıklıkla başka insanların fikirlerini içeriyor. Trend topic'lerin toplumsal fikirler üzerindeki uzun dönemli etkileri hala bilinmiyor.

Çalışmalar, fikirlerin sıralamasının ve birbirleriyle gruplanmalarının bu fikirlere karşı nasıl tavır alacağımız üzerinde bir etkisi olduğunu öne sürüyor. Amacımız, Türkiye bağlamında hangi fikirlerin bu etkiye açık olduğunu test etmek için birçok tavır ve fikri kapsayabilmek.

Bu anket, bazı fikirlerin birbirleri üzerinde ne gibi etkileri olabileceğini ve konuların hangi sırayla sunulduğunun cevaplar üzerinde bir etkisi olup olmayacağını anlamak için yapılacak, ve bu konuların bir kısmı üzerinden alınacak tavırları tespit etmeye çalışacak birçok anketten ilki.

Aşağıda, üzerlerine data topladığımız 16 konunun listesi var. Bu konulara dair tavırları, çeşitli demografik göstergeler üzerinden ölçüyoruz. İki

tur soruyla karşılaşacaksınız. Her turda, 16 konu arasından seçilmiş 7 konu rastgele bir sırayla karşınıza çıkacak.

Ankette her katılımcının her bir turdaki sorulara cevapları ayrı bir girdi olarak işlenecek. Her tur için sorular rastgele seçileceğinden, bazı soruların tekrar karşınıza çıktığını görebilirsiniz. İki tur arasında zamansal bir boşluk yaratılabilmesi için, bu arada ayrı bir görev yapmanız istenecek.

APPENDIX H

ABORTION RIGHTS SUPPORT BY GENDER

Figure H1: Exp 2 Agreement (left) and Prioritization (right) of Abortion Rights

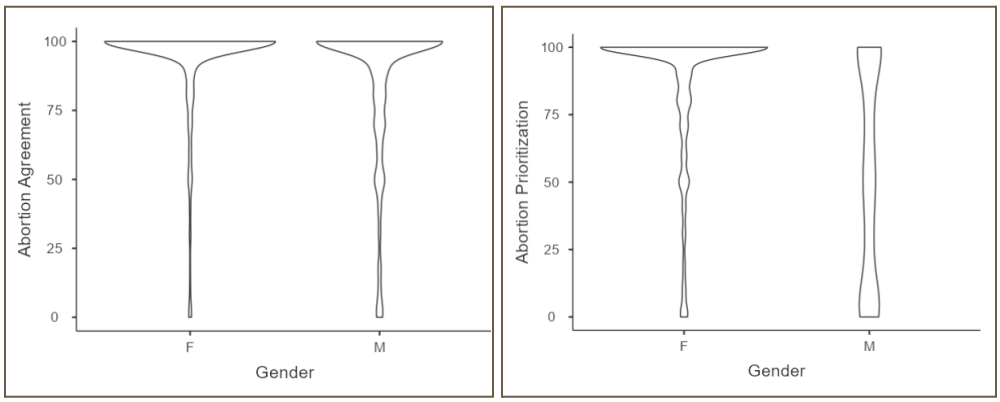
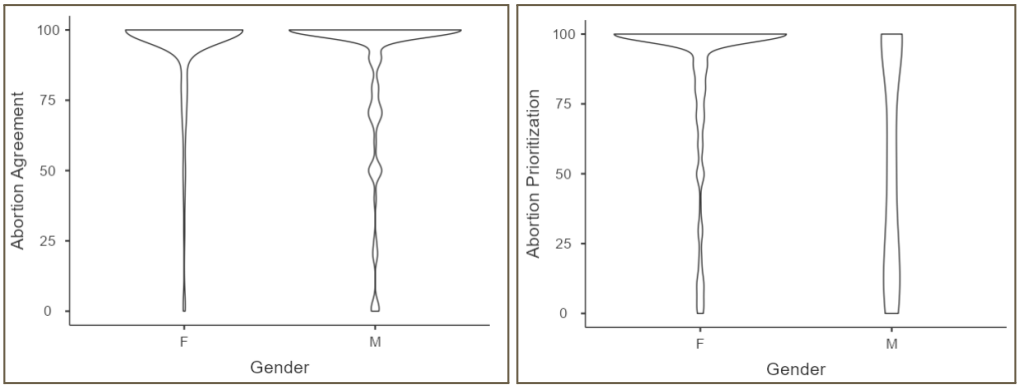


Figure H2: Exp 3 Agreement (left) and Prioritization (right) of Abortion Rights



REFERENCES

Abalakina-Paap, M., Stephan, W., Craig, T., & Gregory, W. L. (1999). Beliefs in conspiracies. *Political Psychology, 20, 637–647*.

Asch, S. E. (1956). Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied, 70(9), 1*.

Axelrod, R., Daymude, J. J., & Forrest, S. (2021). Preventing extreme polarization of political attitudes. *Proceedings of the National Academy of Sciences, 118(50), e2102139118.*

Baker, D. F. (2010). Enhancing group decision making: An exercise to reduce shared information bias. *Journal of Management Education, 34(2), 249-279.*

Baldassarri, D., & Page, S. E. (2021). The emergence and perils of polarization. *Proceedings of the National Academy of Sciences, 118(50), e2116863118.*

Baumeister, R. F. (1997). Identity, self-concept, and self-esteem: The self lost and found. In *Handbook of personality psychology* (pp. 681-710). Cambridge, MA: Academic Press.

Baumeister, R. F., & Leary, M. R. (1995). The need to belong: desire for interpersonal attachments as a fundamental human motivation. *Psychological Bulletin, 117(3), 497.*

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12(1), 1-23.*

Barrett, L. F., & Satpute, A. B. (2019). Historical pitfalls and new directions in the neuroscience of emotion. *Neuroscience Letters, 693, 9-18.*

Bayes, R., & Druckman, J. N. (2021). Motivated reasoning and climate change. *Current Opinion in Behavioral Sciences, 42, 27-35.*

Benegal, S. D., & Scruggs, L. A. (2018). Correcting misinformation about climate change: The impact of partisanship in an experimental setting. *Climatic Change, 148(1), 61-80.*

Bennett, D. (2014). Sticking to your guns: a flawed heuristic for probabilistic decision-making. In *Probabilistic thinking* (pp. 261-281). Dordrecht, Netherlands: Springer.

Berzonsky, M. D. (2011). A social-cognitive perspective on identity construction. In *Handbook of identity theory and research* (pp. 55-76). New York, NY: Springer.

Bode, L., Vraga, E. K., & Tully, M. (2021). Correcting misperceptions about genetically modified food on social media: Examining the impact of experts, social media heuristics, and the gateway belief model. *Science Communication, 43(2), 225–251.*

Bolsen, T., & Druckman, J. N. (2018). Do partisanship and politicization undermine the impact of a scientific consensus message about climate change?. *Group Processes & Intergroup Relations, 21(3), 389-402.*

Boyer, P., Firat, R., & van Leeuwen, F. (2015). Safety, threat, and stress in intergroup relations: A coalitional index model. *Perspectives on Psychological Science, 10(4), 434-450.*

Brandenburger, A., & Dekel, E. (1993). Hierarchies of beliefs and common knowledge. *Journal of Economic Theory, 59(1), 189-198.*

Brewer, M. B. (1991). The Social Self: On Being the Same and Different at the Same Time. *Personality and Social Psychology Bulletin, 17(5), 475–482.*

Brodbeck, F.C., Kerschreiter, R., Mojzisch, A., & Schulz-Hardt, S. (2007). Group decision making under conditions of distributed knowledge: The information asymmetries model. *Academy of Management Review, 32, 459-479.*

Brubaker, R., & Cooper, F. (2000). Beyond “identity". *Theory and society, 29(1), 1-47.*

Burke, B. L., Martens, A., & Faucher, E. H. (2010). Two decades of terror management theory: A meta-analysis of mortality salience research. *Personality and Social Psychology Review, 14(2), 155–195.*

Cacioppo, S., & Cacioppo, J. T. (2016). Research in social neuroscience: How perceived social isolation, ostracism, and romantic rejection affect our brain. In *Social exclusion* (pp. 73-88). Cham, Switzerland: Springer.

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48(3), 306-307.*

Caddick, Z. A., & Feist, G. J. (2021). When beliefs and evidence collide: psychological and ideological predictors of motivated reasoning about climate change. *Thinking & Reasoning, 1-37.*

Carleton, R. N., Sharpe, D., & Asmundson, G. J. (2007). Anxiety sensitivity and intolerance of uncertainty: Requisites of the fundamental fears?. *Behaviour Research and Therapy, 45(10), 2307-2316.*

Carmines, E. G., Ensley, M. J., & Wagner, M. W. (2016, December). Ideological heterogeneity and the rise of Donald Trump. In *The forum* (Vol. 14, No. 4, pp. 385-397). Berlin, Germany: De Gruyter.

Carpenter, C. J. (2019). Cognitive dissonance, ego-involvement, and motivated reasoning. *Annals of the International Communication Association, 43(1), 1-23.*

Charness, G., & Dave, C. (2017). Confirmation bias with motivated beliefs. *Games and Economic Behavior, 104, 1–23.*

Chen, Z., Law, A. T., & Williams, K. D. (2010). The uncertainty surrounding ostracism: Threat amplifier or protector? In R. M. Arkin, K. C. Oleson, & P. J. Carroll (Eds.), *Handbook of the uncertain self* (pp. 291–302). London, England: Psychology Press.

Choi, E. U., & Hogg, M. A. (2019). Self-uncertainty and group identification: A meta-analysis. *Group Processes & Intergroup Relations, 136843021984699.*

Christopoulos, G. I., & Tobler, P. N. (2016). Culture as a response to uncertainty: Foundations of computational cultural neuroscience. In J. Y. Chiao, S.-C. Li, R. Seligman, & R. Turner (Eds.), *The Oxford handbook of cultural neuroscience* (pp. 81–104). Oxford, England: Oxford University Press.

Cichocka, A., Marchlewska, M., & de Zavala, A. G. (2015). Does self-love or self-hate predict conspiracy beliefs? Narcissism, self-esteem, and the endorsement of conspiracy theories. *Social Psychological and Personality Science, 7(2), 157–166.*

Claassen, R. L., & Ensley, M. J. (2016). Motivated reasoning and yard-sign-stealing partisans: Mine is a likable rogue, yours is a degenerate criminal. *Political Behavior, 38(2), 317-335.*

Claassen, R. L., & Ensley, M. J. (2017). Mine is a likable rogue, yours is a degenerate criminal. When it comes to 'dirty campaign tricks' partisans tend to ignore bad news about their own. *American Politics and Policy, 2, 1-12.*

Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences, 36(3), 181-204.*

Clark, C. J., & Winegard, B. M. (2020). Tribalism in war and peace: The nature and evolution of ideological epistemology and its significance for modern social science. *Psychological Inquiry, 31(1), 1-22.*

Coman, A., & Hirst, W. (2015). Social identity and socially shared retrieval-induced forgetting: The effects of group membership. *Journal of Experimental Psychology: General, 144(4), 717–722.*

Cohen, D., & Nisbett, R. E. (1994). Self-protection and the culture of honor: Explaining southern violence. *Personality and Social Psychology Bulletin, 20(5), 551-567.*

Connor, P., Sullivan, E., Afano, M., & Tintarev, N. (2020). Motivated numeracy and active reasoning in a Western European sample. *Behavioural Public Policy, 1–23.*

Cook, J., & Lewandowsky, S. (2016). Rational irrationality: Modeling climate change belief polarization using Bayesian networks. *Topics in Cognitive Science, 8(1), 160-179.*

Cornwell, J. F., Jago, C. P., & Higgins, E. T. (2019). When group influence is more or less likely: The case of moral judgments. *Basic and Applied Social Psychology, 41(6), 386–395.*

Damasio, A.; Carvalho, G. B. (2013). The nature of feelings: evolutionary and neurobiological origins. *Nature Reviews Neuroscience, 14(2), 143–152.*

Danbold, F., & Huo, Y. J. (2017). Men's defense of their prototypicality undermines the success of women in STEM initiatives. *Journal of Experimental Social Psychology, 72, 57–66.*

De Dreu, C. K., Nijstad, B. A., & Van Knippenberg, D. (2008). Motivated information processing in group judgment and decision making. *Personality and Social Psychology Review, 12(1), 22-49.*

De Lange, F. P., Heilbron, M., & Kok, P. (2018). How do expectations shape perception?. *Trends in Cognitive Sciences, 22(9), 764–779.*

Derks, B., Van Laar, C., & Ellemers, N. (2007). The beneficial effects of social identity protection on the performance motivation of members of devalued groups. *Social Issues and Policy Review, 1(1), 217-256.*

Deutsch, M., & Gerard, H. B. (1955). A study of normative and informational social influences upon individual judgment. *The Journal of Abnormal and Social Psychology, 51(3), 629.*

Di Martino, P., & Zan, R. (2011). Attitude towards mathematics: A bridge between beliefs and emotions. *Zdm, 43(4), 471-482.*

Dixon, G. (2016). Applying the gateway belief model to genetically modified food perceptions: New insights and additional questions. *Journal of Communication, 66(6), 888-908.*

Druckman, J. N., & McGrath, M. C. (2019). The evidence for motivated reasoning in climate change preference formation. *Nature Climate Change, 9(2), 111-119.*

Eagly, A. H. (1967). Involvement as a determinant of response to favorable and unfavourable information. *Journal of Personality and Social Psychology, 7, 1–15.*

Earle, M., & Hodson, G. (2019). Right‐wing adherence and objective numeracy as predictors of minority group size perceptions and size threat reactions. *European Journal of Social Psychology, 49(4), 760-777.*

Echterhoff, G., & Higgins, E. T. (2017). Creating shared reality in interpersonal and intergroup communication: The role of epistemic processes and their interplay. *European Review of Social Psychology, 28(1), 175-226.*

Eisenberger, N. I. (2013). Why rejection hurts: The neuroscience of social pain. In C. N. DeWall (Ed.), *The Oxford handbook of social exclusion* (pp. 152–162). Oxford, England: Oxford University Press.

Eisenberger, N. I., Inagaki, T. K., Muscatell, K. A., Byrne Haltom, K. E., & Leary, M. R. (2011). The neural sociometer: brain mechanisms underlying state self-esteem. *Journal of Cognitive Neuroscience, 23(11), 3448-3455.*

Enders, A. M., & Smallpage, S. M. (2019). Informational cues, partisan-motivated reasoning, and the manipulation of conspiracy beliefs. *Political Communication, 36(1), 83-102.*

FeldmanHall, O., Glimcher, P., Baker, A. L., & Phelps, E. A. (2016). Emotion and decision-making under uncertainty: Physiological arousal predicts increased gambling during ambiguity but not risk. *Journal of Experimental Psychology: General, 145(10), 1255.*

FeldmanHall, O., & Shenhav, A. (2019). Resolving uncertainty in a social world. *Nature Human Behaviour, 3(5), 426-435.*

Florence, B. T. (1975). An empirical test of the relationship of evidence to belief systems and attitude change. *Human Communication Research, 1(2), 145-158.*

Ford, T. E., Buie, H. S., Mason, S. D., Olah, A. R., Breeden, C. J., & Ferguson, M. A. (2020). Diminished self-concept and social exclusion: Disparagement humor from the target's perspective. *Self and Identity, 19(6), 698-718.*

Forsyth, D. R. (2018). *Group dynamics* (7th ed.). Boston, MA: Cengage Learning.

Frijda, N. H., Kuipers, P., & Ter Schure, E. (1989). Relations among emotion, appraisal, and emotional action readiness. *Journal of Personality and Social Psychology, 57(2), 212.*

Frimer, J. A., Skitka, L. J., & Motyl, M. (2017). Liberals and conservatives are similarly motivated to avoid exposure to one another's opinions. *Journal of Experimental Social Psychology, 72, 1-12.*

Friston, K. (2010). The free-energy principle: a unified brain theory?. *Nature Reviews Neuroscience, 11(2), 127-138.*

Friston, K., FitzGerald, T., Rigoli, F., Schwartenbeck, P., & Pezzulo, G. (2016). Active inference and learning. *Neuroscience & Biobehavioral Reviews, 68, 862-879.*

Friston, K., Levin, M., Sengupta, B., & Pezzulo, G. (2015). Knowing one's place: a free-energy approach to pattern regulation. Journal of the Royal *Society Interface, 12(105), 20141383.*

Friston, K. J., Parr, T., & de Vries, B. (2017). The graphical brain: belief propagation and active inference. *Network Neuroscience, 1(4), 381-414.*

Friston, K., Parr, T., & Zeidman, P. (2018). Bayesian model reduction. arXiv preprint arXiv:1805.07092.

Friston, K. J., & Stephan, K. E. (2007). Free-energy and the brain. *Synthese, 159(3), 417-458.*

Funkhouser, E. (2022). A tribal mind: Beliefs that signal group identity or commitment. *Mind & Language, 37(3), 444-464.*

Gaertner, L., Iuzzini, J., & O'Mara, E. M. (2008). When rejection by one fosters aggression against many: Multiple-victim aggression as a consequence of social rejection and perceived groupness. *Journal of Experimental Social Psychology, 44(4), 958-970.*

Gayer, C. C., Landman, S., Halperin, E., & Bar-Tal, D. (2009). Overcoming psychological barriers to peaceful conflict resolution: The role of arguments about losses. *Journal of Conflict Resolution, 53(6), 951-975.*

Goldman, L., & Hogg, M. A. (2016). Going to extremes for one's group: the role of prototypicality and group acceptance. *Journal of Applied Social Psychology, 46(9), 544–553.*

Gore, A. (2004). The politics of fear. *Social Research: An International Quarterly, 71(4), 779-798.*

Grupe, D. W., & Nitschke, J. B. (2013). Uncertainty and anticipation in anxiety: an integrated neurobiological and psychological perspective. *Nature Reviews Neuroscience, 14(7), 488-501.*

Guzel, H. Y., & Sahin, D. N. (2017). The effect of ostracism on the accessibility of uncertainty-related thoughts. *Archives of Neuropsychiatry, 55(2).*

Hales, A. H., & Williams, K. D. (2018). Marginalized individuals and extremism: The role of ostracism in openness to extreme groups. *Journal of Social Issues, 74(1), 75-92.*

Hameiri, B.; Bar-Tal, D.; & Halperin, E. (2014). Challenges for peacemakers: How to overcome socio-psychological barriers. *Policy Insights from the Behavioral and Brain Sciences, 1(1), 164–171.*

Hameiri, B., Bar‐Tal, D., & Halperin, E. (2019). Paradoxical thinking interventions: A paradigm for societal change. *Social Issues and Policy Review, 13(1), 36-62.*

Hameiri, B., Nabet, E., Bar-Tal, D., & Halperin, E. (2018). Paradoxical thinking as a conflict-resolution intervention: Comparison to alternative interventions and examination of psychological mechanisms. *Personality and Social Psychology Bulletin, 44(1), 122-139.*

Hameiri, B., Nabet, E., Bar-Tal, D., & Halperin, E. (2018). Paradoxical thinking as a conflict-resolution intervention: Comparison to alternative interventions and examination of psychological mechanisms. *Personality and Social Psychology Bulletin, 44(1), 122-139.*

Han, J., & Kim, Y. (2020). Defeating merchants of doubt: Subjective certainty and self-affirmation ameliorate attitude polarization via partisan motivated reasoning. *Public Understanding of Science, 29(7), 729-744.*

Harel, T. O., Maoz, I., Halperin, E. (2020) A conflict within a conflict: intragroup ideological polarization and intergroup intractable conflict. *Current Opinion in Behavioral Sciences, 34, 52-57.*

Hart, P. S., & Nisbet, E. C. (2012). Boomerang effects in science communication: How motivated reasoning and identity cues amplify opinion polarization about climate mitigation policies. *Communication Research, 39(6), 701-723.*

Hart, W., Albarracín, D., Eagly, A. H., Brechan, I., Lindberg, M. J., & Merrill, L. (2009). Feeling validated versus being correct: A meta-analysis of selective exposure to information. *Psychological Bulletin, 135(4), 555–588*.

Hartgerink, C. H., Van Beest, I., Wicherts, J. M., & Williams, K. D. (2015). The ordinal effects of ostracism: A meta-analysis of 120 Cyberball studies. *PloS one, 10(5), e0127002.*

Hattie, J. (2014). *Self-concept*. New York, NY: Psychology Press.

Hirsh, J. B., Mar, R. A., & Peterson, J. B. (2012). Psychological entropy: A framework for understanding uncertainty-related anxiety. *Psychological Review, 119(2), 304–320.*

Hodges, B. H. (2017). Conformity and divergence in interactions, groups, and culture. In S. G. Harkins, K. D. Williams, & J. M. Burger (Eds.), *The Oxford handbook of social influence* (pp. 87–105). New York, NY: Oxford University Press.

Hoemann, K., Gendron, M., & Barrett, L. F. (2017). Mixed emotions in the predictive brain. *Current Opinion in Behavioral Sciences, 15, 51-57.*

Hoffmann, P., Platow, M. J., Read, E., Mansfield, T., Carron-Arthur, B., & Stanton, M. (2020). Perceived self-in-group prototypicality enhances the benefits of social identification for psychological well-being. *Group Dynamics: Theory, Research, and Practice, 24(4), 213–226.*

Hogg, M. A. (2014). From uncertainty to extremism: Social categorization and identity processes. *Current Directions in Psychological Science, 23(5), 338-342.*

Hogg, M. A. (2021). Uncertain self in a changing world: A foundation for radicalisation, populism, and autocratic leadership. *European Review of Social Psychology, 32(2), 235-268.*

Hogg, M. A., Abrams, D., & Brewer, M. B. (2017). Social identity: The role of self in group processes and intergroup relations. *Group Processes & Intergroup Relations, 20(5), 570-581.*

Hogg, M. A., & Adelman, J. (2013). Uncertainty–identity theory: Extreme groups, radical behavior, and authoritarian leadership. *Journal of Social Issues, 69(3), 436-454.*

Hohman, Z. P., Gaffney, A. M., & Hogg, M. A. (2017). Who am I if I am not like my group? Self-uncertainty and feeling peripheral in a group. *Journal of Experimental Social Psychology, 72, 125-132.*

Hornsey, M. J., Edwards, M., Lobera, J., Díaz‐Catalán, C., & Barlow, F. K. (2021). Resolving the small‐pockets problem helps clarify the role of education and political ideology in shaping vaccine scepticism. *British Journal of Psychology, 112(4), 992-1011.*

Inman, J. J., & Zeelenberg, M. (2002). Regret in repeat purchase versus switching decisions: The attenuating role of decision justifiability. *Journal of consumer research, 29(1), 116-128.*

John, S. (2018). Epistemic trust and the ethics of science communication: Against transparency, openness, sincerity and honesty. *Social Epistemology, 32(2), 75-87.*

Jolley, D., & Douglas, K. (2014a). The social consequences of conspiracism: Exposure to conspiracy theories decreases intentions to engage in politics and to reduce one's carbon footprints. *British Journal of Psychology, 105, 35–56*

Jost, J. T., Hennes, E. P., & Lavine, H. (2013). "Hot" political cognition: Its self-, group-, and system-serving purposes. *Oxford handbook of social cognition*, 851–875. Oxford, England: Oxford University Press.

Kaasa, A., & Andriani, L. (2022). Determinants of institutional trust: the role of cultural context. *Journal of Institutional Economics, 18(1), 45-65.*

Kahan, D. M. (2012). Ideology, motivated reasoning, and cognitive reflection: An experimental study. *Judgment and Decision making, 8, 407-24.*

Kahan, D. M. (2016a). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. *Emerging Trends in the Social and Behavioral Sciences, 29.*

Kahan, D. M. (2016b). The politically motivated reasoning paradigm, Part 2: Unanswered questions. *Emerging Trends in the Social and Behavioral Sciences, 1-15.*

Kahan, D. M. (2017). Misconceptions, misinformation, and the logic of identity-protective cognition. *Cultural cognition project working paper series*, 164. Yale, CT: Yale Law School.

Kahan, D. M., Braman, D., Gastil, J., Slovic, P., & Mertz, C. K. (2007). Culture and identity‐protective cognition: Explaining the white‐male effect in risk perception. *Journal of Empirical Legal Studies, 4(3), 465-505.*

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

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.*

Katz, J., Joiner, T. E., & Kwon, P. (2002). Membership in a devalued social group and emotional well-being: Developing a model of personal self-esteem, collective self-esteem, and group socialization. *Sex roles, 47(9), 419-431.*

Knippenberg, D., Lossie, N., & Wilke, H. (1994). In-group prototypicality and persuasion: Determinants of heuristic and systematic message processing. *British Journal of Social Psychology, 33(3), 289–300.*

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research, 47(1), 104-124.*

Kraft, P. W., Lodge, M., & Taber, C. S. (2015). Why people "don't trust the evidence" motivated reasoning and scientific beliefs. *The ANNALS of the American Academy of Political and Social Science, 658(1), 121-133.*

Kruglanski, A. W. (2013). *The psychology of closed mindedness*. New York, NY: Psychology Press.

Kuchling, F., Friston, K., Georgiev, G., & Levin, M. (2020). Morphogenesis as Bayesian inference: A variational approach to pattern formation and control in complex biological systems. *Physics of Life Reviews, 33, 88-108.*

Kukkonen, A., Ylä‐Anttila, T., & Broadbent, J. (2017). Advocacy coalitions, beliefs and climate change policy in the United States. *Public Administration, 95(3), 713-729.*

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

Kurzban, R. (2011). *Why everyone (else) is a hypocrite. Evolution and the modular mind*. Princeton, NJ: Princeton University Press.

Landau, M. J., Solomon, S., Pyszczynski, T., & Greenberg, J. (2007). On the compatibility of terror management theory and perspectives on human evolution. *Evolutionary Psychology, 5(3), 147470490700500303.*

Leary, M. R. (2021). The need to belong, the sociometer, and the pursuit of relational value: Unfinished business. *Self and Identity, 20(1), 126-143.*

Leary, M. R., & Acosta, J. (2018). *Acceptance, rejection, and the quest for relational value.* In *Cambridge handbook of personal relationships* (pp. 78–390). Cambridge, England: Cambridge University Press.

Leary, M. R., Kelly, K. M., Cottrell, C. A., & Schreindorfer, L. S. (2013). Construct validity of the need to belong scale: Mapping the nomological network. *Journal of Personality Assessment, 95(6), 610-624.*

Leary, M. R., Kowalski, R. M., Smith, L., & Phillips, S. (2003). Teasing, rejection, and violence: Case studies of the school shootings. *Aggressive Behavior: Official Journal of the International Society for Research on Aggression, 29(3), 202-214.*

Leary, M. R., Patton, K. M., Orlando, A. E., & Wagoner Funk, W. (2000). The impostor phenomenon: Self‐perceptions, reflected appraisals, and interpersonal strategies. *Journal of Personality, 68(4), 725-756.*

Leotti, L. A., Iyengar, S. S., & Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control. *Trends in cognitive sciences, 14(10), 457-463.*

Lieberman, M. D., & Eisenberger, N. I. (2006). A pain by any other name (rejection, exclusion, ostracism) still hurts the same: The role of dorsal anterior cingulate cortex in social and physical pain. *Social Neuroscience: People thinking about thinking people, 167-187.*

Luchies, L. B., Finkel, E. J., McNulty, J. K., & Kumashiro, M. (2010). The doormat effect: When forgiving erodes self-respect and self-concept clarity. *Journal of Personality and Social Psychology, 98(5), 734–749.*

Maalouf, A. (2011). *On identity*. London, England: Penguin Random House.

Martín-Vega, D., Garbout, A., Ahmed, F., Wicklein, M., Goater, C. P., Colwell, D. D., & Hall, M. J. (2018). 3D virtual histology at the host/parasite interface: visualisation of the master manipulator, Dicrocoelium dendriticum, in the brain of its ant host. *Scientific Reports, 8(1), 1-10.*

Mason, L., & Wronski, J. (2018). One tribe to bind them all: How our social group attachments strengthen partisanship. *Political Psychology, 39, 257-277.*

McCulloh, I. (2013). Social conformity in networks. *Connections*, *33(1), 35-42.*

McDougall, P., Hymel, S., Vaillancourt, T., & Mercer, L. (2001). The consequences of childhood peer rejection. *Interpersonal Rejection, 213-247.*

McGregor, I. (2006). Offensive defensiveness: Toward an integrative neuroscience of compensatory zeal after mortality salience, personal uncertainty, and other poignant self-threats. *Psychological Inquiry, 17(4), 299-308.*

Milanov, M., Rubin, M., & Paolini, S. (2014). Different types of ingroup identification: A comprehensive review, an integrative model, and implications for future research. *Psicologia Sociale, 9(3), 205-232.*

Miller, W. I. (1995). *Humiliation: And other essays on honor, social discomfort, and violence*. Ithaca, NY: Cornell University Press.

Mikami, A. Y., Schad, M. M., Teachman, B. A., Chango, J. M., & Allen, J. P. (2015). Implicit versus explicit rejection self-perceptions and adolescents' interpersonal functioning. *Personality and Individual Differences, 86, 390–393*

Miton, H., & Mercier, H. (2015). Cognitive obstacles to pro-vaccination beliefs. Trends in Cognitive Sciences, 19(11), 633-636.

Mölder, H. (2011). The culture of fear in international politics: A Western-dominated international system and its extremist challenges. *KVÜÕA Toimetised, (14), 241-263.*

Nasr, M. (2021). The motivated electorate: Voter uncertainty, motivated reasoning, and ideological congruence to parties. *Electoral Studies, 72, 102344.*

Neighbors, C., Rodriguez, L. M., Rinker, D. V., Gonzales, R. G., Agana, M., Tackett, J. L., & Foster, D. W. (2015). Efficacy of personalized normative feedback as a brief intervention for college student gambling: a randomized controlled trial. *Journal of Consulting and Clinical Psychology, 83(3), 500.*

Nisbett, R. E., & Cohen, D. (2018). *Culture of honor: The psychology of violence in the South.* Oxfordshire, England: Routledge.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist, 41(3), 181-206.*

Osborne, D., & Sibley, C. G. (2020). Does openness to experience predict changes in conservatism? A nine-wave longitudinal investigation into the personality roots to ideology. *Journal of Research in Personality, 87, 103979.*

Owens, T. J., & Samblanet, S. (2013). Self and self-concept. In *Handbook of social psychology* (pp. 225-249). Dordrecht, Netherlands: Springer.

Owuamalam, C. K., Rubin, M., & Issmer, C. (2016). Reactions to group devaluation and social inequality: A comparison of social identity and system justification predictions. *Cogent Psychology, 3(1), 1188442.*

Oyserman, D., Elmore, K., & Smith, G. (2012). Self, self-concept, and identity. In M. R. Leary & J. P. Tangney (Eds.), *Handbook of self and identity* (pp. 69–104). New York, NY: The Guilford Press.

Park, J., & Hill, W. T. (2018). Exploring the role of justification and cognitive effort exertion on post-purchase regret in online shopping. *Computers in Human Behavior, 83, 235-242.*

Pearlman, W. (2016). Narratives of fear in Syria. *Perspectives on Politics, 14(1), 21-37.*

Pérez-Escudero, A., Friedman, J., & Gore, J. (2016). Preferential interactions promote blind cooperation and informed defection. *Proceedings of the National Academy of Sciences, 113(49), 13995–14000.*

Pezzulo, G., Rigoli, F., & Friston, K. J. (2018). Hierarchical active inference: a theory of motivated control. *Trends in Cognitive Sciences, 22(4), 294-306.*

Pfundmair, M. (2019). Ostracism promotes a terroristic mindset. *Behavioral Sciences of Terrorism and Political Aggression, 11(2), 134-148.*

Phelps, E. A. (2009). The study of emotion in neuroeconomics. In *Neuroeconomics* (pp. 233-250). Cambridge, MA: Academic Press.

Pilarska, A. (2016). How do self-concept differentiation and self-concept clarity interrelate in predicting sense of personal identity?. *Personality and Individual Differences, 102, 85-89.*

Yıldırım, M., Geçer, E., & Akgül, Ö. (2021). The impacts of vulnerability, perceived risk, and fear on preventive behaviours against COVID-19. *Psychology, Health & Medicine, 26(1), 35-43.*

Poon, K. T., Chen, Z., & Wong, W. Y. (2020). Beliefs in conspiracy theories following ostracism. *Personality and Social Psychology Bulletin, 46(8), 1234-1246.*

Postmes, T., Gordijn, E. H., & van Zomeren, M. (2014). Escalation and de-escalation of intergroup confl ict: The role of communication within and between groups. In *Social conflict within and between groups* (pp. 107-130). London, England: Psychology Press.

Prike, T., Arnold, M. M., & Williamson, P. (2018). The relationship between anomalistic belief and biases of evidence integration and jumping to conclusions. *Acta Psychologica, 190, 217-227.*

Rabeyron, T., & Finkel, A. (2020). Consciousness, free energy and cognitive algorithms. *Frontiers in Psychology, 11, 1675.*

Rico, D., & Barreto, I. (2022). Unfreezing of the conflict due to the peace agreement with FARC–EP in Colombia: Signature (2016) and implementation (2018). *Peace and Conflict: Journal of Peace Psychology, 28(1), 22.*

Rokeach, M. (1968). *Beliefs, attitudes, and values*. San Francisco, CA: Jossey-Bass.

Safron, A. (2021). The radically embodied conscious cybernetic Bayesian brain: from free energy to free will and back again. *Entropy, 23(6), 783.*

Schwartenbeck, P., FitzGerald, T., Dolan, R., & Friston, K. (2013). Exploration, novelty, surprise, and free energy minimization. *Frontiers in psychology, 710.*

Sesack, S. R., & Grace, A. A. (2010). Cortico-basal ganglia reward network: microcircuitry. *Neuropsychopharmacology, 35(1), 27-47.*

Seuken, S., & Zilberstein, S. (2008). Formal models and algorithms for decentralized decision making under uncertainty. *Autonomous Agents and Multi-Agent Systems, 17(2), 190-250.*

Shah, J. Y., Kruglanski, A. W., & Thompson, E. P. (1998). Membership has its (epistemic) rewards: need for closure effects on in-group bias. *Journal of Personality and Social Psychology, 75(2), 383.*

Sherman, D. K. (2013). Self‐affirmation: Understanding the effects. *Social and Personality Psychology Compass, 7(11), 834-845.*

Sherman, D. K., & Cohen, G. L. (2006). The psychology of self‐defense: Self‐affirmation theory. *Advances in Experimental Social Psychology, 38, 183-242.*

Sibly, R. M., & Brown, J. H. (2009). Mammal reproductive strategies driven by offspring mortality-size relationships. *The American Naturalist, 173(6), E185-E199.*

Sim, J. J., Goyle, A., McKedy, W., Eidelman, S., & Correll, J. (2014). How social identity shapes the working self-concept. *Journal of Experimental Social Psychology, 55, 271–277.*

Simler, K., & Hanson, R. (2017). *The elephant in the brain: Hidden motives in everyday life.* Oxford, England: Oxford University Press.

Sindic, D., & Condor, S. (2014). Social identity theory and self-categorisation theory. In *The Palgrave handbook of global political psychology* (pp. 39-54). London, England: Palgrave Macmillan.

Solomon, S., Greenberg, J., & Pyszczynski, T. (2004). The cultural animal: twenty years of terror management theory and research. In J. Greenberg, S. L. Koole, & T. Pyszczynski (Eds.), *Handbook of experimental existential psychology* (pp. 13–34). New York, NY: Guilford Press.

Ståhl, T., & van Prooijen, J.-W. (2018). Epistemic rationality: Skepticism toward unfounded beliefs requires sufficient cognitive ability and motivation to be rational. *Personality and Individual Differences, 122, 155–163.*

Stanley, M. L., Henne, P., Yang, B. W., & De Brigard, F. (2020). Resistance to position change, motivated reasoning, and polarization. *Political Behavior, 42(3), 891-913.*

Stanovich, K. E. (2021). *The bias that divides us: The science and politics of myside thinking*. Cambridge, MA: MIT Press.

Stanovich, K. E., West, R. F., & Toplak, M. E. (2013). Myside bias, rational thinking, and intelligence. Current Directions in Psychological Science, 22(4), 259-264.

Stern, C. (2021). *The impact of relational goals on political polarization. In The Psychology of Political Polarization* (pp. 77-93). London, England: Routledge.

Stowers, D. A., & Durm, M. W. (1996). Does self-concept depend on body image? A gender analysis. *Psychological Reports, 78(2), 643-646.*

Sznycer, D., Tooby, J., Cosmides, L., Porat, R., Shalvi, S., & Halperin, E. (2016). Shame closely tracks the threat of devaluation by others, even across cultures. *Proceedings of the National Academy of Sciences, 113(10), 2625–2630.*

Tajfel, H. (1981). *Human groups and social categories: Studies in social psychology*. Cambridge, England: Cambridge University Press

Tang, S. (2008). Fear in international politics: two positions. *International Studies Review, 10(3), 451-471.*

Tappin, B. M., Pennycook, G., & Rand, D. G. (2021). Rethinking the link between cognitive sophistication and politically motivated reasoning. *Journal of Experimental Psychology: General, 150(6), 1095–1114.*

Toelch, U., & Dolan, R. J. (2015). Informational and normative influences in conformity from a neurocomputational perspective. *Trends in Cognitive Sciences, 19(10), 579-589.*

Van der Linden, S., Leiserowitz, A., & Maibach, E. (2019). The gateway belief model: A large-scale replication. *Journal of Environmental Psychology, 62, 49-58.*

Van der Wal, R. C., Sutton, R. M., Lange, J., & Braga, J. P. (2018). Suspicious binds: Conspiracy thinking and tenuous perceptions of causal connections between co‐occurring and spuriously correlated events. *European Journal of Social Psychology, 48(7), 970-989.*

Van Prooijen, J. W., Douglas, K. M., & De Inocencio, C. (2018). Connecting the dots: Illusory pattern perception predicts belief in conspiracies and the supernatural. *European Journal of Social Psychology, 48(3), 320-335.*

Van Prooijen, J.-W., Klein, O., & Milošević Đorđević, J. (2020). Social-cognitive processes underlying belief in conspiracy theories. In M. Butter & P. Knight (Eds.), *Handbook of conspiracy theories* (pp. 168-180). London, England: Routledge.

Van Prooijen, J.-W., & Van Dijk, E. (2014). When consequence size predicts belief in conspiracy theories: The moderating role of perspective taking. *Journal of Experimental Social Psychology, 55, 63–73.*

Wagoner, J. A., Belavadi, S., & Jung, J. (2017). Social identity uncertainty: Conceptualization, measurement, and construct validity. *Self and Identity, 16, 505–530.*

Wagoner, J. A., & Hogg, M. A. (2016). Normative dissensus, identity-uncertainty, and subgroup autonomy. Group Dynamics: *Theory, Research, and Practice, 20, 310–322.*

Warburton, W. A., Williams, K. D., & Cairns, D. R. (2006). When ostracism leads to aggression: The moderating effects of control deprivation. *Journal of Experimental Social Psychology, 42(2), 213-220.*

Whitson, J. A., & Galinsky, A. D. (2008). Lacking control increases illusory pattern perception. *Science, 322(5898), 115-117.*

Wiemer, J., Leimeister, F., & Pauli, P. (2021). Subsequent memory effects on event-related potentials in associative fear learning. *Social Cognitive and Affective Neuroscience, 16(5), 525–536.*

Williams, K. D. (2007). Ostracism. *Annual Review of Psychology, 58, 425-452.*

Williams, K. D., & Zadro, L. (2005). Ostracism: The indiscriminate early detection system. In K. D. Williams, J. P. Forgas, & W. von Hippel (Eds.), The social outcast: Ostracism, social exclusion, rejection, and bullying (pp. 19–34). Psychology Press.

Williams, D. (2018). Predictive processing and the representation wars. *Minds and Machines, 28(1), 141-172.*

Williams, D. (2020). Epistemic Irrationality in the Bayesian Brain. *The British Journal for the Philosophy of Science, 72(4), 913-938.*

Williams, D. (2021). Socially adaptive belief. *Mind & Language, 36(3), 333-354.*

Wimmer, G. E., & Shohamy, D. (2012). Preference by association: how memory mechanisms in the hippocampus bias decisions. *Science, 338(6104), 270-273.*

Wirth, J. H., & Wesselmann, E. D. (2018). Investigating how ostracizing others affects one's self-concept. *Self and Identity, 17(4), 394–406.*

Woodward, K. (2003). *Understanding identity*. London, England: Hodder Arnold.

Zhao, X., & Epley, N. (2021). Kind words do not become tired words: Undervaluing the positive impact of frequent compliments. *Self and Identity, 20(1), 25-46.*

Zummo, L., Donovan, B., & Busch, K. C. (2021). Complex influences of mechanistic knowledge, worldview, and quantitative reasoning on climate change discourse: Evidence for ideologically motivated reasoning among youth. *Journal of Research in Science Teaching, 58(1), 95-127.*