Beyond Justification: A Position on Intentional Framing in Natural Argumentation

Leonhard Schütz-Bauer[1,2]

[1] *University of Applied Sciences*
[2] *University of Vienna*

Abstract

I argue for the position that argumentation theory should treat justificatory reason-giving ("Rechtfertigen", legitimating a position before a normative order) and explanatory reason-giving ("Begründen", making transparent why one thinks p) as distinct orientations of intention, neither a matter of degree nor reducible to the commitments a speaker can be shown to have incurred. Argumentation's proper aim is transparency and truth rather than provability, so intentions resisting third-party verification does not undermine the distinction. It is what one should expect if the distinction is real. I argue for the stronger consequence that justificatory reason-giving is truth-obstructing whenever it substitutes for explanatory reason-giving, with one principled exception, and that this bears on how LLM output, optimized to be defensible rather than transparent, should be evaluated.

Keywords

argumentation theory, justification, explanation, intention, adversariality, large language models

1. The position

Argumentation theory's vocabulary (claims, verdicts, burden of proof) imports a juridic model in which a proponent defends a position before a normative authority. My position is that this framing, when it silently governs how an exchange is conducted rather than merely how it is described, can preclude the exchange's capacity to track truth, regardless of its logical form or surface cooperativeness. I locate the mechanism one level below adversariality as such [1, 2], in the justificatory intention: the orientation toward licensing a position rather than disclosing one's actual grounds for holding it.

2. Intentions, not commitments

A first objection asked how intentions can be determined, suggesting the model be anchored instead in externalized commitments, in the manner of pragma-dialectics. I resist this move. Commitment-tracking is already a move within the justificatory paradigm this paper opposes, since pragma-dialectics reconstructs argumentative moves as commitments incurred against procedural, resolution-oriented norms: it asks what a speaker is entitled or obliged to defend, not what she actually thinks. Answering the demand for intention with a commitment-based criterion concedes the paper's target thesis before 'defending' it, since it makes verifiability before an audience the price of theoretical legitimacy, which is the juridic move under critique. Pragma-dialectics, and Waltonian dialogue types insofar as they specify what a dialogue-type licenses a speaker to assert [3], belong on the justificatory side of the ledger this argument draws. They function as data for the distinction rather than as its foundation.

Submission for review
*Corresponding author.

2.1. The irrelevance of epistemic access

That intentions cannot always be determined from the outside is true and has no bearing on the distinction's reality or theoretical importance. Argumentation, on the position argued for here, answers to transparency and truth rather than provability. Whether an exchange in fact discloses what a speaker thinks does not depend on whether a third party, or an annotator can verify it. Demanding proceduralized, third-person-verifiable criteria as a condition of taking the distinction seriously reintroduces the juridic standard – justifiability before an audience – that the paper argues distorts natural argument. A friend's motive for an apology may be unknowable to me. That does not mean there is no difference between an apology meant to repair a relationship and one meant only to end an uncomfortable conversation. It means only that I may never find out which one I received.

2.2. The example, revised

[1]"It was my decision, I had every right to leave" is a justification in the sense at stake here, and it is not a refusal to argue rather than the former, as the earlier reviewer read it. On the stipulated sense used throughout, justification as legitimation, to assert a right is to assert legitimacy, which is what justifying a position consists in. The utterance succeeds as a justification by declining to explain the decision, offering entitlement in place of grounds rather than failing to give an explanation. This is the diagnostic case for the whole model. The same request for reasons can be met with a justification, which is legitimation, or an explanation, which is transparency about grounds. These are two different answers to one question, and neither is a stronger or weaker version of the other.

3. Justification is truth-obstructing, with one exception

I hold the stronger claim. Justifying that p, defending one's entitlement to assert or to have done p, is truth-obstructing wherever it substitutes for making transparent why one thinks p, in all cases and not merely some. Justification, on this stipulated sense, answers a different question than explanation does, so answering the wrong question cannot count as a weaker version of answering the right one. The qualification is a limiting case rather than a restriction of scope. Consider a speaker whose own standard of entitlement is honesty itself, for whom nothing legitimates a position except having actually disclosed her real grounds for it. For that speaker, Rechtfertigen and Begründen coincide, because legitimation is only available by way of transparency. Here the two orientations become extensionally identical while remaining conceptually distinct, and that is a limiting case of the thesis but not a clean counterexample to it.

4. Toward operationalization

However, this does not make the model empirically idle, but rather relocates what empirical work can responsibly claim. What resists proceduralization is the detection of orientation from surface behavior alone, since that project imports the commitment-based paradigm under critique and would make third-person verifiability the test of a distinction defined against that very standard. What remains tractable is the study of downstream, content-level consequences once orientation is fixed by independent means. Experimentally manipulated framings, self-reported orientation, naturally occurring cases where a speaker's aim is independently documented, such as mediation transcripts or retrospective interviews etc., can test whether justificatory versus explanatory orientation predicts what gets said next and how an exchange closes. The scheme's

[1]Attached below the literature section and taken from the initial paper.

validity criterion is replication of the content-level and trajectory effects the model predicts, given orientation fixed independently of the text being analyzed, rather than inter-annotator agreement on orientation read off the text itself, which would beg the question against the position argued for here.

5. Implications for evaluating LLM output

On the account argued for here, LLMs are justificatory in the relevant sense constitutively rather than derivatively. Training objectives that optimize outputs against evaluative, preference-based reward are training for legitimacy before an audience, which is the justificatory orientation by definition rather than a contingent tendency that better training could remove. Intention is not reducible to commitment or behavior, so whether an LLM explanatorily intends anything is arguably the wrong kind of question to put to a system with no standpoint from which to disclose what it thinks. There is no transparency to be had, only outputs shaped to survive challenge. This restores the paper's original diagnosis without needing to settle, from the outside, what if anything an LLM really has. The same refusal to demand third-person verification of intention that grounds §2 and §2.1 (from the original paper) applies here too.

6. Conclusion

The justification/reason-giving distinction is a distinction of intention rather than of provable commitment. Its resistance to third-party verification follows from taking argumentation's aim to be transparency and truth rather than legitimacy before an audience. I have restated the central example in its original form, argued for the stronger truth-obstructing thesis with one principled exception, and reinstated the claim that LLM output is justificatory by construction rather than by contingent shortfall. What remains for further work is empirical study of the downstream, content-level consequences of orientation once it is fixed by means independent of the text under analysis. Detecting intention from text is not a goal the position argued for here can accept.

Declaration on Generative AI

During the preparation of this work, the author(s) used ClaudeAI in order to: Grammar and spelling check. After using this tool, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the publication's content.

References

[1] J. Moulton. A Paradigm of Philosophy: The Adversary Method. In S. Harding and M. B. Hintikka (eds.), *Discovering Reality*, pages 149–164. Reidel, Dordrecht, 1983.

[2] D. Romizi. Adversariality in Argument(ation) and Its Implications for Teaching Argumentation Courses: A Reformist Approach. *Teaching Philosophy*, 49(1):61–95, 2026.

[3] D. Walton. *The New Dialectic: Conversational Contexts of Argument*. University of Toronto Press, 1998.

A. Diagrams referenced in §2.2

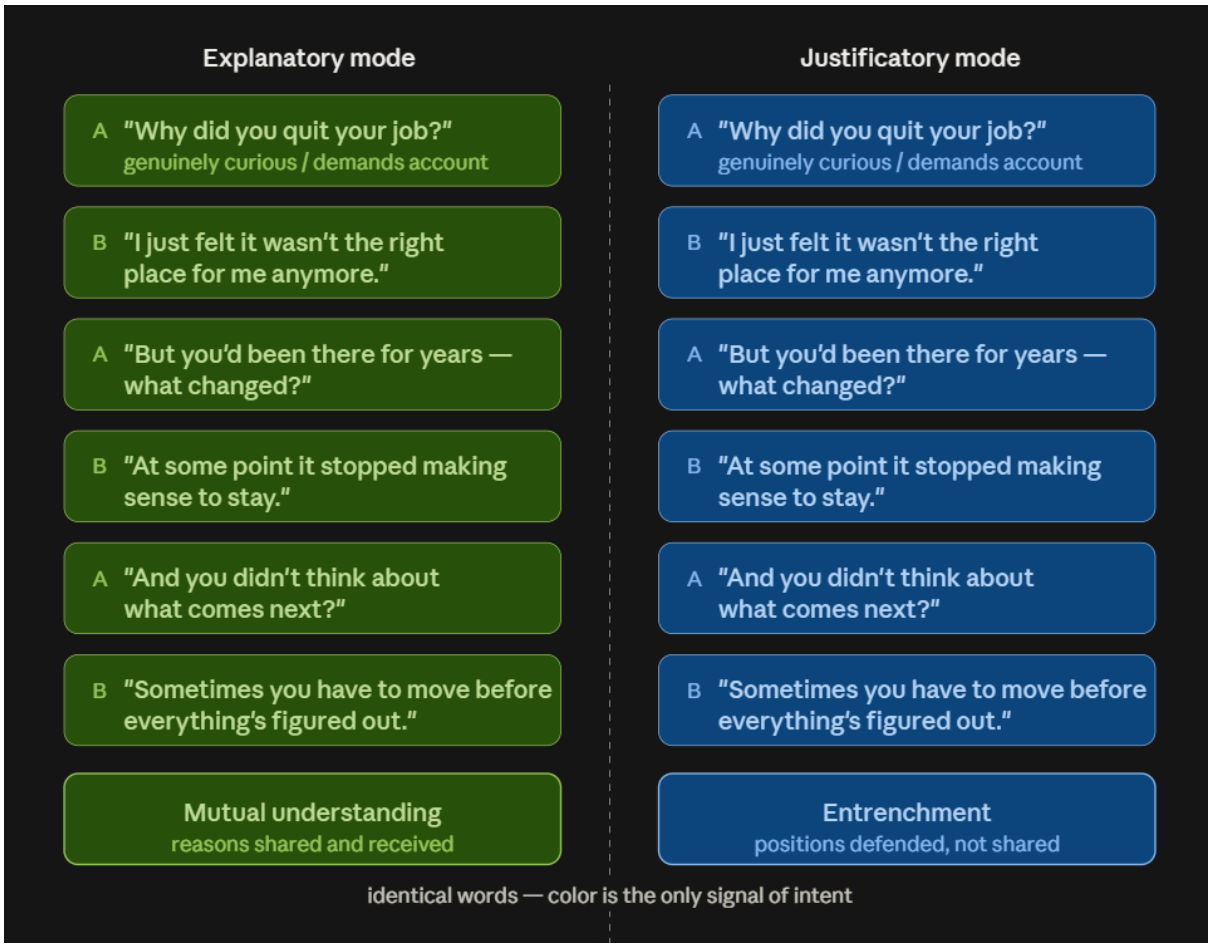

Figure 1: Explanatory vs. justificatory mode with identical surface wording.

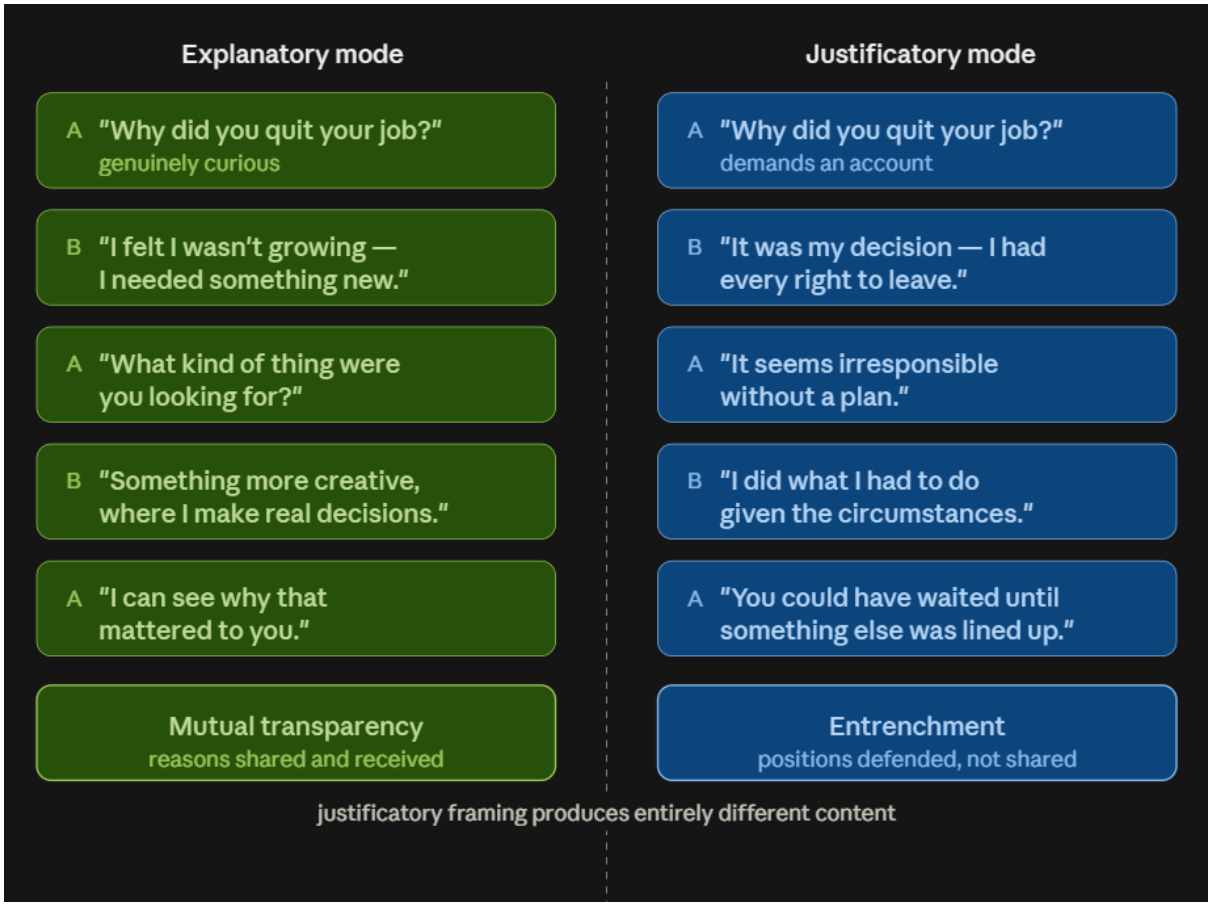

Figure 2: Explanatory vs. justificatory framing producing different content.