From Thought to Action: Rational Agency and the Normative Structure of Deliberation

Ranjeet Kumar Verma *

Posted Date: 5 May 2025

doi: 10.20944/preprints202505.0176.v1

Keywords: Rational Agency; Deliberation; Normativity; Practical Reason; Moral Responsibility; Autonomy

 doi:10.20944/preprints202505.0176.v1

Article

From Thought to Action: Rational Agency and the Normative Structure of Deliberation

Ranjeet Kumar Verma

Independent Researcher; mokshthephilosopher@gmail.com

Abstract: This paper explores the philosophical terrain between deliberative thought and intentional action, proposing a normative account of rational agency grounded in the structural integrity of practical deliberation. Drawing on Donald Davidson's theory of action and reasons as causes, the argument moves beyond causal accounts by integrating Christine Korsgaard's conception of practical identity and reflective endorsement. The paper contends that action is not merely the endpoint of desire-guided cognition but the realization of a normative commitment forged through self-governed deliberation. Through an analysis of Harry Frankfurt's hierarchy of volitions and John McDowell's notion of second nature, it is argued that the capacity to act is inherently tied to our responsiveness to reasons and our ability to see certain considerations as reasons for action. Furthermore, the paper examines how disruptions in this normative structure—such as weakness of will, moral conflict, or existential ambiguity—can be addressed through philosophical counseling understood not as advice-giving but as an exercise in restoring deliberative clarity. In this light, rational agency is framed as a dynamic achievement rather than a given faculty, emerging through an agent's ongoing negotiation with the space of reasons. By articulating the normative underpinnings of deliberation, this paper contributes to contemporary debates in moral psychology, action theory, and the philosophy of mind, offering an account of agency that is both analytically robust and existentially grounded.

Keywords: rational agency; deliberation; normativity; practical reason; moral responsibility; autonomy

1. Introduction

The capacity for human agents to move from reflection to action is not simply a matter of psychological transition or causal necessity; it is a deeply normative phenomenon that lies at the heart of rational agency. While empirical disciplines such as neuroscience and psychology often seek to explain action through causal chains—neuronal firings, affective states, or behavioral patterns—philosophy offers a more fundamental question: *What justifies the transition from thought to action as a rational act?* This paper contends that the structure of deliberation is not merely a cognitive computation or a sequence of preferences but a site of normative engagement, where reasons are weighed, endorsed, and authorized through self-reflection.

The central claim advanced here is that deliberation is best understood as a normatively structured activity, where agents act not because they are caused to do so but because they see themselves as bound by reasons they endorse. In this light, action emerges from the internal logic of normativity, rather than from mere desire, conditioning, or neural determinants. Donald Davidson famously proposed that reasons can be causes, arguing that "the primary reason for an action is its cause" (Davidson, 1963). While this insight was revolutionary in bridging mental content with physical causation, it left open the question of how reasons gain their *normative force*. Why should a reason cause action, not merely by pushing or pulling the agent, but by being *recognized as a reason*?

Christine Korsgaard provides a more internalist answer, arguing that "normativity arises from our nature as reflective beings" (Korsgaard, 2009). On her view, agents must endorse the principles they act upon as laws they give to themselves—what she calls practical identity, the roles and self-conceptions that confer meaning upon action. For instance, a person might refrain from lying not merely due to fear of social consequences, but because they identify as a truthful person; the reason gains its authority from being integrated into the agent's sense of self. "To act is to determine what

 doi:10.20944/preprints202505.0176.v1

counts as a reason, and to take that reason as your law," Korsgaard writes (1996, p. 100), placing agency within the domain of self-legislation.

Similarly, Harry Frankfurt's notion of higher-order volitions helps illuminate the inner structure of this normative commitment. According to Frankfurt (1971), what distinguishes persons from wantons is the ability to *identify with certain desires*—to endorse or reject them from a reflective standpoint. This capacity for volitional endorsement introduces a normative layer to motivation. It is not merely that we want something, but that we want to want it. The alignment of first-order desires with second-order volitions is the hallmark of freedom and autonomy in Frankfurt's framework. For example, if an individual desires to stop smoking and reflectsively endorses this desire as consistent with their values, the resulting action—quitting smoking—is not merely a product of causal desire, but a rational commitment.

Yet neither Korsgaard nor Frankfurt fully explains how *reasons themselves* acquire their justificatory force within a broader landscape of normativity. John McDowell advances this conversation by highlighting the idea of second nature, where rational capacities are not innate but cultivated through education and ethical upbringing (McDowell, 1996). He posits that mature human agents can respond to reasons "as such"—not simply as motivators but as intelligible demands within a shared ethical life. For McDowell, deliberation is not a computational weighing of options but an attunement to a normative space shaped by forms of life. In this sense, reason-responsiveness is not mechanistic but existentially embedded.

Consider a concrete example: a doctor deciding whether to disclose a terminal diagnosis to a patient. The decision is not governed by brute desire or fear, nor is it a simple cost-benefit analysis. Rather, it involves deliberation over reasons grounded in professional ethics, empathy, truthfulness, and respect for autonomy. The physician recognizes the authority of these reasons not as external pressures but as internalized commitments arising from their role and identity as a healer. This case exemplifies what this paper aims to defend: the deliberative transition from thought to action is authorized by the agent's endorsement of normative reasons, situated within a larger ethical and practical framework.

Thus, the phenomenon of acting for a reason cannot be fully accounted for by empirical science alone; it demands a philosophical account that honors the normative structure of deliberation. In what follows, this paper will articulate such an account by critically engaging with the theories of Korsgaard, Davidson, Frankfurt, and McDowell. It will argue that rational agency involves a dynamic synthesis of reflective endorsement, identity-constituting commitments, and norm-responsive attunement. Deliberation, therefore, is not a prelude to action but its condition of intelligibility.

By locating the source of action in the normativity of thought, this paper aims to contribute to the philosophical understanding of agency, responsibility, and the moral psychology of deliberation—core themes that remain central to contemporary debates in analytic philosophy of mind and action theory.

2. Literature Review

The philosophical literature on the relation between thought and action traverses' multiple disciplinary intersections—ranging from the philosophy of mind and action theory to moral psychology and metaethics. Central to these debates is the question of how reasons, as normative entities, exert their authority in the formation of intentional action. This paper intervenes in this discourse by analyzing how rational agency is not merely causally responsive to reasons but normatively structured through deliberation. To this end, it engages with four major philosophical figures whose contributions offer distinctive insights while also revealing significant gaps: Donald Davidson, Harry Frankfurt, Christine Korsgaard, and John McDowell.

Donald Davidson's seminal theory of *anomalous monism* posits that while every action is caused by mental events, including reasons, there are no strict psycho-physical laws linking reasons to actions (Davidson, 1970). In his influential paper "Actions, Reasons, and Causes" (1963), Davidson argues that the primary reason for which an agent acts can be identified as the cause of that action. This move was critical in naturalizing action explanation within the broader causal framework of science. However, his model has been subject to extensive critique for its inability to account for the *normative authority* of reasons. As critics have noted, Davidson's account leaves unanswered why certain considerations count as reasons *to the agent*, rather than merely as causal precursors. While his theory

 doi:10.20944/preprints202505.0176.v1

elegantly explains how reasons can function as causes, it does not explain how they acquire their force as *justifications*.

Harry Frankfurt introduces a more inward-facing model of agency through his concept of *higher-order volitions* (Frankfurt, 1971). For Frankfurt, what distinguishes persons from mere wantons is the ability to reflectively endorse or repudiate one's desires. This capacity for second-order volition imbues human agency with a distinctive reflexivity: one is not simply moved by desires but actively chooses which desires to act upon. However, Frankfurt's account, while normatively richer than Davidson's, still faces challenges. Notably, it offers little explanation of how higher-order volitions gain their normative authority. Why should the reflective endorsement of a desire count as sufficient for its rationality or ethical rightness? Frankfurt himself admits that identification is a psychological process that may or may not align with moral or rational norms (Frankfurt, 1988). Thus, while his model clarifies the structure of volitional hierarchy, it does not fully resolve the question of normativity.

Christine Korsgaard addresses this gap by offering a *constitutivist* account of normativity—one that grounds the authority of moral and practical norms in the structure of agency itself. According to Korsgaard, to act at all is to act under the idea of a law; that is, rational agents must regard their reasons as binding upon themselves (Korsgaard, 1996). The source of normativity, on this view, lies in the reflective nature of human beings, who cannot act without endorsing a principle as a reason. She writes, "The reflective structure of human consciousness…makes the question 'Is this reason really a reason?' inescapable" (Korsgaard, 2009, p. 2). Korsgaard's notion of *practical identity*—the self-conception that confers unity and coherence on one's actions—plays a central role here. For example, a person who identifies as a parent, citizen, or moral agent acts in ways that reflect those identities, thereby binding themselves to the reasons internal to those roles. Her account thus offers a powerful solution to the problem of normative authority, though it has been critiqued for its reliance on internal coherence without external criteria of correctness.

John McDowell advances this discussion by challenging the overly formalistic and instrumental models of rationality often found in analytic philosophy. His work on *second nature* (McDowell, 1996) repositions normativity as a product of ethical enculturation and lived experience. For McDowell, mature rational agents are not merely rule-followers but participants in a shared *form of life*, within which reasons appear as intelligible and binding. He resists the idea that reasons can be reduced to either causal forces or subjective endorsements. Instead, practical reasoning involves *seeing* certain features of a situation as reasons, a skill developed through habituation and ethical education. For instance, a seasoned judge may *see* the morally salient aspects of a case not through logical deduction but through cultivated ethical perception. McDowell thus offers a *perceptual model* of normativity, where deliberation is a matter of attunement rather than computation.

Despite their differences, all four thinkers converge on the idea that agency involves a complex relation to reasons—one that cannot be explained solely through causal mechanisms. Yet none offers a fully integrated model of how deliberation itself functions as a normative process that enables the transition from thought to action. This paper seeks to synthesize their contributions into a unified account that views deliberation as an activity governed by normativity at every level: from the internal coherence of volition (Frankfurt), to the self-constituting structure of agency (Korsgaard), to the responsiveness to ethical reasons shaped by second nature (McDowell), all undergirded by Davidson's insight that reasons play an explanatory role.

In doing so, this study contributes not merely to a descriptive model of agency but to a normative theory of deliberative action—one that can accommodate the complexity of human motivations, the demands of moral reflection, and the situated nature of ethical life. It argues that rational agency should be conceived not as the capacity to act for reasons in general, but as the capacity to deliberate normatively—that is, to recognize, endorse, and act upon reasons as authoritative within a reflective structure of self-governance.

3. Methodology

This paper adopts a conceptual and analytical methodology, grounded in the method of *reflective equilibrium*, to explore the normative underpinnings of deliberative agency. Rather than pursuing empirical verification or psychological modeling, the inquiry proceeds through critical engagement with existing philosophical accounts—especially those of Donald Davidson, Christine Korsgaard, Harry Frankfurt, John McDowell, R. Jay Wallace, and Nomy Arpaly—and aims to develop a constructive theoretical framework that both synthesizes and transcends their insights.

 doi:10.20944/preprints202505.0176.v1

The method of reflective equilibrium enables a dialectical balance between intuitions about particular cases of agency and more general principles about what it means to act for reasons. This involves an iterative process: refining our understanding of intentional action in light of conceptual tensions, counterexamples, and theoretical desiderata, while also modifying background commitments where necessary to achieve coherence.

The paper draws on thought experiments, drawn from both classical sources (e.g., akratic action, weakness of will) and contemporary discussions (e.g., moral motivation without conscious endorsement, cases of rational regret), in order to pressure-test theoretical models of deliberation. These examples help illuminate the difference between merely causally effective mental states and those that function as authoritative reasons for action.

Additionally, conceptual analysis plays a central role, particularly in clarifying contested notions such as "normativity," "commitment," "practical identity," and "deliberative authority." Rather than assuming shared definitions, the paper seeks to explicate the internal structure and logical interrelations of these concepts, showing how they underwrite a distinctively normative account of agency.

By maintaining a focus on the first-personal standpoint, this methodology avoids reductive explanations that abstract away from the deliberative perspective of the agent. In so doing, it preserves the philosophical significance of rational agency as an irreducibly normative phenomenon—one that requires philosophical analysis rather than empirical reduction or behavioral description. The result is a model of deliberation that aspires to be not only conceptually rigorous but also sensitive to the real demands placed on agents in conditions of evaluative complexity and moral responsibility.

3.1. The Structure of Deliberation

Deliberation, at its philosophical core, is not simply a process of choosing among alternatives based on instrumental utility or subjective desire. Rather, it is a normatively guided activity, one that presupposes a conception of the agent as a self-reflective bearer of reasons. In this sense, to deliberate is not only to consider what one wants, but to assess what one ought to do—what is justifiable within a space of reasons that one recognizes as authoritative.

As Christine Korsgaard argues, "to act is to choose, and to choose is to value" (Korsgaard, 1996, p. 93). But such valuing is not arbitrary—it emerges from an agent's practical identity, the self-conception that allows one to ask not only *what* to do but *why* to do it. Deliberation thus involves a kind of internal dialogue in which the agent occupies a normative standpoint, testing the coherence of various impulses, principles, and commitments. This structure requires that agents not only have desires or goals but also the capacity to evaluate and endorse certain reasons as binding.

Donald Davidson's influential theory of action (1980) famously posits that reasons can be causes, yet his model—anchored in *anomalous monism*—does not fully capture the normative dimension of why some reasons justify actions while others merely explain them. An agent may be moved to act by a desire, but what distinguishes deliberative action is the endorsement of a reason as one that *ought* to count. As R. Jay Wallace (1994) notes, the very idea of responsibility presupposes that agents act under the guidance of reasons they recognize as normatively salient.

This commitment to normative appraisal distinguishes mere decision-making from deliberation. For example, consider an individual deciding whether to tell a painful truth or to remain silent. From a merely consequentialist frame, one might weigh potential outcomes; but within the deliberative structure of rational agency, the agent must assess not just outcomes, but duties, relationships, and principles—such as honesty, care, or justice—that may override mere preference. The agent asks: *What would it mean for me, as someone who values truthfulness or respects others' autonomy, to act in this way?* In doing so, the agent does not merely consult desires, but interrogates their fit within a larger evaluative framework.

Harry Frankfurt's (1988) theory of second-order volitions enriches this model by showing how agents are not just pushed by desires but take stands on them. Yet, Frankfurt's framework lacks a sufficient account of how such commitments become normatively binding. Korsgaard fills this gap by arguing that "the normative force of a reason lies in its relation to the identity of the agent" (Korsgaard, 2009, p. 20). In deliberation, the agent is engaged in a kind of self-legislation, giving reasons that are not merely motivational but justificatory. This is a crucial distinction: one may be motivated by fear, but one cannot justify a cowardly act merely by citing fear. Justification appeals to reasons that withstand scrutiny under standards the agent endorses.

 doi:10.20944/preprints202505.0176.v1

John McDowell (1996), through his notion of *second nature*, deepens this understanding by emphasizing that mature practical reasoners are not merely rational calculators but have undergone a process of habituation into the moral and rational space of reasons. Deliberation, in this sense, is a practice situated in a normative tradition—it reflects not just personal preference but a socially and historically embedded sense of the good.

Finally, Nomy Arpaly (2003) challenges the assumption that only explicitly endorsed reasons can be normative. Her account of inverse akrasia—cases where agents do the right thing despite not endorsing the right reasons—highlights that sometimes-moral insight can operate beneath reflective awareness. Yet even these cases point to a deeper structure of rational agency: agents are attuned, even implicitly, to moral salience, and their responsiveness can still be normatively assessed.

In summary, deliberation is not a computational algorithm nor a raw negotiation of competing desires. It is a rationally structured activity through which agents determine what they ought to do by consulting a web of values, commitments, and identities. It presupposes not only the capacity for reflection but also the authority to bind oneself to reasons that survive such reflection. It is in this capacity—to deliberate, endorse, and act for reasons—that we find the distinctive normativity of rational agency.

3.2. The Normativity of Reasons

The normative dimension of reasons does not operate as an external imposition on the will, nor does it arise merely from sociocultural convention or psychological conditioning. Rather, normativity is constitutive of agency itself. To act as a rational agent is not merely to be moved by internal impulses or external stimuli, but to locate oneself within a space of justifications, where actions are guided by reasons that one can endorse as binding. This endorsement is not an optional feature of agency—it is, as Christine Korsgaard (1996) famously argues, what constitutes the self as a unified agent capable of owning its actions.

Korsgaard's constitutivist account insists that the authority of reason arises from the very structure of practical identity. "Acting on a reason," she writes, "is acting in a way that you can justify to yourself" (Korsgaard, 1996, p. 101). For her, normativity does not derive from some antecedent moral law or external command, but from the agent's need to unify herself through reflective endorsement of principles. This process of self-constitution entails a critical distance from immediate inclinations and a capacity to adopt maxims that express one's values—maxims that become binding precisely because they are necessary for maintaining one's practical unity.

This conception of normativity challenges both instrumentalist and naturalist accounts of action. Instrumentalist models, such as those stemming from Humean traditions, reduce reasons to mere tools for achieving antecedently given ends. But on Korsgaard's view, ends themselves require justification—and that justification can only be supplied by the agent's reflective endorsement of the principles they embody. In this way, reasons are not mere instruments but conditions of autonomous agency.

Moreover, the normativity of reasons is not simply descriptive, as in some contemporary naturalist approaches that attempt to explain norm-guided behavior through evolutionary or neurobiological functions. While such models may account for how human beings come to exhibit rational capacities, they often fall short of explaining why some reasons have binding authority. As John McDowell (1996) observes, the domain of reasons is "not law-governed like nature, but rule-governed in the space of concepts." For McDowell, normative responsiveness is part of a second nature—a cultivated attunement to reasons that arises from moral education and social practice, not mechanistic causality.

Consider an illustrative case: an individual faces the decision to betray a friend for personal gain. The agent feels the pull of self-interest but also experiences guilt at the thought of violating a bond of trust. What transforms this from mere inner conflict to a deliberative choice is the agent's capacity to reflectively assess the reasons involved. The betrayal may be advantageous, but it may also undermine the agent's identity as a loyal friend or ethical person. If the agent refrains from betrayal, not merely out of fear or habit but because loyalty is a value she endorses as part of who she is, then the reason has normative force—it binds because it emerges from her reflective self-understanding.

Korsgaard's claim that we act under the "idea of freedom"—that we must see ourselves as capable of acting on principle—also invites engagement with Kantian themes. Yet unlike Kant's formalism, Korsgaard grounds normativity in the agent's need for practical unity, thus avoiding the

 doi:10.20944/preprints202505.0176.v1

abstraction sometimes criticized in Kant. Her approach aligns more closely with the Aristotelian idea that ethical life involves habituating oneself into a rational form of life, but adds the modern insight that such habituation requires continual reflective endorsement.

In contrast, Nomy Arpaly (2003) complicates this picture by emphasizing that moral worth does not always align with reflective endorsement. Her account of inverse akrasia—cases where agents do the right thing despite not endorsing the right reasons—suggests that responsiveness to moral salience can be non-reflective yet still normatively significant. This does not negate Korsgaard's framework but refines it: normativity may sometimes exceed our reflective grasp, but it remains intimately tied to the agent's capacity to recognize and respond to reasons, even when such recognition is intuitive rather than deliberative.

Thus, the authority of reasons lies not in their external imposition, but in their internal necessity for maintaining the coherence of agency. As Korsgaard (2009) writes, "If there is no law you must give yourself, there is no self to give it." The agent is, in this view, both the legislator and the subject of normativity—a dual role that affirms autonomy while preserving the bindingness of reasons.

To act, then, is to submit oneself to a structure of norms one recognizes as governing the space of reasons. It is this submission—not to external compulsion, but to self-imposed law—that constitutes rational freedom. The normativity of reasons is the medium through which action becomes not just possible, but meaningful—the space where the agent no longer merely reacts but takes responsibility for what she does.

3.3. Practical Identity and Self-Commitment

Rational agency cannot be understood in abstraction from the agent's self-conception. Every choice is made not from a neutral vantage point but from within a practical identity—a self-understanding shaped by the roles, values, and aspirations to which an agent commits. As Christine Korsgaard (1996) articulates, "Your identity is constituted by what you regard as reasons to act." In this sense, practical identity is not merely descriptive (e.g., I am a teacher or a parent) but normative: it provides the framework within which certain reasons count as reasons for that agent.

The central claim of this section is that rational choice is intelligible only through the lens of self-commitment to particular roles and values. When an agent deliberates about whether to tell the truth, confront a friend, or resign from a job, their decisions are filtered through the normative lens of who they take themselves to be. To betray a friend, for example, is not just to violate a social rule but to fail to live up to a self-understanding—to act in a way that undermines one's integrity as a loyal person. Thus, practical identity is the medium through which reasons acquire their authority.

Harry Frankfurt's (1988) conception of second-order desires and volitions forms a critical precursor to this view. For Frankfurt, what distinguishes persons from mere wantons is the capacity to reflect on their desires and endorse some over others, thereby creating a hierarchy that structures the will. A person's identity, on this account, is not simply a matter of what one wants but what one wants to want—the desires one endorses as expressive of one's true self. However, Frankfurt leaves somewhat opaque the normative grounds on which such endorsement takes place. Why should an agent treat certain desires or commitments as authoritative?

Korsgaard (2009) addresses this lacuna by embedding Frankfurt's insight within a broader constitutivist framework: the act of self-constitution, of binding oneself to principles that structure one's agency, is what gives second-order volitions their normative force. To endorse a desire is to incorporate it into a practical identity—an identity that must be maintained through coherent, deliberative action. For Korsgaard, this is not a static process but a dynamic self-commitment: "To value anything at all, you must value your own humanity, and you must do so practically, by treating yourself as an end" (Korsgaard, 1996, p. 122).

An agent's practical identity therefore serves as both the source and the standard of normativity. It is what enables the agent to recognize some reasons as binding while dismissing others as irrelevant. Consider the case of a doctor who is offered a bribe to prioritize a patient. The temptation may appeal to self-interest, but if the agent's identity is shaped by professional ethics and a commitment to justice, then the bribe cannot be endorsed without fracturing the coherence of that identity. Her refusal is not merely a matter of preference but a reaffirmation of her normative self-conception.

This model also resonates with John McDowell's (1996) emphasis on the cultivation of second nature. McDowell argues that moral perception is not reducible to rule-following but is infused with a background of ethical training and habituation. Practical identity, on this view, is not imposed by

 doi:10.20944/preprints202505.0176.v1

external codes but formed through participation in forms of life that shape the agent's responsiveness to reasons. The deliberating self is not a detached rational calculator but an embedded agent, whose identity mediates how the normative force of reasons is experienced and internalized.

However, one might object that practical identities are often conflicted, unstable, or imposed. Social roles can be constraining, and an individual may experience conflicting demands between identities (e.g., as a parent vs. a professional). The model advanced here does not deny this. Rather, it emphasizes that deliberation itself is the means by which agents negotiate these tensions. Rational agency is the capacity to assess, revise, and endorse identities reflectively—what Frankfurt might call "taking responsibility for the self" and what Korsgaard calls "self-constitution." Indeed, it is precisely because identities are complex and sometimes in tension that the normative structure of deliberation becomes essential.

Moreover, drawing on contemporary contributions from R. Jay Wallace (2006), we can deepen this picture by emphasizing the moral expectations embedded within identity. Wallace argues that our commitments generate "normative expectations" that govern our interactions with others. In his account of responsibility, identity is not merely a personal project but a relational structure embedded in moral practices that hold us accountable. To act against a commitment is not simply to act out of character but to violate the normative expectations that others—and ourselves—have formed on the basis of our avowed identity.

In light of this, practical identity is not a passive inheritance but a reflexively sustained project. Through acts of deliberation, agents reaffirm, revise, or repudiate their roles and commitments. It is this process of self-binding that gives rational action its normative intelligibility. Deliberation is not merely a computation of outcomes but a dialogue with the self—a process in which reasons are weighed not only for their consequences but for their coherence with the agent's ongoing project of being a certain kind of person.

In sum, the normative authority of reasons arises not in abstraction from the self but in relation to the practical identity that constitutes the self. Rational agency is a process of self-commitment through reason—a process that binds the agent to her principles not as external laws, but as internal necessities of personhood.

3.4. From Thinking to Doing

The transition from deliberative thought to intentional action marks a critical threshold in the structure of rational agency. While much philosophical attention has been devoted to identifying the nature of reasons and their normative authority, less clarity exists about how agents move from reflectively endorsing a reason to enacting it. This section proposes that action is not a mere output of deliberation but a synthesis of reflective endorsement and embodied habituation, a point crucially informed by John McDowell's concept of *second nature*.

For McDowell (1996), the capacity to act for reasons cannot be reduced to rule-following or algorithmic decision-making. Rather, it is the result of a kind of moral and practical upbringing—a habituation into forms of responsiveness that make certain actions "make sense" from within a rational perspective. He writes, "Our capacities to recognize reasons are not grounded in bare receptivity, but in a second nature shaped by upbringing into reason-governed practices." This shaping is not merely causal; it constitutes the normative orientation necessary for moving from endorsement to action.

This transition—from reflective thought to enactment—can be understood as a process of *practical synthesis*, in which the agent's identity, background habits, and immediate circumstances converge to produce action. The deliberating agent does not merely ask "What ought I to do?" but integrates this question with "What am I doing?" in a temporally unfolding process of self-authorization. Action, then, is not a separate domain from thought, but its culmination. It is what happens when reasons are not only acknowledged but *inhabited*.

A paradigmatic case might be the decision to confront a colleague over an ethical violation. One may deliberate extensively, endorse the principle of integrity, and recognize the professional duty to act. However, the actual confrontation—speaking the words, initiating the moment—is not simply caused by reflection, nor can it be fully explained by the reasons themselves. What enables the agent to move from reflective endorsement to courageous enactment is the integration of rational conviction with embodied readiness. This readiness is not innate but cultivated—a result of prior actions, character formation, and the internalization of values through practice.

 doi:10.20944/preprints202505.0176.v1

Harry Frankfurt's (1988) account of volitional necessity further illuminates this point. He describes agents who, when acting on a deeply held commitment, experience their action as inevitable—not because they lack control, but because their identity and values leave no viable alternative. In such cases, action is not externally compelled but internally necessitated. The bridge from thought to doing, then, is formed not only by the logic of reasons but by the felt necessity of embodying them in a contextually appropriate form.

Still, this synthesis is fragile and contingent. Contemporary research in moral psychology and neuroscience has emphasized the dissonance that can arise between evaluative judgments and actual behavior (Greene & Haidt, 2002; Cushman, 2008). Such findings challenge any simplistic assumption that recognizing a reason will automatically produce corresponding action. However, these challenges do not refute the normative structure proposed here; rather, they underscore the difficulty and achievement involved in aligning thought with action. Failures to act on reasons can reflect not a breakdown of normativity but a deficiency in the agent's practical integration—an instability in identity, habituation, or situational courage.

This gap between thought and doing invites renewed attention to the role of practical imagination, a faculty often underexplored in accounts of agency. To act well, an agent must not only deliberate but envision the enactment of reasons, anticipate the emotional and social terrain, and preemptively reconcile the internal resistances that may arise. In this way, imagination becomes an enabling condition of rational agency, helping to bridge the abstract space of reasons with the concrete space of lived embodiment.

Moreover, as R. Jay Wallace (2006) and Nomy Arpaly (2003) argue, agents are often held morally responsible not merely for reflective rationality but for whether they are responsive to moral reasons in a reliable and authentic way. Arpaly, in particular, critiques the overemphasis on internal reflection and highlights non-deliberative responsiveness—actions that emerge from an internalization of moral sensitivity without discursive deliberation—as genuine expressions of moral agency. This supports the claim that reason-responsiveness must be cultivated, not merely acknowledged, and that such cultivation is what allows reflection to translate into action.

In sum, the movement from thinking to doing is not linear but dialectical and embodied. It involves the mutual interplay of reflective deliberation, habituated sensitivity, emotional preparedness, and situational responsiveness. Rational agency, then, is not reducible to cognition; it is a lived normativity, a practice of binding thought to the world through self-committed, identity-constitutive action.

3.5. Implications and Philosophical Payoff

The reconceptualization of deliberation as a normatively structured, identity-constituting process of rational agency yields significant implications across a range of contemporary philosophical domains. Most notably, it provides a fresh framework for rethinking autonomy, moral responsibility, and critiques of instrumental rationality, while also offering fertile ground for analyzing emerging challenges in AI ethics and cognitive science.

First, in the discourse on autonomy, this account departs from procedural or minimal conceptions that define autonomy merely in terms of non-interference or decision-making capacity. Instead, following Korsgaard (2009) and Frankfurt (1988), autonomy is here re-envisioned as the *capacity to bind oneself to reasons one reflectively endorses*, grounded in a deliberative relation to one's practical identity. This allows for a thicker, more substantive conception of the autonomous agent—not as a disengaged chooser, but as a being who stands in normative relation to their actions, through self-constitution. In this sense, true autonomy is a form of self-legislation, where deliberation is not just a psychological mechanism but a moral activity of shaping one's will.

Second, this view reshapes how we think about moral responsibility. The traditional model often assumes a direct causal connection between internal states (beliefs, desires, intentions) and outward behavior. But as R. Jay Wallace (2006) has argued, responsibility hinges not merely on control or awareness but on normative responsiveness—the agent's ability to recognize and act for the right reasons. By emphasizing the practical integration of identity, habituated sensitivity, and reflectively endorsed reasons, this account explains how responsibility is not exhausted by causal antecedents. Rather, it is earned through an ongoing practice of *reason-guided self-governance*, even when actions do not proceed from explicit deliberation.

 doi:10.20944/preprints202505.0176.v1

Third, this argument intervenes in long-standing critiques of instrumentalism, especially in debates stemming from Humean and Kantian traditions. Instrumental views—where reason is merely a tool for satisfying pre-existing desires—fail to account for how reasons can transform desires or override them altogether. McDowell's (1996) conception of second nature, as well as Korsgaard's constitutivism, provide resources for resisting this reduction. By foregrounding deliberation as normatively loaded, this paper supports the claim that agents are not merely calculators of outcomes but *participants in the space of reasons*—a space shaped by values, commitments, and forms of life. This helps explain phenomena like moral transformation, akrasia, and conscientious dissent, which instrumental theories struggle to accommodate.

Finally, the implications of this account extend into the philosophy of mind and artificial intelligence, where questions of synthetic agency and ethical alignment have become urgent. Contemporary AI systems may simulate deliberation—ranking options, predicting consequences, optimizing behavior—but they lack the normativity that constitutes human agency. They do not bind themselves to reasons, nor do they possess identities to which they are answerable. Even the most advanced AI lacks what Frankfurt calls "second-order volition" and what McDowell understands as "second nature." This raises a critical challenge: without the capacity to experience reasons as binding, can AI ever be said to act, rather than merely behave?

Such questions are not only theoretical but urgent, as AI increasingly mediates decision-making in healthcare, justice, warfare, and governance. By clarifying what it means to move from thought to action—*not just causally, but normatively*—this account offers a benchmark for distinguishing genuine agency from mere computation. In this light, any future claims about artificial moral agents must reckon with the irreducibility of normativity in action.

In sum, the philosophical payoff of this account is twofold. It deepens our understanding of human agency by grounding action in a normative, identity-based structure of deliberation. And it challenges prevailing assumptions in moral psychology, ethics, and AI by demonstrating that agency is not merely about having reasons but about being able to live them.

4. Conclusions

This paper has defended a normative account of rational agency that challenges reductive models which attempt to explain action through psychological or neurophysiological causation alone. By drawing upon and synthesizing insights from Davidson, Frankfurt, Korsgaard, McDowell, and more recent contributions by thinkers like Wallace and Arpaly, it has argued that deliberation is not a prelude to action but an essential expression of agency itself. It is within the act of deliberation—structured by reflective endorsement, practical identity, and habituated responsiveness—that the agent binds themselves to reasons they recognize as authoritative.

Rather than treating reasons as inert causes or preferences, this account frames them as normative commitments that arise from the agent's self-conception, their second-order volitions, and their embeddedness within ethical practices. The transition from thought to action thus becomes intelligible not as a mysterious leap or as an outcome of causal chains, but as the culmination of an agent's rational self-determination.

By foregrounding the constitutive role of normativity, this approach not only clarifies long-standing puzzles in the philosophy of action and moral psychology but also offers resources for contemporary debates in AI, ethics, and autonomy. It articulates a standard by which genuine agency must be evaluated—not merely in terms of behavior or outcomes, but in terms of the internal normative architecture that enables a being to ask: *What ought I to do?* and to answer not just with action, but with conviction.

Ultimately, this view restores deliberation to its rightful place at the heart of human agency: not as a mechanism of choice, but as the form through which freedom, identity, and ethical life are enacted. The capacity to act is inseparable from the capacity to reflect, to endorse, and to inhabit the reasons that make action meaningful.

5. Future Directions

The normative account of deliberative agency defended in this paper opens several promising avenues for future philosophical investigation. First, it invites a more sustained engagement with applied ethics, particularly in domains where rational deliberation intersects with moral education, legal responsibility, and public reasoning. If agency is constituted by reflective endorsement and

 doi:10.20944/preprints202505.0176.v1

practical identity, then fostering agency—whether in educational institutions or therapeutic practices—requires more than behavioral regulation; it demands the cultivation of reflective capacities and the moral imagination.

Second, this framework suggests a new orientation for debates in cognitive science and AI ethics, especially regarding the simulation or realization of rational agency in artificial systems. As discussions about machine autonomy gain momentum, the question becomes not simply whether AI can act, but whether it can deliberate in norm-sensitive ways. Can an artificial agent exhibit genuine responsiveness to reasons—not merely as instrumental inputs, but as normatively binding considerations arising from a self-governing structure? This reframes the challenge of AI alignment: not as programming rules, but as modeling forms of normativity that mirror human deliberative agency.

Third, the paper opens space for further analysis of autonomy in hybrid agents, including neurodivergent individuals or human-technology interfaces, where the boundaries of self-constitution, reason-responsiveness, and moral responsibility are increasingly contested. The conceptual toolkit developed here may thus offer fresh insights into neuroethical, posthuman, and technologically mediated forms of practical reasoning.

In sum, the proposed view not only enriches theoretical accounts of action but also bears on urgent interdisciplinary conversations—calling for a deeper integration of normative philosophy, moral psychology, and the design of deliberative systems, both human and artificial.

References

1. Anscombe, G. E. M. *Intention*. Harvard University Press, 1957.
2. Arpaly, Nomy. *Unprincipled Virtue: An Inquiry into Moral Agency*. Oxford University Press, 2003.
3. Bratman, Michael E. *Intention, Plans, and Practical Reason*. Harvard University Press, 1987.
4. Churchland, Patricia. *Braintrust: What Neuroscience Tells Us about Morality*. Princeton University Press, 2011.
5. Davidson, Donald. 'Actions, Reasons, and Causes'. *The Journal of Philosophy* 60, no. 23 (1963): 685–700.
6. Frankfurt, Harry G. 'Freedom of the Will and the Concept of a Person'. *The Journal of Philosophy* 68, no. 1 (1971): 5–20.
7. ——. *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press, 1988.
8. Greene, Joshua. *Moral Tribes: Emotion, Reason, and the Gap Between Us and Them*. Penguin, 2013.
9. Korsgaard, Christine M. *The Sources of Normativity*. Edited by Onora O'Neill. Cambridge University Press, 1996.
10. ——. *Self-Constitution: Agency, Identity, and Integrity*. Oxford University Press, 2009.
11. McDowell, John. 'Virtue and Reason'. *The Monist* 62, no. 3 (1979): 331–350.
12. ——. *Mind and World*. Harvard University Press, 1994.
13. Scanlon, T. M. *What We Owe to Each Other*. Harvard University Press, 1998.
14. Wallace, R. Jay. *Responsibility and the Moral Sentiments*. Harvard University Press, 1994.
15. ——. *The View from Here: On Affirmation, Attachment, and the Limits of Regret*. Oxford University Press, 2013.