The Polarization Error: An Analysis of the Misattribution of Definitional and Instrumental Disputes as Axiological Conflicts in Public Discourse

Tomáš KOLLÁRIK

Abstract

This study analyzes a specific cognitive bias and argumentative fallacy for which the author introduces the original term polarization error. It is a faulty abductive inference in which an agent interprets a disagreement at the non-axiological level (instrumental or definitional) as a fundamental disagreement at the axiological level. The aim of this paper is not to deny the existence of genuine value conflicts, but to demonstrate how pseudo-value disputes imperceptibly intermingle with legitimate conflicts in a polarized society. The author argues that the essence of this error lies in the projection of one's own knowledge onto the opponent and the attribution of bad faith, which leads to a form of epistemic misrecognition (in the sense of A. Honneth). Failure to recognize this type of dispute transforms political agonism into destructive antagonism and sets in motion the polarization spiral. The study illustrates this mechanism through examples from contemporary public discourse and proposes a corrective through the Principle of Charity as a necessary condition for functional agonism.

Keywords

polarization error, pseudo-value disputes, recognition theory, agonism, abductive inference, polarization spiral

1. Introduction: A Rift That Need Not Be

The current social atmosphere in Slovakia, as well as in the broader geopolitical context of Western democracies, is characterized by a deep and paralyzing rift. This rift, which cuts across families, workplaces, and public spaces, is not merely a matter of opinion; it has acquired a distinctly moral character. Hostile factions no longer perceive each other simply as groups with different political programs or economic preferences, but as bearers of moral evil. An opponent in contemporary discourse is not someone who holds a different view on tax reform or public administration; an opponent is someone who "hates the country," "supports murderers," or "wants to destroy civilization." The consequence of this mindset is a refusal of basic respect, an inability to listen to the other side's arguments, and an a priori unwillingness to seek compromise—because compromise with evil is perceived as moral failure.

The present study does not aim to claim that all social conflicts are merely misunderstandings. Its contribution lies in identifying a specific type of dispute—pseudo-value disputes—which easily and frequently intermingle with genuine conflicts in a polarized atmosphere. These disputes are not grounded in a fundamental conflict of values but in a specific error of judgment, for which we introduce the term polarization error. This error consists in the misattribution of an opponent's motivations and in the projection of one's own cognitive models onto their reasoning. The danger is that if society cannot distinguish genuine value disputes from pseudo-value disputes, the polarization spiral is needlessly set in motion, blocking resolution even of problems where agreement would be possible.

The aim of this text is to analytically define the essence of the polarization error as a failure of abductive reasoning, to distinguish it from established cognitive biases (FAE, Straw Man), to situate it within the framework of recognition theory (Honneth) and agonism (Mouffe), and to illustrate its operation through concrete examples.

2. The Polarization Error: Theoretical Framework

2.1. The Nature of the Error: Failure of Abductive Inference

For the purposes of this paper, we define the polarization error as a faulty abductive inference (inference to the best explanation), the outcome of which is a conclusion about the opponent's moral failure (axiological level), even though the actual difference lies in a different cognitive model (a set of beliefs about facts and causality) from which the opponent operates.

The logical structure of this error consists not merely in projecting one's own model of reality, but specifically in the substitution of the opponent's unstated premises. The agent takes the opponent's enthymeme (an argument with suppressed parts) and fills in the blanks not with the opponent's premises, but with premises derived from the agent's own cognitive model. The reasoning process unfolds as follows:

1. Observed phenomenon (Fact): The opponent proposes solution B.
2. Substitution of Unstated Premises (Faulty projection): The opponent's proposal is an enthymeme relying on a suppressed premise (e.g., *"B leads to a positive outcome"*), which makes it rational within their model. The agent ignores this premise. Instead, the agent fills the gap with a premise from their own knowledge system, in which solution B leads to catastrophe. The agent fails to realize that the opponent might genuinely hold a different opinion or operate within a distinct framework, and therefore projects their own causal logic onto the opponent: *"The opponent also knows (what I know) that B leads to catastrophe."*
3. Faulty abduction (Conclusion): The agent seeks the best explanation for the opponent's behavior. The reasoning proceeds: *"If the opponent possesses the same information as I do (knows that B is harmful), yet proposes B nonetheless, then the only explanation is that the opponent desires harm (is evil)."*

A key aspect of this failure is the attribution of bad faith. The agent does not believe in the authenticity of the opponent's arguments. Since from the agent's perspective the negative consequences of the opponent's proposal are obvious and indisputable (because the agent does not see the opponent's different cognitive model), they reject the possibility that the opponent is merely mistaken. The reasoning leads to the conclusion: "They are not stupid, so they must be evil." The opponent's rational arguments are then perceived merely as a "smokescreen" (rationalization) for their true, nefarious intentions.

This mechanism corresponds with findings from moral psychology, specifically Jonathan Haidt's (2012) social intuitionist model. Haidt employs the metaphor of the elephant (automatic, emotional, and intuitive processes) and the rider (controlled, rational processes). He argued that moral judgments do not arise from rational analysis but are primarily the result of rapid intuitions of the "elephant." The role of the "rider" (reason) is often not to steer but merely to defend and rationalize what the elephant has already instinctively decided. In this context, the polarization error is a manifestation of such an a priori intuition ("the opponent is evil") that precedes comprehension of their arguments. Dan Kahan (2012) complements this picture with his theory of cultural cognition. He demonstrates that people do not analyze facts neutrally but filter them through the need to protect their group identity. The polarization error thus arises at the moment when an agent fails to see that the opponent is not denying "obvious facts" out of malice, but is operating with a different set of "facts" (a different model of reality) that have passed through a different identity filter.

2.2. Socio-Philosophical Context: The Denial of Recognition

The polarization error is not merely an isolated cognitive error but has a fundamental impact on the social ontology of relationships. In the terminology of Axel Honneth (1995), we can interpret this phenomenon as a specific form of misrecognition, specifically in the sphere of legal/moral recognition.

This concerns primarily the sphere of legal recognition, in which Honneth defines an individual's moral integrity through the capacity to be recognized as a rational agent capable of moral judgments. Once we commit the polarization error, we not only perceive the opponent negatively but actively reclassify them. We no longer consider them an equal partner who is mistaken (which would be a legitimate dispute at the level of rationality), but an agent with malicious intentions whose arguments are irrelevant. Through this cognitive act, we legitimize the denial of their status as a moral agent. The polarization error thus functions as a mechanism that justifies the transition from discussion to combat—the opponent is no longer a subject of rights but an object with whom one does not argue but whom one suppresses.

2.3. Relationship to Established Cognitive Biases: Logical, Cognitive, and Argumentative Profiles

For precise theoretical positioning of the polarization error, it is necessary to distinguish three levels of analysis. The polarization error is not identical with the fallacies discussed below but represents a specific phenomenon that can be examined from the perspectives of logic (propositions), cognition (mind), and argumentation theory (dialectical interaction).

A. **Logical Profile (Propositional Structure):** At this level, we abstract from social context and examine the purely logical structure of the inference. At its core is a faulty abductive inference consisting in incorrect selection among competing hypotheses.

B. **Cognitive Profile (Social Cognition):** The error in selecting the correct hypothesis relates to the phenomenon of the *fundamental attribution error* **(FAE)** (Ross 1977) and to the theory of **naive realism** (Ross and Ward 1996). **Naive realism explains the agent's conviction in the objectivity of their own perception.** The error in judgment results from incorrect handling of the variables *cognition* (factual belief) and *intention* (purpose). The inference erroneously derives *evil intent* from the premise of *harmful consequence*, while in the process of abduction it illegitimately excludes the variable of

different cognition (error or different model). From the perspective of social cognition, it is a specific subspecies of FAE that focuses exclusively on moral dispositions (good vs. evil) in the context of ideological dispute and is conditioned by epistemic projection (the assumption of shared knowledge). While FAE says "they are late because they are irresponsible," the polarization error says "they vote for X because they want to destroy the country."

C. **Argumentative Profile (Dialectical Level)**: From the perspective of argumentation theory, the polarization error is a specific violation of the rules of discourse. At this level, it overlaps with classic argumentative fallacies:

- **Straw Man**: This fallacy consists in replacing the original argument with a weaker argument that is much easier to refute (creating a caricature of the position).
- **Ad Hominem and Poisoning the Well**: In the dynamics of the polarization error, the Straw Man opens the possibility of distorting the opponent's motivation and character. The distorted interpretation of the thesis (Straw Man) serves as a necessary springboard: only when the opponent's position appears absurd or harmful does space emerge for constructing a fictitious (evil) motive. Since the opponent is (based on the logical error) identified as an agent with malicious intentions, their participation in discussion is a priori illegitimate. The argumentative move of "poisoning the well" here serves not merely as a rhetorical trick but as a defense mechanism against perceived evil.

2.4. Analytical Tool and Heuristic

How do we distinguish when we are dealing with a polarization error (pseudo-value dispute) and when with a genuine conflict of values? We propose a simple heuristic based on the test of shared axiological maximum.

1. Axiological level: Do the parties agree on the ultimate goal? (E.g., "We want innocent people not to die.")

2. Instrumental/Definitional level: Do they differ in their proposed approach to resolving the situation, or in the definition of relevant terms through which the dispute is described?

- **Example of pseudo-dispute (Polarization error)**: If both parties want prosperity, but one is convinced of an approach involving low taxes while the other advocates an approach involving high taxes, this is an instrumental dispute. Accusations of "theft" or "social callousness" constitute a polarization error.
- **Example of genuine value conflict**: If one side claims that "Race X is superior to Race Y and has the right to subjugate it," while the other maintains that "All human beings are equal," there is no shared axiological maximum. This is not an error in knowledge (cognitive model) but a fundamental conflict of values. The polarization error applies primarily to the first type of dispute, which nevertheless dominates democratic discourse while being masked as the second type.

3. Application: False Value Conflicts in Practice

The following section applies the above model to four key topics. The aim is to demonstrate how pseudo-value disputes intermingle with public discourse.

3.1. Instrumental Dispute: The War in Ukraine

Public debate about the geopolitical conflict in Ukraine is predominantly framed through emotive labels such as "warmonger" versus "Russophile" or "traitor." The rhetoric of both sides suggests that the adversary desires death and suffering. Upon deeper analysis, however, we find that the overwhelming majority of the population shares the same values at the axiological level: Peace is qualitatively better than war; the security of one's own state is a priority; human life should not be wasted.

The dispute is actually conducted at the instrumental level, which stems from diametrically different interpretations of the situation (different causal models). On one side stands Model A,

which identifies the cause of war as the imperial expansion of an aggressor. In this model, the instrumental premise holds: "Any concession (B) leads to strengthening the aggressor and ultimately to greater suffering and expansion of the conflict." On the other side stands Model B, which sees the cause in the geopolitical provocation of great powers. In this model, the premise holds: "Supplying weapons (A) merely prolongs suffering without any chance of military victory."

The polarization error occurs at the moment of projection: a proponent of Model A believes that a proponent of Model B knows (possesses the same insight) that concession will strengthen the aggressor, yet advocates concession nonetheless. The logical conclusion of this faulty inference is that the opponent must be a traitor who desires the fall of their own country. What is ignored is the fact that in the opponent's cognitive model, concession is precisely the path to preserving lives.

3.2. Instrumental Dispute: The Brown Bear

A similar dynamic operates in the topic of nature conservation, which is highly polarizing in the Central European context. The discourse pits "eco-terrorists" against "nature killers." Yet here too there exists a surprising consensus at the axiological level: It is a tragedy when a bear kills a person (protection of life). It is morally wrong when a person kills protected animals without reason (protection of nature).

The dispute is primarily about diagnosing the cause (different interpretation of reality). One group operates with data according to which the cause of conflicts is changes in human behavior (garbage, bait stations), and therefore culling does not solve the problem. The other group proceeds from a model in which the cause is biological overpopulation exceeding the carrying capacity of the territory, and therefore population management is necessary for safety. A nature conservationist commits the polarization error if they think the hunter wants to kill out of bloodthirsty pleasure. A hunter is wrong if they think the conservationist does not care about people. Both sides miss each other in their interpretation of facts, not in their basic values.

3.3. Definitional Dispute: Abortion

The topic of abortion appears to be the purest value conflict. It is often framed as a conflict between the "culture of death" (child murderers) and "enemies of freedom" (enslavers of women). Paradoxically, even here both mainstream camps agree on the axiological maximum: Human life is valuable and should not be ended without reason. The disagreement lies at the definitional level: From when does a fetus have the status of a person? For the pro-life camp, the starting point is from conception; for the pro-choice camp, this status is acquired gradually.

It is necessary to address the objection that this is not merely about definition but about a genuine conflict in the hierarchy of values (woman's autonomy vs. fetal life). Many indeed acknowledge that abortion is wrong but prioritize freedom of choice. However, this hierarchy is sustainable only because of the different definitional starting point. If proponents of choice defined a fetus as a born five-year-old child, the autonomy argument would lose its weight (society does not accept killing a child for the convenience or freedom of a parent). The fact that abortion is perceived as an "acceptable evil" thus still stems from the conviction that the fetus has not yet achieved full ontological status.

The polarization error in this case consists in questioning the sincerity of the opponent's conviction. A proponent of prohibition often assumes that the opponent deep down knows it is a child but suppresses this knowledge and accepts the developmental definition merely as expedient rationalization. Recognition that the opponent authentically believes in a different definition of reality (and therefore is not a "murderer" in the intentional sense) is key to reducing polarization.

3.4. Economic Dispute: Taxes and Redistribution

The dispute over tax rates is a classic example of an apparent moral conflict. Proponents of higher taxes are labeled thieves; proponents of lower taxes are called social monsters. In reality, the axiological goal of both groups is identical: a prosperous society, low unemployment, and a high standard of living.

The dispute is purely instrumental, based on different theoretical models. Model A (Keynesian) maintains that the path to prosperity is demand stimulation and public investment. Model B (neoclassical) maintains that the path is supply stimulation and low tax burdens. The polarization error arises from substituting instrumental strategy for moral intent—we assume that if someone chooses a tool that (in our view) harms the poor, then their goal is to harm the poor.

4. Escalation of Dispute: From Media Image to the Spiral of Violence

4.1. Framing and the Reproduction of the Error

The polarization error is not merely an individual psychological failure but is systematically reproduced and amplified by the media environment through framing. According to Robert Entman (1993), framing involves the selection of certain aspects of reality and making them salient, thereby defining the problem and moral evaluation.

In the context of political dispute, media often select conflict potential and the moral consequences of a decision because these generate greater attention. Conversely, the instrumental logic (the complex cognitive premises and models) that led an agent to their decision is often suppressed or entirely omitted as too technical. The recipient thus receives information that is structurally incomplete: they see the negative consequence of the opponent's action but lack access to its epistemic justification. This gap is automatically filled by the recipient through the polarization error—since they do not see the rational reason (instrumental logic), they assume an irrational or evil motive.

Let us illustrate this with an example of fiscal consolidation. If a politician proposes reducing social transfers based on an economic model that predicts this will stimulate employment in the long term (instrumental logic), the media output is often limited to the headline: "Government takes money from the poorest". This frame omits the causal mechanism of the measure and emphasizes only its immediate negative impact. The recipient, who is denied access to the instrumental logic (why are they doing this?), fills the empty space in motivation with the

simplest explanation: "They are doing it because they are antisocial and evil". Media thus not only reflect polarization but industrially produce it by delivering to the public ready-made "conclusions about the moral defect" of opponents without providing the key to their thinking.

4.2. The Polarization Spiral and the Legitimization of Violence

Here a fundamental shift occurs: the distorted image of reality (the level of media representation) enters the real world and shapes the behavior of political actors (the level of interaction). Once actors adopt this media-reinforced image of the opponent as "evil," interaction ceases to be dialogue and transforms into a mechanism we may call the polarization spiral (a mechanism related to Sunstein's "law of group polarization", 2002).

The mechanism proceeds as follows:

1. Actor A (under the influence of the polarization error) labels Actor B as "evil."

2. Based on this designation, Actor A adopts radical measures (because one does not negotiate with evil; evil is suppressed), often violating ordinary democratic norms.

3. Actor B observes this radical behavior by Actor A. Since they do not see the faulty inference behind it (do not see that A is acting in self-defense against perceived evil), they interpret this behavior as proof that Actor A is an aggressor and the real "evil."

4. Actor B responds with a counterattack, which in turn confirms and reinforces Actor A's original prejudice.

The spiral winds toward the legitimization of violence. If an opponent is perceived through the lens of the polarization error as the embodiment of evil, violence against them (for example, an assassination attempt on a politician) transforms in the eyes of a radicalized individual from a crime into an act of heroism and salvation of society.

5. Conclusion: From Charity to Agonism

The aim of this study was not to naively call for friendship in politics. The analysis of the polarization error is meant, paradoxically, to enable genuine conflict.

Here we arrive at an apparent normative tension between the Principle of Charity and Chantal Mouffe's (2013) concept of agonism. Mouffe criticizes liberal consensus and argues that democracy needs conflict. She distinguishes, however, between antagonism (struggle between enemies who seek to destroy each other) and agonism (struggle between opponents who recognize the legitimacy of their conflicting claims).

Our thesis is: The elimination of the polarization error through the Principle of Charity is a necessary condition for the transition from antagonism to agonism. As Aikin and Talisse (2014) remind us, the manner in which we conduct disputes defines the quality of our democracy. If we conduct disputes in the "tone of war," democracy fails. Robert Talisse (2019) further warns against "overdoing democracy," whereby politics colonizes even non-political spheres of life, precisely because we see moral enemies in our opponents.

The Principle of Charity (a concept developed by Donald Davidson 1984, with roots in the work of W. V. O. Quine 1960) is understood here not as an ethical imperative to "be nice" but as an epistemic necessity for interpretation. If we attribute to an opponent only irrationality or pure evil (polarization error), we cannot engage them in agonistic dispute—we can only destroy them (antagonism). Only if we assume that their positions stem from a coherent (though in our view mistaken) model of the world does the opponent become a legitimate adversary. Correcting the polarization error thus does not lead to the abolition of conflict but to its civilization.

List of sources used

AIKIN, Scott F. & Robert B. TALISSE, (2014). Why We Argue (And How We Should): A Guide to Political Disagreement. New York: Routledge.

DAVIDSON, Donald, (1984). Inquiries into Truth and Interpretation. Oxford: Clarendon Press.

ENTMAN, Robert M., (1993). Framing: Toward Clarification of a Fractured Paradigm. Journal of Communication, vol. 43 (1993), no. 4, pp. 51–58.

HAIDT, Jonathan, (2012). The Righteous Mind: Why Good People Are Divided by Politics and Religion. New York: Pantheon Books.

HONNETH, Axel, (1995). The Struggle for Recognition: The Moral Grammar of Social Conflicts. Cambridge: Polity Press.

KAHAN, Dan M., (2012). Cultural Cognition as a Conception of the Cultural Theory of Risk. In: Sabine ROESER, Rafaela HILLERBRAND, Per SANDIN & Martin PETERSON (eds.). Handbook of Risk Theory. Dordrecht: Springer.

MOUFFE, Chantal, (2013). Agonistics: Thinking The World Politically. London: Verso.

QUINE, Willard Van Orman, (1960). Word and Object. Cambridge: MIT Press.

ROSS, Lee & Andrew WARD, (1996). Naive Realism in Everyday Life: Implications for Social Conflict and Misunderstanding. In: T. BROWN, E. REED & E. TURIEL (eds.). Values and Knowledge, pp. 103–135. Hillsdale, NJ: Lawrence Erlbaum Associates.

ROSS, Lee, (1977). The Intuitive Psychologist and His Shortcomings: Distortions in the Attribution Process. In: L. BERKOWITZ (ed.). Advances in Experimental Social Psychology, vol. 10, pp. 173–220. New York: Academic Press.

SUNSTEIN, Cass R., (2002). The Law of Group Polarization. Journal of Political Philosophy, vol. 10 (2002), no. 2, pp. 175–195.

TALISSE, Robert B., (2019). Overdoing Democracy: Why We Must Put Politics in Its Place. Oxford: Oxford University Press.

Author

Tomáš Kollárik

tomas.kollarik@savba.sk

Department of Normativity and Human Rights Research, Institute of Philosophy,

Slovak Academy of Sciences, v. v. i., Bratislava

Klemensova 19

811 09 Bratislava 1

Slovak Republic

Tomáš Kollárik studied philosophy at the Faculty of Philosophy of Comenius University in Bratislava, where he obtained his PhD. His research focuses on the theory of argumentation, autonomy, and questions of rationality in public discourse. He works as a researcher at the Department of Normativity and Human Rights Research at the Institute of Philosophy of the Slovak Academy of Sciences, v.v.i., and is engaged in analytical philosophy, critical thinking, and epistemic problems of polarization.