The essence of "pathological polarization" in postmodern society

La esencia de la "polarización patológica" en la sociedad postmoderna

A essência da "polarização patológica" na sociedade pós-moderna

Ana María Ruiz-Ruano García

Facultad de Ciencias de la Educación, Departamento de Psicología Evolutiva y de la Educación, Universidad de Granada,

España

https://orcid.org/0000-0002-7260-0588

amruano@ugr.es

Jorge López Puga

Facultad de Psicología, Departamento de Personalidad, Evaluación y Tratamiento Psicológico, Universidad de Granada,

España

https://orcid.org/0000-0003-0693-0092

jlpuga@ugr.es

Abstract

Polarization is one of the most pressing challenges facing democratic societies today. While not inherently harmful, polarization becomes detrimental when it crosses certain thresholds—disrupting social cooperation, escalating conflict, and ultimately leading to the dehumanization of public discourse. This paper explores the concept of *pathological polarization*, understood as a toxic form of societal clustering that undermines social harmony. Two speculative yet potentially fruitful ideas are introduced to inspire future research. First, the gregariousness/individualism tension (GIT) is proposed as a psychological framework for understanding the emergence of polarized identities. Second, overconfidence in inductive reasoning is examined as a cognitive source of bias that may reinforce polarized thinking. Finally, the paper outlines preliminary proposals for mitigating pathological polarization, with the aim of fostering epistemically successful and socially harmonious communities.
Keywords: Social conflict, politics, education, well-being, democracy, reasoning, attitudes, beliefs.

Resumen

La polarización es uno de los desafíos más urgentes que enfrentan las sociedades democráticas en la actualidad. Aunque no es intrínsecamente perjudicial, la polarización se vuelve dañina cuando supera ciertos umbrales, interrumpiendo la cooperación social, intensificando los conflictos y conduciendo a la deshumanización del discurso público. Este artículo explora el concepto de polarización patológica, entendida como una forma tóxica de agrupamiento social que socava la armonía colectiva. Se presentan dos ideas especulativas pero potencialmente fructíferas para inspirar futuras investigaciones.

En primer lugar, se propone la tensión gregarismo/individualismo (GIT) como un marco psicológico para comprender la emergencia de identidades polarizadas. En segundo lugar, se examina la sobreconfianza en el razonamiento inductivo como fuente cognitiva de sesgos que pueden reforzar el pensamiento polarizado. Finalmente, se esbozan propuestas preliminares para mitigar la polarización patológica, con el objetivo de fomentar comunidades epistemológicamente exitosas y socialmente armoniosas.

Palabras clave: conflicto social, política, educación, bienestar, democracia, razonamiento, actitudes, creencias.

Resumo

A polarização é um dos desafios mais prementes enfrentados pelas sociedades democráticas atualmente. Embora não seja necessariamente prejudicial, torna-se nociva quando ultrapassa certos limites — interrompendo a cooperação social, intensificando conflitos e levando à desumanização do discurso público. Este artigo explora o conceito de *polarização patológica*, entendida como uma forma tóxica de agrupamento social que compromete a harmonia coletiva. Duas ideias especulativas, mas potencialmente promissoras, são apresentadas para inspirar pesquisas futuras. Primeiro, propõe-se a tensão entre gregário e individualismo (GIT) como um modelo psicológico para compreender o surgimento de identidades polarizadas. Em seguida, examina-se o excesso de confiança no raciocínio indutivo como uma fonte cognitiva de viés que pode reforçar o pensamento polarizado. Por fim, são delineadas propostas preliminares para mitigar a polarização patológica, com o objetivo de promover comunidades epistemicamente bem-sucedidas e socialmente harmoniosas.

Palavras-chave: conflito social, política, educação, bem-estar, democracia, raciocínio, atitudes, crenças.

Introduction

Speaking about polarization is inherently difficult, as it challenges the possibility of adopting a truly neutral stance. Metaphorically, neutrality resembles a flat encephalogram —an absence of life. Positioning oneself equidistant from opposing extremes may shield one from conflict, but it also isolates, attracting neither allies nor adversaries. In this sense, equidistance can appear suspicious or even antagonistic. As Greene (2004) suggests, complete detachment is nearly impossible; we are inevitably drawn into one side or another. Yet, striving for a balanced perspective remains essential if we are to resolve social conflicts sensibly. When society fails to reach calm consensus on critical issues, it risks squandering time and energy —and, more gravely, undermining its own capacity to function.

Polarization has been described as "one of society's greatest ills" (Jung et al., 2019, p. 301). Before

proceeding, a brief clarification is needed regarding how the term *society* will be used throughout this essay. This is not a trivial matter, given the conceptual scope of the arguments that follow. Here, society refers not merely to a specific nation, culture, or even the so-called "global society," but to humanity as a whole. To capture this broader, more inclusive perspective, we propose the neologism humansphere —a conceptual parallel to the biosphere, which denotes the totality of life on Earth. While the biosphere encompasses all living organisms, the humansphere extends beyond biological existence to include human-created artifacts, institutions, and symbolic systems. A book, for instance, though not alive, belongs to the humansphere by virtue of its human origin and cultural significance.

As Jung et al. (2019) emphasize, polarization may be understood as a disease, a disorder, or even a social infection. Kirmayer (2024) similarly frames it as a form of social pathology, marked by self-destructive ideologies that erode civil society and undermine collective well-being. Polarization weakens social cohesion and obstructs cooperation (Jost et al., 2022). It fractures trust within the humansphere, damaging interpersonal relationships, institutional processes, and the functioning of social subsystems (Patent, 2022). One of the most alarming consequences of polarization emerges during emergencies. Polarized societies often fail to respond effectively to crises -whether natural disasters or armed conflicts- thereby exacerbating human suffering. When polarization becomes extreme, the shared construction of reality begins to fragment. The boundary between facts and opinions blurs, enabling the proliferation of misinformation and fake news. In such contexts, competing efforts to impose official narratives often arise (Berger & Luckmann, 1967). This erosion of consensus on critical issues increases the likelihood of radical, even violent, group formation (Jung et al., 2019). More troubling still, polarization fosters a corrosive sense of dehumanization (Jost et al., 2022). It also takes a psychological toll, contributing to rising levels of anxiety, depression, and stress-related disorders within civil society (Makri, 2024).

Polarization is not a new phenomenon, but what is novel are the vehicles through which polarized and polarizing ideas now propagate. Today, much of this dynamic unfolds within the framework of social media interactions (Kirmayer, 2024). These platforms accelerate the spread of polarizing content and facilitate a high rate of ideological mutation, rendering the phenomenon increasingly complex and difficult to anticipate. Social media presents a twofold challenge. On one hand, it satisfies fundamental human needs for connection and communication (Ruiz-Ruano et al., 2023). On the other, it amplifies polarization through the networked architecture of global communication. Several structural properties of social networks help explain this dynamic: the scale-free nature of online platforms (Barabási, 2009; Barabási & Bonabeau, 2003), the principle of preferential attachment in evolving connections (Barabási & Albert, 1999), and the small-world topology of social ties (Watts & Strogatz, 1998). These features enable rapid diffusion and reinforcement of polarized narratives. Moreover, misinformation and disinformation play a central role in igniting and sustaining polarization within these networks (Ball & Maxmen, 2020; Kirmayer, 2024). As we will argue, the interplay between technological affordances and

cognitive vulnerabilities constitutes a fertile ground for the intensification of polarization in the humansphere.

In this article, we aim to describe *pathological polarization* from a psychosocial perspective. The concept should be understood in light of what Jung et al. (2019) refer to as *epistemologically healthy societies* (p. 310), and within a harmonizing and functional conception of mental health (Frances & Widiger, 2012). Pathological polarization refers to a societal condition marked by a profound failure in epistemological coordination —where society can no longer converge on a shared understanding of facts, whether historical or contemporary. In such contexts, the boundary between reality and fiction becomes socially, not just individually, indiscernible. This may even give rise to the illusion that something can be simultaneously true and false. By analogy, extreme polarization resembles a form of *social mental disorder*, as it mirrors symptoms observed in individuals with severe psychopathology. Importantly, we do not invoke "pathology" from a strictly realist or nominalist standpoint, nor do we adopt a constructivist-pessimist view of social organization. Rather, we conceptualize pathology here as the loss of systemic equilibrium. In this sense, pathological polarization is a dysfunctional process that disrupts the balance and harmony of the humansphere.

In what follows, we focus on two key features that contribute to pathological polarization: (a) the perverse imbalance between gregariousness and individualism, and (b) the overreliance on inductive reasoning as a primary mode of thought. We begin with the premise that polarization is not inherently detrimental. In fact, under certain conditions, it can foster social and political progress by challenging stagnant norms and stimulating debate. However, when polarization inflicts acute or chronic harm -whether through immediate destruction or delayed dysfunction- and when it obstructs collective efforts to address urgent societal issues, it becomes necessary to scrutinize its dynamics with critical attention.

The remainder of this text is organized into five sections. In the next section, we examine the psychology of polarization, presenting both classical and contemporary definitions. Section 3 explores the tension between gregariousness and individualism as a central dynamic in pathological polarization within postmodern societies. Section 4 addresses the cognitive dimension, focusing on how polarization is shaped by thinking biases —particularly the overreliance on inductive reasoning. Finally, in Section 5, we offer concluding reflections aimed at identifying conditions under which polarization might be transformed into a constructive force for individuals and for the humansphere as a whole.

Psychology of polarization

The term *polarization* carries multiple meanings within social psychology. The most widely preserved interpretation in textbooks refers to *group polarization* (see, for example, Hogg & Vaughan, 2018). This phenomenon occurs when a group discusses a socially contentious issue -such as war, migration during economic crises, or nuclear energy in the context of climate change- and the group's average attitude

shifts toward a more extreme position. Group polarization is typically observed through a three-phase experimental design. First, individual attitudes toward a given issue are assessed. Then, participants engage in group discussion. Finally, individual attitudes are reassessed. What is consistently found is that the post-discussion average becomes more extreme than the pre-discussion baseline. For instance, if the group's initial average support for nuclear energy is moderate (e.g., 50 out of 100), it may rise significantly (e.g., to 89) after deliberation. This form of polarization is critical for understanding decision-making in complex social contexts -such as jury deliberations- where group dynamics can amplify preexisting biases and lead to skewed outcomes. In such cases, the final verdict may reflect not a balanced synthesis of evidence, but a polarized echo of initial individual leanings.

Although *group polarization* remains a foundational concept in social psychology, other forms -such as attitude polarization (Sherif, 1958) and political polarization (Converse, 1958)- were also explored in the mid-20th century. As Converse (1958) noted, certain expressions of political polarization are shaped by social class, and the rhetoric surrounding class divisions continues to fuel polarization today. Given the conceptual diversity of the term, the following subsections aim to clarify key dimensions of contemporary polarization. We adopt the classification proposed by Jost et al. (2022), distinguishing among *ideological*, *partisan*, and *affective polarization*. Subsequently, in Sections 3 and 4, we introduce two psychosocial mechanisms that, in our view, partially explain the emergence of pathological polarization: the tension between gregariousness and individualism, and the paradoxical role of inductive reasoning. Finally, in Section 5, we offer conclusions and propose strategies to mitigate pathological polarization and promote healthier forms of social dissent.

Ideological Polarization

Issue or *ideological polarization* refers to a form of societal division in which individuals or groups adopt opposing, often extreme, positions on issues critical to social organization. This type of polarization is frequently modeled statistically as a bimodal distribution, where two distinct peaks represent opposing ideological camps. The distribution may be symmetric, with both sides moving away from the center to a similar degree, or asymmetric, where one group remains near the center while the other shifts toward an extreme.

The psychology of issue or ideological polarization can be meaningfully understood through the lens of attitudes. In this view, polarization is not merely a structural or political phenomenon, but an attitudinal process. The concept of attitude conflict has been introduced to describe situations marked by "competitive disagreement with regard to beliefs, values, and preferences, characterized by parties' intolerance of each other's positions" (Minson & Dorison, 2022, p. 182). Attitudes, as foundational constructs in social psychology, play a central role in shaping how individuals and groups engage with polarizing issues. According to Hogg and Vaughan (2018, p. 154), an attitude is both:

a) A relatively enduring organisation of beliefs, feelings and behavioural tendencies towards socially significant objects, groups, events or symbols.

b) A general feeling or evaluation -positive or negative- about some person, object or issue.

As noted above -and as Allport (1935) suggested nearly a century ago- attitudes are multidimensional and relatively stable psychological constructs. This stability implies that attitudes tend to persist over time unless disrupted by significant external influences. The definitions cited also emphasize that attitudes are evaluative in nature: they reflect favorable or unfavorable orientations toward specific aspects of reality. A final critical feature of attitudes is their dimensionality. Attitudes are not unidimensional judgments but rather complex evaluations that span multiple psychological domains. The prevailing view in contemporary social psychology conceptualizes attitudes as comprising three interrelated components: cognitive, emotional, and behavioral (Ajzen & Fishbein, 2005; Feldman, 1998; Franzoi, 2005). This tripartite model provides a useful framework for understanding how ideological polarization becomes deeply embedded in both individual and collective psychology.

Attitude formation is a complex process shaped by multiple psychosocial mechanisms (Hogg & Vaughan, 2018). Attitudes are learned and gradually integrated into cognitive, emotional, and behavioral response patterns throughout development (Fishbein & Ajzen, 1975). For instance, early traumatic experiences -as suggested by psychoanalytic theory- can strongly influence the fixation of certain attitudes, sometimes persisting across the lifespan (Allport, 1935). However, trauma is not the only pathway. Classical conditioning, as demonstrated by Pavlov (1927), also plays a role: attitudes may form through repeated associations between neutral stimuli and emotionally charged or biologically significant events. In addition, vicarious learning -observing and internalizing the attitudes of others- further contributes to attitudinal development. This is why Allport (1935) emphasized the importance of parents and significant peers in shaping attitudes. Ultimately, our orientations toward socially significant issues are often modeled on those we perceive as relevant or authoritative.

A critical element in understanding contemporary ideological polarization is the role of *beliefs*. Beliefs exert a powerful influence on both attitudes and behavioral intentions (Ajzen & Fishbein, 1980). What makes ideological polarization particularly intense today is its connection to fundamental or existential concerns (Lenci, 2023). People do not polarize over trivial preferences -such as whether apples taste better than oranges- but over deeply consequential questions about how society should be organized. Polarization intensifies when debates focus on issues such as the pursuit of peace during wartime, the meaning and legitimacy of private property, or matters of religious faith. These are not merely abstract topics; for many, they are existential anchors —sources of meaning that shape their understanding of life itself. As a result, when opposing views challenge these core beliefs, the potential for polarization becomes not only likely but almost inevitable.

Partisan Polarization

The term partisan refers to strong allegiance to a person, principle, or political party, often without critical evaluation (Walter et al., 2008). Partisan polarization, then, describes a form of political extremism in which individuals increasingly align their thinking with one end of the political spectrum, often disregarding nuance or compromise. This phenomenon can be interpreted as a contemporary expression of long-standing social conflicts. For instance, the class-based confrontations during the French Revolution may be viewed through the lens of early political polarization. In this context, we might ask: were those uprisings fueled by a form of social polarization? Converse (1958) appears to suggest so when introducing the concept of class polarization.

While political discourse has undoubtedly become more complex over time, it still carries a familiar resonance —an "aged taste," reminiscent of historical struggles, like that of matured cheese or wine. Across Europe, the United States, and many other regions, political landscapes have long been structured around a binary spectrum: left and right. In the United States, these positions are often associated with the Democratic and Republican parties, respectively. However, as Ruiz-Ruano and Puga (2022) point out, this dichotomy is not always descriptively accurate. The labels Democrat and Republican often obscure the nuances of political orientation. Similarly, the liberal-conservative distinction can be misleading. In some contexts, "liberalism" is paradoxically associated with conservative ideologies. Nevertheless, this dichotomy does highlight a deeper ideological divide, one that Jost et al. (2022) link to the concept of system justification. Conservative ideologies -typically aligned with the political right- tend to resist systemic change, favoring the preservation of the status quo. In contrast, liberal or progressive ideologies advocate for structural transformation in pursuit of social improvement. What underlies this divide is the enduring political counterpoint that history has repeatedly documented.

Partisan alignment is arguably more complex than ideological polarization, as it is not confined to a single attitudinal dimension but spans a multidimensional space of political, cultural, and moral issues. As Makri (2024) suggests, partisan animosity can erode public health by intensifying and entrenching stress responses. One of the most dangerous consequences of extreme political polarization is the reduction of political discourse to a binary "us versus them" framework. When political dynamics are framed in such oversimplified terms, society risks descending into what might be called an *unhealthy political war*. In these conditions, in-group and out-group dynamics become distorted. Proportionality is lost, and moral judgments are no longer applied consistently: the same behavior may be condemned or excused depending on whether it is performed by an ally or an opponent. Individual identity is subsumed under group affiliation, and people are no longer treated as individuals but as representatives of a collective. This deindividuation paves the way for prejudice and the most dangerous forms of discrimination.

Affective Polarization

Affective polarization refers to the emergence of strong positive or negative emotions toward social or political groups, independent of ideological content or argumentation. In this form of polarization, what matters is not the substance of beliefs, but rather the emotional distinction between *ingroup* and *outgroup* membership. Recent studies suggest that affective polarization has intensified over the past decades in countries such as the United States, Switzerland, Denmark, Canada, and New Zealand, while it appears to have declined in others, including Japan, Australia, the United Kingdom, Norway, Sweden, and Germany (Boxell et al., 2024). Extreme manifestations of affective polarization can escalate into partisan violence, as exemplified by the public riots at the United States Capitol (May Sidik, 2023). These events underscore the potential for emotional group-based divisions to destabilize democratic institutions and civil order.

The forms of polarization discussed above -ideological, partisan, and affective- are not isolated phenomena. Rather, they interact and reinforce one another in feedback loops that can severely disrupt social functioning. What is at stake is not only the utilitarian efficiency of society, but also its deeper sense of social harmony. As Jost et al. (2022) observe, emotion influences ideology, ideology shapes emotion, and both are entangled with political party preferences. These recursive dynamics can have devastating consequences: they foster stereotypes, prejudice, in-group favouritism, out-group derogation, and ultimately, dehumanization. In extreme cases, such processes may escalate into open physical conflict. To understand how such destructive dynamics take root, we now turn to a deeper psychosocial tension —one that lies at the heart of postmodern society: the fragile balance between gregariousness and individualism.

Gregariousness/Individualism Tension

Describing, predicting, and preventing unhealthy polarization remains a formidable challenge, largely due to its inherent complexity (Ladyman et al., 2013). Polarization is not a static phenomenon —it evolves over time, shaped by countless individual opinions and collective interactions. Like a wave crashing against the shore, it can carry immense energy and cause significant damage when amplified by social forces. Given this complexity, offering conceptual tools to understand polarization is no simple task. In this section, we explore the possibility that polarization is, at least in part, conditioned by a fundamental psychosocial tension: the dynamic interplay between gregariousness and individualism, as theorized by Gil (2021, 2022, 2024).

Human beings cannot survive without social support. This is not merely a statistical observation —it is a fundamental truth about our species. Unlike certain animals, such as some turtles or snakes, which

probabilistically survive without parental care, humans require sustained social interaction from birth. For example, some turtle species return to the shore each season to lay eggs, and once hatched, the baby turtles must reach the sea alone, evading predators. Only a small fraction survive to adulthood. Their survival is a matter of probability. In contrast, human survival is contingent on social care. There are no documented cases of human infants surviving without support beyond the first few days of life. The tragic case of *feral children*, such as Genie, illustrates the profound consequences of social neglect —severe impairments in functioning and integration into society. These examples underscore a central point: the human individual is not autonomous at birth. We are, from the outset, embedded in a social matrix. In this sense, the individual is inseparable from society.

The opposite extreme of the society–individual continuum is equally delicate. A society that fails to recognize individual or natural rights is now widely regarded as a questionable, if not unhealthy, form of social organization. Aldous Huxley (1946, p. 46) famously imagined a dystopian world in which "[…] everyone belongs to everyone else" —a mechanized society where individuals exist solely to serve collective functions, and personal freedom is systematically erased. In such a system, individuals are expected to conform entirely to social directives, leaving no room for dissent or deviation. Ideas deemed disruptive to social order are condemned to extinction. This vision illustrates the dangers of suppressing individuality in favour of rigid collectivism, and it highlights the fragile balance societies must maintain between cohesion and autonomy.

In reality, no society fully embodies the extreme models described above. Most social systems exist somewhere between these poles, which is why societies are often characterized as more individualistic or collectivistic. These orientations are frequently linked to economic models -such as capitalism and communism- but our concern here is not with the sociological or economic nature of society per se. While these domains are undoubtedly interconnected, our focus is on the psychological dimension of the self. The tension between gregariousness and individualism offers a compelling lens through which to examine polarization. This perspective aligns coherently with the dynamics discussed earlier and may help explain why this psychosocial dichotomy contributes to the emergence of polarization in postmodern societies.

The gregariousness/individualism tension (GIT) refers to the underlying forces shaping the psychological architecture of human beings. As a theoretical lens, GIT seeks to explain the historical emergence and variation of the psychological self (Gil, 2021). Within this framework, we distinguish two archetypal configurations —what we term *selfitypes*: the gregarious self and the individualist self. These selfitypes, though historically rooted, coexist to varying degrees in contemporary society (Gil, 2022). The *gregarious self* embodies an in-group integrated psychological orientation, reminiscent of prehistoric hominid herds of hunters. In such social structures, the group supersedes the individual; personal identity is subsumed under collective cohesion. The individual, in this context, is not a

standalone entity but a behavioral node contributing to group survival. A proto-individualist self emerged alongside work specialization (Gil, 2024), marking a qualitative shift in self-perception. Individuals began to define themselves through their functional roles within society —what they did became who they were. This transformation reframed the psychological self as a socially embedded function, no longer merely a member of a herd but a productive unit with differentiated value. Although the theory is still evolving, and detailed accounts of the individualist self remain forthcoming, we can anticipate its contours. GIT posits that the *individualist self* is a relatively recent development in the psychological evolution of our species (Gil, 2021). As we continue to refine this framework, we recognize that the tension between gregariousness and individualism is not merely historical —it is a living dialectic shaping our polarized present.

While Social Identity Theory and Self-Categorization Theory have been widely employed to explain polarization and its underlying processes -given their focus on intergroup relations and self-categorization dynamics (Hogg & Vaughan, 2018)- it is worth considering what explanatory potential the GIT framework might offer. In what follows, we engage in a speculative exercise, drawing on the published foundations of GIT theory, to explore how this perspective could illuminate the mechanisms of pathological polarization. Our aim is not only to deepen our understanding of polarization and its consequences but also to consider possible avenues for mitigating conflictual dynamics in contemporary society.

We may speculate that, from the perspective of GIT theory, polarization arises from the persistent tension between the gregarious and individualist archetypes that coexist -albeit in varying proportions-within contemporary individuals and groups. GIT would thus frame polarization as a natural phenomenon, serving adaptive functions for individuals, groups, and societies. This view aligns with recent proposals by Sharot et al. (2023), who argue for the utility of beliefs: attitudes and opinions possess value because they help individuals navigate survival challenges in adaptive ways. In this sense, expressing and defending particular attitudes or beliefs carries tangible consequences -both positive and negative- in modern societies. As a result, belief change is subject to selection pressures, with certain attitudes persisting or fading according to their adaptive utility.

Several questions arise when we examine polarization through the lens of GIT theory. To begin, we might ask: (1) In what ways is polarization a consequence of the contemporary archetypes proposed by GIT theory? (2) Do different self-archetypes exert selective influences on polarization processes? While a comprehensive answer to these questions lies beyond the scope of this manuscript -especially given the current theoretical limitations- we believe that offering tentative responses may provide valuable directions for future research.

When considering the relationship between GIT theory and polarization, we must remember that GIT is

fundamentally concerned with the historical evolution of the psychological self. The self, as conceptualized by GIT, is shaped by natural influences operating across evolutionary history. In this light, current trends in polarization can be interpreted as outcomes of ongoing evolutionary processes. Thus, both the emergence of the self and the phenomenon of polarization are, at minimum, historically intertwined with the present moment. Building on this perspective, we can propose hypotheses regarding the relationship between polarization and the archetypes of self. Moreover, it becomes possible to formulate propositions that explain how different self-archetypes may differentially impact the process of societal polarization. While these propositions remain speculative, they offer a conceptual foundation for future empirical research aimed at elucidating the mechanisms by which self-archetypes influence polarization dynamics.

Returning to the concept of "pathological polarization" introduced at the outset, we can now hypothesize -within the GIT framework- how polarization might be generated under the influence of a "subversive" individual self. By default, GIT theory posits that the individual self stands in tension with society and social order; thus, polarization processes that threaten or undermine social cohesion are likely to be initiated or amplified by one or more individual selves. In this view, the onset of polarization may be catalyzed by certain "antisystem" individuals, who subsequently persuade others to confront opposing groups in defense of their own interests. In this sense, collective polarization can be traced back to individual motivations rooted in the core of the individual self. As noted by Kirmayer (2024), polarization seems to emerge to activate or potentiate personal political agendas. However, as polarized groups become more extreme, a phenomenon akin to "toxic membership" may emerge: the group itself develops a quasi-personality, behaving as a unified entity that employs self-preservation strategies to endure over time. We use the term "toxic membership" to describe situations in which individuals relinquish their personal identities in order to fully align with the interests of a group, even when the group's behavior is ethically harmful —both to its own members, to those outside the group, and to society at large. In such cases, group belonging no longer serves the individual but instead drives members toward confrontation with dissenting groups. This toxicity can be seen as the starting point for the most extreme forms of group conflict, potentially escalating to violence and threatening the integrity of human life. This dynamic poses significant risks to societal stability, mirroring the way the individualist archetype may threaten social order. While this account is necessarily speculative and incomplete, we believe it offers a potentially fruitful perspective for conceptualizing and investigating polarization processes in future research (May Sidik, 2023).

Having outlined the speculative mechanisms by which GIT theory may account for pathological polarization, we now turn to a related but distinct issue: the paradoxes inherent in inductive reasoning, and how they may further enrich our understanding of polarization in contemporary society.

The Inductive Reasoning Paradox

Building on our previous, largely speculative account grounded in GIT theory, it becomes clear that psychosocial and humanities-based explanations offer valuable but partial insights into polarization. As Jung et al. (2019) argue, addressing the complex phenomenon of pathological polarization requires an integrated approach that combines scientific and humanistic perspectives. In this context, the paradoxes inherent in inductive reasoning emerge as a crucial, yet often overlooked, dimension that may further complicate -and potentially enrich- our understanding of polarization in contemporary society.

One possible factor influencing contemporary social polarization is a form of "overconfidence" in inductive reasoning. Inductive reasoning -unlike deductive reasoning- enables us to construct theories, ideas, or models based on specific cases or observations (Wason, 1966). This cognitive process is so fundamental to human thought that Jeffreys (1931) argued deductive reasoning itself is merely a consequence of knowledge accumulated through repeated inductive inferences. Over time, this perspective has become widely accepted, sometimes even regarded as unequivocally true. However, an exclusive reliance on inductive reasoning can lead to significant errors in judgment. As we discuss below, such errors may be contributing factors to the emergence and persistence of pathological polarization in society.

A classical demonstration of the limitations of inductive reasoning was presented by Wason (1960). In his experiment, participants were asked to infer the rule behind a sequence of numbers -2, 4, and 6- by proposing additional numbers and receiving feedback on whether their suggestions conformed to the rule. After several trials, participants were asked to articulate the rule they believed generated the original sequence. Despite producing numbers consistent with the rule, most failed to identify its simplicity: the rule was merely "ascending numbers." The results revealed a cognitive bias toward confirmation —participants persistently tested hypotheses that supported their initial assumptions, rather than exploring alternative explanations. In other words, they exhibited a form of overconfidence in their immediate experience and inductive conclusions. This tendency to confirm rather than challenge one's beliefs has profound implications for understanding polarization. As Honner (2019) has recently argued, similar reasoning pitfalls continue to challenge default reliance on inductive logic, especially when individuals interpret complex social realities through narrow experiential lenses. Such cognitive rigidity may contribute to the entrenchment of polarized attitudes, reinforcing belief systems that resist revision even in the face of contradictory evidence.

The example discussed above also serve as simplified models of scientific reasoning, particularly when scientific methods rely on evidence acquisition to develop theories and explanations of nature (Mynatt et al., 1977). While it is true that scientific inquiry draws upon both inductive and deductive reasoning (Box, 1976), it is equally true that scientists are not immune to cognitive biases that can distort these processes. As Nuzzo (2014) points out, even within scientific practice, researchers may inadvertently mislead themselves by over-relying on confirmatory evidence. The underlying mechanism at play here

is known as confirmation bias —the tendency to seek, interpret, and generate information that supports pre-existing beliefs or explanations (Hogg & Vaughan, 2018, p. 9). This bias, deeply embedded in human cognition, is one of the key problematic elements that may contribute to the emergence and persistence of pathological polarization.

When we lose the capacity to argue with ourselves, we become vulnerable to polarization. This vulnerability is deeply rooted in the difficulty of counteracting confirmation bias—a cognitive tendency that is particularly potent because inductive reasoning, one of the foundational processes through which we construct our understanding of the world, is highly susceptible to it. When we rely solely on what appears self-evident to our minds, we risk reinforcing existing beliefs without critical scrutiny, thereby deepening polarization. As Jost et al. (2022) have noted, many of the cognitive biases that fuel pathological political polarization are closely linked to confirmation bias. These include selective exposure to information, selective avoidance, biased assimilation, disconfirmation bias, and the backlash and backfire effects—all of which distort reasoning and entrench belief systems. This is especially concerning in the digital age, where information ecosystems amplify these distortions. Vosoughi et al. (2018) found that falsehoods are 70% more likely to be shared on social media than truthful content. When misinformation is propagated without verification, the mechanisms of polarization are intensified. As Von Solms and Van Nierkerk (2013) aptly observed, the core issue in virtual environments is not merely *data integrity*, but *information integrity* —a distinction that underscores the epistemic fragility of our current media landscape.

Conclusions

As suggested at the outset, adopting a neutral stance may be one of the most effective strategies for mitigating pathological polarization in society. Kirmayer (2024) proposes that cultivating a form of healthy skepticism can help prevent the emergence of extremist and polarized opinions. Yet, as discussed throughout this essay, such neutrality is difficult to maintain and may carry its own costs. At the very least, we hope the remedy is not worse than the disease. René Descartes (1912, p. 20) would likely endorse this moderating approach, having written:

> amid many opinions held in equal repute, I chose always the most moderate, as much for the reason that these are always the most convenient for practice, and probably the best (for all excess is generally vicious), as that, in the event of my falling into error, I might be at less distance from the truth than if, having chosen one of the extremes, it should turn out to be the other which I ought to have adopted.

In this article, we have addressed two key factors that may contribute to pathological polarization. We define pathological polarization as a toxic form of societal clustering that undermines social harmony and promotes antagonism. First, we introduced the gregariousness/individualism tension (GIT) as a

theoretical framework to account for psychological mechanisms that may lead to polarization. Second, we examined inductive reasoning as a cognitive process vulnerable to biases -particularly confirmation bias- that may exacerbate polarized thinking. While both proposals are speculative, we have argued that they offer conceptual value and may serve as fruitful starting points for future research. The question of how to effectively counteract pathological polarization remains open. It would be presumptuous to offer definitive solutions based solely on the ideas presented here. Nevertheless, in the following paragraphs, we briefly discuss some promising approaches to reducing polarization, drawing on recent work in this area.

As Descartes (1912) suggested nearly four centuries ago -his *Discourse on the Method* was originally published in 1637- moderation is a prudent strategy for navigating conflicting opinions. Echoing this sentiment, Ben-Porath (2024) emphasizes that seeking middle ground is a crucial starting point for overcoming polarization. She proposes fostering environments where individuals can engage openly across differences, creating opportunities for dialogue that cultivate appreciation for diversity and generate positive emotional responses to disagreement. This approach reframes depolarization not as a matter of solving problems or directing people toward a predefined center, but as a shared journey toward mutual understanding. In educational contexts, Ruiz-Ruano and Puga (2024) offer a compelling metaphor: educators should act as DJs, mixing seemingly incongruent tracks and bringing together what appears dissimilar. Mistakes are not only expected but welcomed in the process of building an open-minded view of our complex social reality. Those involved in depolarization efforts must be prepared to acknowledge and correct their own errors. While the process can be challenging, Ben-Porath (2024) cautions against avoiding difficult or contentious topics. The key lies in maintaining a constructive attitude—one that prioritizes dialogue over division. In musical terms, the goal is not simply to reach the middle, but to achieve *social harmony*. Ultimately, Ben-Porath advocates for the creation of safe environments that invite individuals to reconsider their beliefs. As Sharot et al. (2023) suggest, belief change is fundamentally driven by environmental change. If we succeed in transforming the conditions under which beliefs are formed and maintained, we may begin to erode the rigid structures of polarized thinking.

The task of depolarization is far from straightforward, as it involves the pursuit of *collective epistemic success* —a concept that, as Jung et al. (2019) suggest, requires shared representations of reality across diverse social groups. Achieving such consensus is particularly challenging when different communities hold conflicting official narratives about facts or historical events. While some research indicates that group dynamics may facilitate depolarization, it remains uncertain whether these strategies can be effectively scaled to the level of cities or nations (May Sidik, 2023). Yet, history offers hopeful examples: societies have, at times, found ways to walk together -hand in hand- even after civil wars. Still, balance is fragile. Equilibrium is hard to achieve and easily lost. For this reason, society must remain vigilant, continuously monitoring its progress toward collective epistemic success. This shared

horizon -grounded in dialogue, mutual understanding, and epistemic humility- must guide our efforts if we are to move forward in harmony.

References

Ajzen, I., & Fishbein, M. (1980). *Understanding attitudes and predicting social behavior*. Prentice-Hall.

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In: D. Albarracín, B. T. Johnson, & M. P. Zanna (eds.), *The handbook of attitudes* (pp. 173-221). Lawrence Erlbaum Associates Publishers.

Allport, G. W. (1935). Attitudes. In: C. Murchison (ed.), *Handbook of social psychology* (pp. 798-844). Clark University Press.

Ball, P., & Maxmen, A. (2020, May 28). The epic battle against coronavirus misinformation and conspiracy theories. *Nature, 581*(7809), 371-374. https://doi.org/10.1038/d41586-020-01452-z

Barabási, A. L. (2009, July 27). Scale-free networks: a decade and beyond. *Science, 325*(5639), 412-413. https://doi.org/10.1126/science.1173299

Barabási, A. L., & Albert, R. (1999, October 8). Emergence of scaling in random networks. *Science, 286*(5439), 509-512. https://doi.org/10.1126/science.286.5439.509

Barabási, A. L., & Bonabeau, E. (2003). Scale-free networks. *Scientific American, 288*(5), 50-59, https://doi.org/10.1038/scientificamerican0503-60

Ben-Porath, S. (2024). How to depolarize your students. *Nature Human Behavior*, *8*(2), 186-189. https://doi.org/10.1038/s41562-023-01801-8

Berger, L. B., & Luckmann, T. (1967). *The social construction of reality. A treatise in the Sociology of Knowledge*. Penguin Books.

Box, G. E. P. (1976). Science and statistics. *Journal of the American Statistical Association*, *71*(356), 791-799. https://doi.org/10.1080/01621459.1976.10480949

Boxell, L., Gentzkow, M., & Shapiro, J. M. (2024). Cross-country trends in affective polarization. *The Review of Economics and Statistics*, *106*(2), 557-565. https://doi.org/10.1162/rest_a_01160

Converse, P. E. (1958). The shifting role of class in political attitudes and behavior. In: E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (eds.), *Readings in social psychology, 3rd edn* (pp. 388-399). Holt, Rinehart and Winston.

Descartes, R. (1912). *A discourse on method. Meditations and principles.* Everyman's Library.

Feldman, R. S. (1998). *Social psychology*. Prentice Hall.

Fishbein, M., & Ajzen, I. (1975). *Belief, attitude, intention, and behavior: An introduction to theory and research*. Addison-Wesley.

Frances, A. J., & Widiger, T. (2012). Psychiatric diagnosis: Lessons from the DSM-IV past and cautions for the DSM-5 future. *Annual Review of Clinical Psychology*, *8*, 109-130. https://doi.org/10.1146/annurev-clinpsy-032511-143102

Franzoi, S. L. (2005). *Social psychology*. McGraw Hill.

Gil, J. (2021). Tensión individualismo-gregarismo en la configuración psicológica del ser humano I: apuntes sobre el surgimiento y las variaciones históricas del yo. *International Journal of Psychology and Psychological Therapy*, *21*(3), 261-288. https://www.ijpsy.com/volumen21/num3/583.html

Gil, J. (2022). Tensión individualismo-gregarismo en la configuración psicológica del ser humano, II: Arquetipos de yo gregario. *International Journal of Psychology and Psychological Therapy*, *22*(2), 99-141. https://www.ijpsy.com/volumen22/num2/611.html

Gil, J. (2024). Tensión individualismo-gregarismo en la configuración psicológica del ser humano, III: emergencia de configuraciones protoindividualistas. *International Journal of Psychology and Psychological Therapy*, *24*(3), 309-344. https://www.ijpsy.com/volumen24/num3/668.html

Greene, G. (2004). *The quiet American*. Vintage.

Hogg, M. A., & Vaughan, G. M. (2018). *Social psychology*. Pearson.

Honner, P. (March 14, 2019). Where proof, evidence and imagination intersect. *Quantamagazine*. https://www.quantamagazine.org/where-proof-evidence-and-imagination-intersect-in-math-20190314/

Huxley, A. (1946). *Brave new world*. Harper & Row.

Jeffreys, H. (1931). *Scientific inference*. Cambridge University Press.

Jost, J. T., Baldassarri, D. S., & Druckman, J. N. (2022). Cognitive-motivational mechanisms of political polarization in social-communicative contexts. *Nature Reviews Psychology*, *1*(10), 560-576. https://doi.org/10.1038/s44159-022-00093-5

Jung, J., Grim, P., Singer, D. J., Bramson, A., Berger, W. J., Holman, B., & Kovaka, K. (2019). A multidisciplinary understanding of polarization. *American Psychologist*, *74*(3), 301-314. https://doi.org/10.1037/amp0000450

Kirmayer, L. J. (2024). The fragility of truth: Social epistemology in a time of polarization and pandemic. *Transcultural Psychiatry, 61*(5), 701-713. https://doi.org/10.1177/13634615241299556

Ladyman, J., Lambert, J., & Wiesner, K. (2013). What is a complex system? *European Journal for Philosophy of Science*, *67*(1), 33-67. https://doi.org/10.1007/s13194-012-0056-8

Lenci, K. (2023). *Learning to depolarize. Helping students and teachers reach across lines of disagreement*. Routledge.

Makri, A. (2024). Anxiety, depression, headaches - Is political polarization bad for your health? *Nature Medicine, 30*(8)*,* 2099-2102. https://doi.org/10.1038/s41591-024-03136-x

May Sidik, M. (2023, March 2). How to tackle political polarization - the researchers trying to bridge divides. *Nature, 615*(7950), 26-28. https://doi.org/10.1038/d41586-023-00573-5

Minson, J. A., & Dorison, C. A. (2022). Toward a psychology of attitude conflict. *Current Opinion in Psychology*, *43*, 182-188. https://doi.org/10.1016/j.copsyc.2021.07.002

Mynatt, C. R., Doherty, M. E., & Tweney, R. D. (1977). Confirmation bias in a simulated research environment: An experimental study of scientific inference. *Quarterly Journal of Experimental*

Psychology, *29*(1), 85-95. https://doi.org/10.1080/00335557743000053

Nuzzo, R. (2014) Scientific method: Statistical errors. *Nature*, *506*(7487), 150-152. https://doi.org/10.1038/506150a

Patent, V. (2022). Dysfunctional trusting and distrusting: Integrating trust and bias perspectives. *Journal of Trust Research*, *12*(1), 66-93. https://doi.org/10.1080/21515581.2022.2113887

Pavlov, I. P. (1927) *Conditioned reflexes*. Oxford University Press.

Ruiz-Ruano, A. M., & Puga, J. L. (2022). *Review of: Building a digital republic to reduce health disparities and improve population health in the United States*. Qeios. https://doi.org/10.32388/H38R8Q

Ruiz-Ruano, A. M., & Puga, J. L. (2024). *Last CLASS a DJ saved my life: The mixing role of contemporaneous educators*. Zenodo. https://doi.org/10.5281/zenodo.14025914

Ruiz-Ruano, A. M., Sánchez-Kuhn, A., Flores, P., & Puga, J. L. (2023). Social expectancy increases skin conductance response in mobile instant messaging users. *Psicothema, 35*(4), 414-422. https://doi.org/10.7334/psicothema2022.362

Sharot, T., Rollwage, M., Sunstein, C. R., & Fleming, S. M. (2023). Why and when beliefs change. *Perspectives on Psychological Science*, *18*(1), 142-151. https://doi.org/10.1177/17456916221082967

Sherif, M. (1958). Group influences upon the formation of norms and attitudes. In: E. E. Maccoby, T. M. Newcomb, & E. L. Hartley (eds.), *Readings in social psychology, 3rd edn* (pp. 219-232). Holt, Rinehart and Winston.

Von Solms, R., & van Niekerk, J. (2013). From information security to cyber security. *Computers and Security, 38,* 97-102. https://doi.org/10.1016/j.cose.2013.04.004

Vosoughi, R., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146-1151. https://doi.org/10.1126/science.aap9559

Walter, E., Woodfor, K., McIntosh, C., Francis, G., Cranz, J, Glennon, L., Side, R., McCarthy, M. & Martinez, R. (2008). *Cambridge advanced learner's dictionary*. Cambridge University Press.

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly Journal of Experimental Psychology*, *12*(3), 129-140. https://doi.org/10.1080/17470216008416717

Wason, P. C. (1966). Reasoning. In: B. Foss (ed.), *New horizons in psychology* (pp. 131-151). Penguin Books.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of 'small-world' networks. *Nature, 393*(6684), 440-442. https://doi.org/10.1038/30918