Scientific Research Publishing

Open Journal of Philosophy, 2025, 15(3), 551-566
https://www.scirp.org/journal/ojpp
ISSN Online: 2163-9442
ISSN Print: 2163-9434

The Participant Stance and the View from Nowhere: Reconciling Strawson's Reactive Attitudes with Agent-Neutral Morality

Leo Lin

Independent Researcher, Baltimore, USA
Email: Leo760Lin@gmail.com

How to cite this paper: Lin, L. (2025). The Participant Stance and the View from Nowhere: Reconciling Strawson's Reactive Attitudes with Agent-Neutral Morality. *Open Journal of Philosophy, 15,* 551-566.
https://doi.org/10.4236/ojpp.2025.153033

Received: June 2, 2025
Accepted: July 28, 2025
Published: July 31, 2025

Abstract

This paper explores the apparent tension between P. F. Strawson's reactive attitude framework—rooted in the interpersonal, participant stance—and the agent-neutral ideals found in utilitarian and deontological moral theories. While Strawson emphasizes emotionally embedded responses like resentment and gratitude as essential to moral responsibility, agent-neutral theories demand impartiality and universal justification. The paper critically examines three central points of friction: the partiality of reactive attitudes, the naturalistic insulation of Strawson's framework from external rational critique, and the potential for moral correction based on impartial principles. It argues that reconciliation is possible through reflective equilibrium, viewing reactive attitudes as epistemically significant but revisable, and by delineating a division of moral labor: Strawson's account captures the practice of moral responsibility, while agent-neutral theories provide normative criteria. Ultimately, the paper defends a dynamic, interpretive compatibility that honors both the engaged, affective reality of moral life and the aspirational impartiality of ethical theory.

Keywords

Free Will, P. F. Strawson, Reactive Attitudes, Agent-Neutral Morality, Moral Responsibility, Participant Stance, Determinism, Compatibilism, Resentment, Reflective Equilibrium, Utilitarianism, Kantian Ethics, Naturalism, Justification, Moral Psychology

1. Introduction

A reactive attitude is said to be "fitting" when directed, in the right degree, to the agent morally responsible for the harm that provoked it. Morality and accountability are integral to social relations because individuals hold one another account-

able. When we feel offended or annoyed at what we think is an injustice done to us, we get angry or irritated, rewarded when kind or nice things are done to us, and angry or annoyed when people we relate with are mistreated. They are profound because they form the basis of our emotional existence and moral perceptions. In a similar vein, these "reactive attitudes" are placed by P. F. Strawson at the core of his theory of moral responsibility. He suggested that our patterns of praising, blaming, punishing, or forgiving are strategic attitudes directed towards actions based on the quality of will of the actions, goodwill, ill goodwill, or no goodwill (Strawson, 1962: pp. 6-7). Strawson offers this framework as an inherent part of our interpersonal engagements, possibly a permanent element of human interaction immune to change due to new and different metaphysics of determinism.[1]

This contrasts with an agent-neutral approach to morality. Utilitarianism or certain forms of Kantian deontology call for the individual to evaluate the rightness or wrongness, the goodness or badness of the states of affairs from an impartial point of view, irrespective of the context or the perspectives of the actors involved in the act (Smart, 1961).[2] These theories look for principles that could be used to enshrine moral assessment that is not bound by individual feelings and interactions. Instead, they aspire to a purity of perspective detached from the earthly plane and establish a definite moral imperative.[3] This apparent tension provokes an important question: Is Strawson's reactive attitude framework, grounded in the participant standpoint and interpersonal commitment, aligned with the impersonal, universal rationality of agent-neutral ethics? Is it possible to reconcile Strawson's significance to our attitudes and practices with theories that pass judgment on actions concerning some objective standards? Despite prima facie conflicts that might be seen, especially in certain aspects of reactive attitudes conflicts and questions about the justification of the framework as partialism, closer analysis shows that Strawson's account can be squared with agent-neutral morality. Again, however, reactive attitudes might best fit into a picture of moral

[1]Strawson offers this framework as an inherent part of our interpersonal engagements, possibly a permanent element of human interaction immune to change due to new and different metaphysics of determinism. This implies that our responsibility practices are not theoretical add-ons but expressions of how we relate to one another at a fundamental level. The force of this view lies in its defiance of revision: even if determinism were proven true, our practices of praise, blame, and moral expectation would remain, not out of ignorance, but because they are constitutive of what it means to live as moral beings among others.

[2]Utilitarianism or certain forms of Kantian deontology call for the individual to evaluate the rightness or wrongness, the goodness or badness of the states of affairs from an impartial point of view, irrespective of the context or the perspectives of the actors involved in the act. This can feel strangely detached—as if moral judgment happens outside of human relationships, stripped of the emotional and social fabric that gives those judgments weight. See J. J. C. Smart (1961) for a classic articulation of utilitarian reasoning, and Immanuel Kant's *Groundwork for the Metaphysics of Morals* for the deontological emphasis on universal moral law and rational duty.

[3]"Purity of perspective" refers to a moral viewpoint that seeks to transcend personal ties, emotions, and situational context, aiming instead for an impartial and rational stance. The phrase "definite moral imperative" evokes the kind of clear, universal moral rules emphasized in Kantian ethics and certain forms of utilitarianism, where the rightness or wrongness of actions is determined independently of individual perspective.

understanding that encompasses broader perspectives and is perhaps best considered in terms of something akin to reflective equilibrium biting the bullet on the compatibility of engaged responsiveness and impartial cognition, even if there is no simple answer to the question how.[4] Expunging these sectarian ontological speculations, especially regarding determinism, makes it possible to observe a more explicit focus on this core problem of the participant approach to value and agent neutrality.

2. Strawson's Reactive Attitude Framework: The Participant Stance

2.1. The Nature and Function of Reactive Attitudes

Strawson's analysis begins by highlighting the profound importance we attach to the attitudes and intentions others display toward us. Our emotional and interpersonal lives are structured around reactions to the consequences of actions and the quality of will they express (Strawson, 1962: pp. 6-7).[5] As Strawson illustrates, the pain of a hand being trodden upon may be identical whether it occurs accidentally or from contemptuous disregard. However, our reaction—resentment—arises primarily in the latter case, registering the perceived ill will or indifference (Strawson, 1962: p. 7).

2.2. Excuses and the Preservation of Participant Status

He identifies a range of "reactive attitudes" stemming from this interpersonal engagement. Personal reactive attitudes, like resentment and gratitude, arise from how others' wills are directed towards us. Impersonal or vicarious analogs, such as indignation and moral admiration, arise from how others' wills are directed toward third parties. Finally, self-reactive attitudes, like guilt and remorse, involve directing these reactions toward our manifested will (Strawson, 1962: p. 7). Common to all is their nature as responses within the network of interpersonal relationships, predicated on viewing others (and ourselves) as participants in this shared moral space. Smart represents the "optimist" position described by Strawson. He argues against libertarian free will and offers a consequentialist/utilitarian justification for practices of praise and blame (Smart, 1961: pp. 302-306). He distinguishes mere "grading" (dispraise) from blame, suggesting that rational blame is essentially grading plus an assessment of responsibility, understood pragmatically in terms of whether the agent's behavior is modifiable by social influences like reward and punishment. Smart seeks to replace what he sees as confused met-

[4]This gesture toward reflective equilibrium points to an attempt at reconciliation, but the paper arguably oversimplifies the challenge by framing the tension between reactive attitudes and impartial cognition as merely a matter of balance. The deeper difficulty lies in whether these fundamentally different moral perspectives—emotionally grounded engagement versus detached evaluation—can be integrated without undermining the core commitments of either. The appeal to equilibrium risks glossing over the structural incommensurability some theorists (e.g., Galen Strawson) insist upon.

[5]"Quality of will" refers to the attitudes, intentions, and degree of regard an agent shows toward others in their actions. It is central to Strawson's account of moral responsibility, where our emotional responses hinge not merely on outcomes but on the meaning behind the agent's conduct.

aphysical notions underpinning ordinary blame with this pragmatic justification (Smart, 1961: pp. 292, 294-295).[6] This contrasts sharply with Strawson, who considers the reactive attitudes and the interpersonal demands they express as fundamental and not reducible to, or justifiable solely by, their utility in regulating behavior. Smart prioritizes agent-neutral consequences, while Strawson grounds responsibility in agent-relative interpersonal reactions.

Strawson contrasts these "participant attitudes" with the "objective attitude." To adopt the objective attitude towards another is to cease viewing them as a fellow participant and instead see them "as an object of social policy; as a subject for what, in a wide range of sense, might be called treatment; as something certainly to be taken account, perhaps precautionary account, of; to be managed or handled or cured or trained." (Strawson, 1962: p. 8). While the objective attitude might involve emotions like pity or fear, it excludes characteristically interpersonal reactions like resentment, gratitude, forgiveness, or reciprocal adult love (Strawson, 1962: pp. 6-7).

Strawson identifies two primary ways our reactive attitudes are modified or suspended. The first involves excuses. Here, we learn information showing that the action did not manifest the quality of will we initially perceived (e.g., "He didn't mean to," "He hadn't realized," "He couldn't help it because he was pushed"). These considerations modify or suspend a particular reactive attitude (resentment might give way to forgiveness or be nullified). Still, they do not remove the agent from interpersonal relationships or make us view them objectively (Strawson, 1962: pp. 6-7).[7] The demand for goodwill remains in place; we merely learned it wasn't breached on this occasion or in the way we thought (Strawson, 1962: pp. 12-13).[8]

2.3. Exemptions and the Objective Attitude

The second involves exemptions. Here, we suspend the reactive attitudes generally towards an individual, either temporarily ("He wasn't himself," "He has been under very great strain") or more permanently ("He's only a child," "He's a hopelessly schizophrenic," "His mind has been systematically perverted"). In these cases, we view the agent as psychologically abnormal or morally undeveloped, as someone "excluded from ordinary adult human relationships" or "incapacitated in some or all respects for ordinary interpersonal relationships." (Strawson, 1962: pp. 10). Adopting the objective attitude here means lifting the usual demands and

[6]J. J. C. Smart argues that traditional metaphysical accounts of free will and blame are incoherent and should be replaced with a forward-looking, utilitarian rationale. Rather than blame resting on desert, it should be justified in terms of its social usefulness in shaping behavior.

[7]This distinction highlights Strawson's view that excuses operate within the reactive framework rather than undermining it—they adjust our moral responses without dissolving the interpersonal stance that grounds them.

[8]See P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 12-13, and Pamela Hieronymi, *Freedom, Resentment, and the Metaphysics of Morals* (Princeton, NJ: Princeton University Press, 2020), 9, for the view that our expectations of goodwill are not necessarily withdrawn when we revise our interpretation of another's actions—only our understanding of whether a breach occurred.

expectations associated with participant status (Hieronymi, 2020: pp. 12-13).

Strawson famously introduces a third way the objective attitude might arise: we can sometimes choose to adopt it as a "resource," perhaps as a refuge from the strains of involvement or as an aid to policy (Strawson, 1962: p. 13). This "use of the resource" becomes central to his later argument. Still, his core picture presents the framework of reactive attitudes and the associated demands constituting ordinary interpersonal life.

2.4. Naturalism and the Insulation from External Justification

This framework is presented with a strong sense of "naturalism." Strawson argues that our "commitment to participation in ordinary inter-personal relationships is… too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction (like determinism) might so change our world." (Strawson, 1962: p. 11). He suggests the entire framework "neither calls for nor permits, an external 'rational' justification." (Strawson, 1962: pp. 22). Like our commitment to inductive reasoning, it is "original, natural, non-rational (not irrational), in no way something we could choose or give up." (Strawson, 1962: p. 22; Hieronymi, 2020: p. 55).[9] It is "given with the fact of human society." (Strawson, 1962: p. 22). This naturalistic stance insulates the framework from certain kinds of external critique, including those potentially stemming from agent-neutral moral theories.

3. Agent-Neutral Morality: The View from Nowhere

3.1. The Agent-Neutral View in Theory

Agent-neutral moral theories stand in apparent contrast to the participant-focused framework Strawson describes. While Strawson emphasizes our reactions to the quality of will manifested within relationships, agent-neutral theories aim to establish moral criteria independent of any particular agent's perspective, desires, or interpersonal commitments.

The core idea of agent-neutrality is that the fundamental reasons for action, or the principles determining rightness and wrongness, are universal in their application (Scanlon, 1986: pp. 152-53; Nagel, 1986: pp. 152-154; Parfit, 1984).[10] What makes an action right or wrong, or a state of affairs good or bad, is determined by factors that do not inherently refer to the specific agent performing the action or the particular relationships they stand in. Dennett (1978) offers a compatibilist

[9]P. F. Strawson, "Freedom and Resentment," *Proceedings of the British Academy* 48 (1962): 22, as discussed in Pamela Hieronymi, *Freedom, Resentment, and the Metaphysics of Morals* (Princeton, NJ: Princeton University Press, 2020), 55. Strawson likens our commitment to moral attitudes—such as holding others responsible—to our reliance on inductive reasoning: a fundamental, non-rational stance that is not chosen but simply part of our human form of life.

[10]On the agent-neutral vs. agent-relative distinction, see Thomas Nagel, *The View from Nowhere* (Oxford University Press, 1986), 152-54, where agent-neutral reasons are defined as those that "can be given by anyone." See also Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984), for a broader framework in which agent-neutral reasons are central to utilitarian ethics.

account that, like Strawson's, seeks to insulate responsibility from the threat of determinism or mechanism (Dennett, 1978: pp. 282-83; Wolf, 1990: pp. 2-3).[11] However, instead of focusing on affective reactive attitudes, Dennett introduces the concept of explanatory "stances" (physical, design, intentional). Responsibility, for Dennett, is primarily located within the intentional stance, where we attribute beliefs, desires, and rationality to agents (Dennett, 1978: pp. 296-97). Excuses arise when this stance becomes inappropriate (e.g., due to manipulation, severe irrationality, or a lower-level stance provides a better explanation). While Strawson's "participant attitude" aligns closely with Dennett's "intentional stance," Dennett's framework offers a more cognitive and predictive rationale for responsibility practices, contrasting with Strawson's emphasis on the inherent nature of our interpersonal emotional responses and commitments (Hieronymi, 2020: pp. 56-56).[12] This objectivity, while valuable for agent-neutral theorizing, risks downplaying the moral and interpersonal depth of attitudes like resentment and forgiveness that Strawson sees as morally constitutive. Dennett's account thus points toward reconciliation, but only by shifting away from the emotionally embedded structure that makes Strawson's framework distinctive (Scanlon, 2008: pp. 122-147).[13]

3.2. Utilitarian and Kantian Models of Impartial Reasoning

Classic utilitarianism provides a clear example. The principle of utility dictates that the right action is the one that maximizes overall happiness or well-being and is impartially considered. The reason any agent has to act is grounded in its contribution to this aggregate good, regardless of whose good it is or what relationship the agent has with those affected. The value being promoted (overall happiness) is agent-neutral, and the reason derived (maximize that value) applies equally to all agents capable of influencing the outcome (Parfit, 1984: pp. 3-5).[14]

Specific forms of deontology also exhibit agent-neutrality. Kant's Categorical Imperative, for instance, demands that one act only according to maxims that one can will to become universal laws. The test is impartial and universal; the resulting duties (such as., not to lie or make false promises) apply to all rational agents ir-

[11]Dennett's approach belongs to a broader tradition of functionalist and cognitive compatibilism. For a contrasting account grounded more explicitly in emotional and interpersonal dynamics, see Susan Wolf, *Freedom Within Reason* (Oxford University Press, 1990), esp. ch. 2-3, where she defends a view of moral responsibility that integrates both reason-responsiveness and moral competence.

[12]While Dennett foregrounds cognitive and predictive utility, it may be an overstatement to read Strawson as entirely affective. As Pamela Hieronymi argues, Strawson's reactive attitudes are deeply tied to evaluative judgments, not just spontaneous emotions—suggesting a more integrated account than the contrast might imply.

[13]This divergence reflects a deeper philosophical question: can responsibility practices be both normatively justified and psychologically grounded? For an effort to bridge emotional and rationalist models of blame, see T. M. Scanlon, "Blame," in *Moral Dimensions: Permissibility, Meaning, Blame* (Harvard University Press, 2008), 122-147.

[14]This is a classic articulation of agent-neutrality: the reason for action is not indexed to any particular agent's perspective or relationship. See Derek Parfit, *Reasons and Persons* (Oxford University Press, 1984), 3-5, for an influential distinction between agent-neutral and agent-relative reasons.

respective of their inclinations or relationships.[15] The moral law itself is presented as an agent-neutral requirement binding on all.

The contrast with Strawson's framework seems stark. While Strawson acknowledges "impersonal" reactive attitudes like indignation, which respond to harms or benefits to others, even these are framed as arising within the participant stance–they are vicarious analogs of personal reactions felt by one participant concerning the interaction between two others (Strawson, 1962: p. 7). Agent-neutral theories, however, seem to demand a perspective outside this web of interpersonal reactions, evaluating actions against an impartial standard. Utilitarianism, for example, might require us to perform an action that harms a loved one if it produces a greater net benefit overall, potentially overriding the reactive attitudes (like loyalty or localized resentment) that Strawson's framework highlights. The demands of impartiality seem capable of conflicting with the partialities inherent in the participant's stance (Strawson, 1962: pp. 23-25).[16]

4. Locating the Tension: Partiality, Justification, and Correction

4.1. Partiality and the Participant Stance

Juxtaposing P. F. Strawson's reactive attitude framework with the principles characteristic of agent-neutral moral theories brings into sharp focus several points of significant potential tension, challenging the notion of their easy coexistence. These tensions revolve primarily around the contrasting perspectives they adopt—the engaged, participant stance versus the impartial, universal viewpoint—and the differing implications for moral justification and the possibility of critique (Strawson, 1962: pp. 6-7).[17] Understanding these friction points is crucial for evaluating whether these distinct approaches to morality can be reconciled.

The most immediate and intuitive tension stems from the inherent partiality embedded within many of the reactive attitudes Strawson places at the heart of moral responsibility. Consider the personal reactive attitudes like resentment and gratitude. Strawson describes that resentment typically arises in response to an injury or disrespect perceived as directed towards oneself, while gratitude often answers benefits conferred upon oneself (Strawson, 1962: pp. 8-9). These reactions are fundamentally perspectival, rooted in the individual's position within an interpersonal interaction. Agent-neutral theories, however, foundationally de-

[15]While Kantian duties are often categorized as agent-neutral because they apply universally, it's worth noting that the reasoning still originates from the agent's own rational will. Thus, even in its universality, the moral law is framed from within the perspective of the agent as a legislator of universal law, rather than from a fully external, observational standpoint.

[16]This tension between impartial moral demands and emotionally grounded interpersonal responses echoes Strawson's broader claim that such attitudes are not easily replaced by purely objective or impersonal stances. He emphasizes that our interpersonal commitments are deeply human, and abandoning them in favor of a wholly impartial perspective would constitute a kind of moral solipsism.

[17]Strawson characterizes the participant stance as one in which we respond with attitudes like resentment, gratitude, and forgiveness, arising naturally within interpersonal relations and informed by expectations of goodwill.

mand impartiality. From a classic utilitarian standpoint, for instance, the moral value of an action is determined by its contribution to the overall good, considered without special weight given to the agent's welfare or the welfare of those close to them. This impartial calculus directly conflicts with the self-focused nature of personal resentment. A utilitarian might conclude that an action causing me significant personal distress and thus provoking my (from a Strawsonian perspective, perhaps justified) resentment is morally right if it maximizes aggregate well-being (Scanlon, 1986: pp. 152-53).[18] In such a scenario, the agent-neutral principle appears to override or dismiss the validity of the partial reactive attitude. The question thus arises: can a moral framework that gives such central importance to these inherently partial, perspectival attitudes truly be consistent with a moral ideal demanding that we precisely transcend such limited perspectives in our ethical judgments? Even Strawson's "impersonal" or vicarious reactive attitudes, such as moral indignation felt on behalf of a third party, remain reactions within the specific web of human relationships and shared expectations (Strawson, 1962: pp. 15-16).[19] As such, they still reflect culturally specific norms or the particular contours of an existing social framework rather than necessarily aligning with the universal, impartial principles sought by agent-neutral theories. The very foundation of Strawson's approach in the participant stance seems prima facie opposed to the "view from nowhere" characteristic of agent-neutrality.

4.2. Justification and Strawson's Naturalist Insulation

This raises a further complexity in the Strawson-agent-neutral tension: while reactive attitudes are perspectival, they are not always idiosyncratic. An individual may feel resentment or indignation and still question whether the feeling is appropriate (Strawson, 1962: p. 6; Hieronymi, 2020: pp. 55-56).[20] This meta-level reflection suggests a possible evaluative framework external to the attitude itself. Just as one might ask whether a color experience reflects a stable property of the object under standard conditions, we can ask whether our reactive attitudes align with shared moral expectations or common human responses. The analogy to color perception is instructive here: we define "green" as what normal observers see under normal conditions, yet it remains a fact of the world that a traffic light's bottom signal is green (Hieronymi, 2020: pp. 55-56).[21] Similarly, while resentment is subject-dependent, we might still say it is objectively fitting or unfitting depending on whether it responds to genuine ill will or unjust treatment. This capacity

[18]This reflects the agent-neutral structure of utilitarianism, where reasons for action derive from states of affairs and apply universally, regardless of the agent's identity or perspective.

[19]Even Strawson's so-called "impersonal" reactive attitudes, such as moral indignation, depend on a shared moral culture and presuppose the standing of individuals within a moral community.

[20]Strawson recognizes that our emotional responses can be subject to moral assessment, distinguishing between having a reaction and judging it to be appropriate. Hieronymi develops this further, arguing that reactive attitudes admit of normative evaluation, not just psychological description.

[21]Hieronymi uses this analogy to argue that, like secondary qualities, reactive attitudes can be both subject-relative and truth-apt: they depend on human sensibility but are still open to objective assessment.

for critical reflection within Strawson's framework—evaluating the fittingness of an emotional response—opens a potential bridge to agent-neutral standards (Strawson, 1962: pp. 6-7; Hieronymi, 2020: p. 68).[22] It implies that reactive attitudes, though born from participant interactions, can be measured against shared norms, making them more compatible with the universalist aspirations of agent-neutral theories.

A second profound tension concerns the justification and foundational status of the respective frameworks. Strawson argues forcefully, invoking a form of naturalism, that the general framework of reactive attitudes—our deep-seated "commitment to participation in ordinary inter-personal relationships"—is something "given with the fact of human society." (Strawson, 1962: pp. 22). In this way, Strawson roots morality in actual human practices—how we respond to others and hold them accountable—rather than starting from universal moral precepts or rules. This contrasts with agent-neutral theories, which construct morality from abstract principles and apply them impartially, often without regard for how moral responsibility is enacted in real relationships. He suggests it is an "original, natural, non-rational" commitment, not something we choose or could give up, and consequently, it "neither calls for, nor permits, an external 'rational' justification." (Strawson, 1962: pp. 22).[23] This naturalistic defense aims to insulate the core practice of holding one another responsible from external critique based on abstract metaphysical or potentially even moral theories. Agent-neutral theories, conversely, typically aspire to provide precisely such a rational justification for their fundamental principles. Whether grounded in the calculation of utility, the demands of rational consistency as in Kantian ethics, or the outcome of a hypothetical contractual agreement, these theories seek to establish their core tenets through reasoned argument accessible from an impartial standpoint (Scanlon, 1982: pp. 103-128).[24] This raises a critical compatibility question: If Strawson's reactive attitude framework is beyond the reach of external rational justification, as his naturalism suggests, how can it mesh with theories whose methodology involves providing such justification for fundamental moral requirements? Does Strawson's claim that the framework is "given" preclude the possibility that an agent-neutral theory, should its rationally justified principles conflict with the deliverances of our reactive attitudes, could provide legitimate grounds for fundamentally criticizing or overriding the entire framework? If the framework is foundational in Strawson's sense, immune to external rational demands, it seems chal-

[22]This possibility complicates the apparent opposition between Strawson's participant stance and agent-neutral theories. If reactive attitudes can be assessed against intersubjective norms, then their justification may not be wholly agent-relative—a point that supports efforts to reconcile moral particularism with broader normative frameworks.

[23]Strawson argues that our commitment to moral responsibility is not a product of rational deliberation, but a natural human stance—one that lies outside the domain of justification and is not subject to voluntary revision.

[24]T. M. Scanlon develops a model of morality grounded in principles that no one could reasonably reject—a structure explicitly designed to justify moral claims from an impartial, agent-neutral standpoint.

lenging to integrate it smoothly with theories that derive their authority precisely from such demands. Hieronymi highlights Strawson's apparent view that the very question of external rational justification for the framework is "unreal," further emphasizing the potential justificatory gulf between the approaches (Hieronymi, 2020: pp. 59, 68).[25]

4.3. Correctability, Norm Critique, and the Role of Dennett and Frankfurt

Finally, tension arises concerning the possibility of moral correction and revision. Agent-neutral theories inherently provide an external standpoint from which our existing practices, norms, and attitudes can be critically evaluated (Scanlon, 1986: pp. 152-153).[26] Utilitarianism can condemn widely accepted social practices if they fail to maximize overall well-being. Frankfurt's (1971) seminal paper focuses on the structure of the will required for personhood and free will, distinguishing persons by their capacity for second-order volitions (desires about which first-order desires move them to act) (Frankfurt, 1971: pp. 6-8).[27] Frankfurt's analysis deeply informs the kind of agency towards whom such attitudes might be appropriate. His concept of identifying with one's desires illuminates the "quality of will" central to Strawson's account. Furthermore, Frankfurt's argument that moral responsibility does not require alternative possibilities (famous "Frankfurt cases" implied and illustrated here by the willing vs. unwilling addict) provides strong support for compatibilist positions like Strawson's, which similarly sideline the metaphysical debate about determinism's impact on choices (Strawson, 1962: pp. 6-7; Frankfurt, 1971: pp. 13-14).[28] However, Frankfurt concludes by analyzing the agent's internal structure rather than Strawson's focus on interpersonal practice. This inward focus offers conceptual tools for grounding responsibility in self-governance, making Frankfurt's theory potentially more agent-neutral in spirit. Yet it also distances his model from the social, affective engagement Strawson considers essential, suggesting that the reconciliation comes with a shift in emphasis—from relationships to internal coherence (Strawson, 1962: pp. 15-16; Frankfurt, 1971: pp. 15-16).[29]

[25]Hieronymi underscores Strawson's view that seeking an external rational justification for our moral responsibility practices is a misguided endeavor—what he sees as an "unreal" question—highlighting the deep conceptual divide between reactive and more detached, justificatory approaches.

[26]This evaluative structure is central to agent-neutral theories: they are defined not just by whom moral reasons apply to, but by their capacity to offer reasons independent of any particular perspective or practice.

[27]Frankfurt argues that persons differ from non-persons in their capacity for second-order volitions—desires concerning which desires should move them to act—thereby grounding moral responsibility in a hierarchical model of agency.

[28]Frankfurt's claim that moral responsibility does not depend on access to alternative possibilities undermines the metaphysical objection often leveled at compatibilism—supporting Strawson's pragmatic focus on responsibility as a lived human practice.

[29]Frankfurt's account shifts the basis of responsibility from shared social norms and expectations to an agent's internal structure of volitional alignment. This inward turn stands in contrast to Strawson's view, where reactive attitudes are rooted in interpersonal life and emotional responsiveness.

Kantianism can critique societal norms that cannot be universalized without contradiction. This critical function implies that our reactive attitudes are potentially subject to correction based on these overarching, impartial moral principles. For example, we might resent someone whose actions, while harmful to us, were morally required by an agent-neutral principle (perhaps causing lesser harm to prevent a greater catastrophe). In such a case, the agent-neutral theory would suggest that the resentment, however naturally felt, is morally inappropriate or misplaced, not because of an internal excuse or exemption of the kind Strawson describes (like accident or incapacity), but because an external, impartial standard justified the action itself. Does Strawson's framework accommodate this kind of fundamental, externally motivated correction? He explicitly states there is "endless room for modification, redirection, criticism, and justification" within the framework, driven perhaps by internal pressures like the demand for consistency or a deeper understanding of the relationships involved (Strawson, 1962: p. 22; Hieronymi, 2020: p. 80).[30] However, this internal flexibility may not extend to critique from external, agent-neutral standards. Strawson's emphasis on the natural, deeply rooted commitment to the existing framework seems to resist the idea that it could be fundamentally challenged or reshaped by appeal to such external criteria; his naturalism suggests the framework itself is too fundamental to be overturned simply because it conflicts with an abstract principle (Strawson, 1962: pp. 14-15). T. M. Scanlon's distinction between appraisals based on the "Quality of Will" (akin to Strawson's focus) and those based on the "Value of Choice" (more outcome-focused and potentially linkable to agent-neutral criteria) further suggests that different kinds of moral considerations might operate according to distinct justificatory logics, hinting at a potential compartmentalization rather than a seamless integration of these frameworks (Scanlon, 1986: pp. 183-184).

These tensions concerning partiality, justification, and the scope of moral correction indicate that asserting a straightforward consistency between Strawson's reactive attitude framework and agent-neutral morality is problematic. The participant stance, emphasizing engaged, perspectival reactions and its naturalistic grounding, appears prima facie to sit uneasily with the impartial, reason-giving aspirations characteristic of prominent agent-neutral ethical theories. Any successful reconciliation must directly address these points of friction.

5. Paths to Reconciliation: Reflective Equilibrium and Division of Labor

5.1. Reflective Equilibrium and Attitudinal Fit

Despite the significant points of tension identified between P. F. Strawson's reactive attitude framework and the principles of agent-neutral morality, several interpretive strategies and theoretical lines of argument suggest that compatibility, or at least a functional coexistence, might be achievable. These approaches seek to

[30]Strawson acknowledges that while the framework of moral responsibility is not externally justified, it allows for extensive internal revision—shaped by demands for coherence, reflection, and richer interpersonal understanding, as Hieronymi emphasizes.

bridge the gap between the engaged participant stance and the impartial viewpoint, primarily by re-evaluating the role and status of reactive attitudes and the nature of Strawson's underlying naturalism.

One highly promising avenue for reconciliation involves viewing the reactive attitudes not as the ultimate determinants of moral truth or justification but as crucial epistemic guides or starting points within a broader process of ethical reasoning. (Hieronymi, 2020: pp. 55-56).[31] Our intuitive, affect-laden responses—the sting of resentment at perceived unfairness, the warmth of gratitude for unexpected kindness, the swell of indignation at cruelty witnessed—can be understood as generally reliable, though fallible, trackers of morally salient features in our interactions (Strawson, 1962: pp. 6-7).[32] Any plausible moral theory, including those striving for agent neutrality, must ultimately account for and make sense of these core human responses to perceived goodwill or ill will. In this light, the reactive attitudes provide the initial data, the "considered judgments" about particular cases, that often fuel the process of seeking more general moral principles.[33] This mirrors the methodology of reflective equilibrium, a common approach in ethical theorizing where we move back and forth between our intuitions about specific scenarios and the general principles we formulate, revising each in light of the other until a coherent balance is achieved. Our reactive attitudes supply the rich, experientially grounded intuitions from which we might seek to extract or test general principles. These principles could very well be agent-neutral in their final form. These derived principles might then lead us, in turn, to critically re-evaluate and refine some of our initial reactive tendencies, perhaps identifying certain forms of resentment as excessive or specific patterns of indignation as reflecting bias rather than impartial justice. This model supports Pamela Hieronymi's analysis of Strawson's social naturalism, emphasizing the explicit room Strawson leaves for internal criticism, refinement, and the incorporation of ideals within the framework (Hieronymi, 2020: p. 80). While Strawson's framework originates in subjective attitudes, many of these attitudes are judged by shared standards of fittingness, making it possible to reach broadly agent-neutral moral assessments from within the participant stance. This kind of evaluation is analogous to how we might assess color: while colors are defined by typical human responses under normal conditions, it is still a fact that the bottom light on a traffic signal is green. Similarly, while resentment originates in a subjective feeling, we might agree—under shared social and rational norms—that it is or is not appropriate in a given situation. This shows that Strawson's framework may accommodate agent-neutral

[31]Hieronymi emphasizes that reactive attitudes are not merely emotional outbursts but carry normative significance; they reflect our implicit expectations of others and can serve as starting points for principled moral reflection.

[32]Strawson presents attitudes like resentment and gratitude as responsive to perceived goodwill or ill will, indicating that they function as socially embedded moral perceptions rather than arbitrary emotions.

[33]This mirrors the method of reflective equilibrium, a process in moral philosophy where considered judgments about particular cases and abstract principles are revised in mutual adjustment. The approach treats our intuitive responses, like reactive attitudes, as legitimate sources of ethical insight that can be refined through rational deliberation.

evaluative standards, even while remaining grounded in the interpersonal realm (Strawson, 1962: pp. 4-5, 21-22).[34] For example, one might initially feel resentment after being insulted, but later ask: was the insult intentional? Was it serious? Was I overly sensitive? These are principled assessments that involve evaluating whether the emotion fits the situation—not based on raw feeling, but on shared standards of fairness, intent, or harm. According to Hieronymi's reading, this means reactive attitudes are not brute emotions, but contain implicit normative judgments that can be revised or justified through reflection. Thus, Strawson's framework permits internal moral reasoning while remaining rooted in emotional engagement. Based on this interpretation, the framework of attitudes is not static or merely reflective of current norms; it possesses an inherent dynamism capable of evolving through reflection and potentially incorporating insights derived from or aligning with agent-neutral moral ideals (Strawson, 1962: pp. 15-16; Hieronymi, 2020: pp. 68-69).[35] The attitudes inform the principles, and the tenets discipline the attitudes.

5.2. Division of Moral Labor and Framework Reform

A second strategy for achieving compatibility involves proposing a division of moral labor between the two approaches, assigning them distinct but complementary roles within the overall moral landscape. Based on this view, Strawson's framework can be understood primarily as a descriptive and phenomenological account of the psychology and practice of holding one another responsible as participants in interpersonal life. It brilliantly captures what it feels like and what it means within our social world to blame, forgive, resent, feel indignant, or express gratitude. It elucidates the structure of the participant's stance. On the other hand, age is primarily concerned with providing the substantive criteria for the rightness or wrongness of the actions themselves, independent of our immediate reactions. Compatibility is then achieved by recognizing these distinct domains: the agent-neutral theory determines whether a moral norm has been violated (e.g., whether an action maximized utility or respected universal duties), while Strawson's framework explains the nature and conditions of the appropriateness of our engaged, attitudinal responses when such a violation is perceived as manifesting a certain quality of will (ill will, indifference, etc.) (Strawson, 1962: pp. 6-7, 15-16).[36] For example, utilitarianism might deem breaking a promise wrong because it leads to suboptimal consequences. If that promise-breaking also displayed a disregard for the promisee's reliance, then reactive attitudes like resentment or indignation, as described by Strawson, become intelligible and potentially appropri-

[34]Despite its openness to normative reflection, Strawson's framework remains anchored in the interpersonal structure of human moral life, resisting the urge to reduce moral responsibility to abstract metaphysical conditions.

[35]Both Strawson and Hieronymi allow for a reflective space within moral responsibility. While attitudes originate in emotional responses, their normative force—and potential revision—depends on intersubjective standards and reasoned evaluation.

[36]Strawson grounds the appropriateness of reactive attitudes in their responsiveness to the perceived quality of will—such as ill will, indifference, or good will—rather than in objective rule violation.

ate responses within the participant's stance (Strawson, 1962: pp. 16-17).[37] The agent-neutral theory identifies the breach of obligation, while Strawson's account explains the significance and character of our interpersonal reaction to the will behind that breach. This aligns well with Strawson's persistent focus on the quality of will as the object of the reactive attitudes; they are not merely responses to outcomes or rule violations per se but to the intentions, attitudes, and degree of regard manifested by the agent in their actions (Strawson, 1962: pp. 6-7, 12).

Furthermore, Strawson's social naturalism might be interpreted as conducive to compatibility. Suppose the core "natural human commitment" is, as Hieronymi suggests, a commitment to participate in some system of mutual expectation and demand grounded simply in the necessities of social co-existence (Hieronymi, 2020: pp. 28-29, 62). In that case, this foundational commitment does not necessarily preclude the specific content of that system from evolving to incorporate impartial, agent-neutral standards. The basic commitment might be to structured, evaluative social interaction. Still, the particular demands we make, the specific qualities of will we expect, and the precise character of our reactive attitudes could be significantly shaped by ongoing reasoned reflection, cultural change, and moral argument—including arguments explicitly appealing to agent-neutral ideals like fairness, equality, or overall well-being (Strawson, 1962: pp. 6-7; Hieronymi, 2020: pp. 67-68).[38] The "framework" might be a natural given from our sociality. Still, its internal structure could be highly malleable and responsive to reasoned critique, including critiques originating from an impartial perspective. Hieronymi's defense of social naturalism, highlighting the internal dynamics of pressure and counter-pressure driven by consistency and the competing needs and interests of participants, supports this view of a potentially evolving framework capable of incorporating more impartial considerations without betraying its naturalistic roots (Hieronymi, 2020: pp. 80, 84-85).

Finally, the apparent conflict over the justification of the framework might be softened by applying a similar distinction. Perhaps Strawson is correct that the framework as a whole—the bare fact that humans engage interpersonally via systems involving reactive attitudes and associated demands—needs no external rational justification and permits none, being a fundamental aspect of our social nature (Strawson, 1962: p. 22).[39] It is simply how creatures like us navigate a shared world. However, this immunity from external justification for the framework's existence need not extend to the specific norms, standards, and practices operating within that framework at any given time. These internal components—what counts as adequate regard, which capacities excuse or exempt, and how con-

[37]The moral significance of promise-breaking in Strawson's view depends not merely on the act itself but on the underlying attitude it expresses—whether it manifests indifference, contempt, or respect for others' expectations.

[38]Strawson acknowledges that reactive attitudes can be modified and refined through reflection and social evolution. Hieronymi develops this idea, arguing that the system of reactive attitudes accommodates pressures toward consistency and responsiveness to moral argument.

[39]Strawson argues that the participant stance is not something we adopt by choice or justify from outside; it is part of "what it is to be a human being."

sistently principles are applied—remain open to scrutiny, justification, and revision. Agent-neutral principles, therefore, could find a crucial role not in justifying (or failing to justify) the entire edifice of interpersonal engagement from the outside but in serving as powerful tools for internal criticism and refinement. They could provide the impartial criteria needed to evaluate existing social norms, challenge inconsistencies, argue for the inclusion of new ideals (like broader conceptions of equality or fairness), and guide the ongoing evolution of the standards of regard. Agent neutrality thus finds its place not as an external judge of the participant's stance but as a source of principles for improving our conduct (Hieronymi, 2020: pp. 68-69).[40]

However, acknowledging these pathways to reconciliation does not eliminate all difficulties. The partiality of personal reactive attitudes remains a challenge. While indignation might align well with agent-neutral condemnations of injustice, can intense personal resentment following a minor slight (but a slight nonetheless) always be squared with an impartial assessment of overall harm or universal principle? Furthermore, if a deep conflict arises where our "natural commitments" (as Strawson sees them) consistently pull against a strongly justified agent-neutral requirement, it's unclear which should yield. Does the natural facticity of the attitude grant it immunity, or does the rational force of the impartial principle demand a revision, even if psychologically costly or "practically inconceivable"? Strawson seems to lean towards the former, but this remains a point of profound philosophical tension (Strawson, 1962: pp. 14-15). The precise way our engaged, often partial, responses serve as reliable guides to impartial truths, as suggested by the reflective equilibrium model, also requires much more detailed explication than can be provided here.

Acknowledgements

I would like to thank Professor Hilary Bok for her invaluable guidance and for the thought-provoking class that inspired and shaped this paper.

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

Dennett, D. C. (1978). *Brainstorms: Philosophical Essays on Mind and Psychology.* Bradford Books.

Frankfurt, H. G. (1971). Freedom of the Will and the Concept of a Person. *The Journal of Philosophy, 68,* 5-20. https://doi.org/10.2307/2024717

Hieronymi, P. (2020). *Freedom, Resentment, and the Metaphysics of Morals.* Princeton University Press. https://doi.org/10.23943/princeton/9780691194035.001.0001

[40]Rather than offering an external justification of moral practices, agent-neutral theories can play a vital internal role—helping to refine, challenge, or improve norms within the reactive framework. This reinterpretation positions impartiality as a supplement to, not a replacement for, Strawson's interpersonal moral psychology.

Nagel, T. (1986). The View from Nowhere. Oxford University Press.

Parfit, D. (1984). *Reasons and Persons.* Oxford University Press.

Scanlon, T. M. (1982). Contractualism and Utilitarianism. In A. Sen, & B Williams (Eds.), *Utilitarianism and Beyond* (pp. 103-128). Cambridge University Press. https://doi.org/10.1017/cbo9780511611964.007

Scanlon, T. M. (1986). *The Significance of Choice. Tanner Lectures on Human Values.* Delivered at Brasenose College, Oxford University.

Scanlon, T. M. (2008). *Moral Dimensions: Permissibility, Meaning, Blame* (pp. 122-147). Harvard University Press. https://doi.org/10.4159/9780674043145

Smart, J. J. C. (1961). I.—Free-Will, Praise and Blame. *Mind, LXX,* 291-306. https://doi.org/10.1093/mind/lxx.279.291

Strawson, P. F. (1962). Freedom and Resentment. *Proceedings of the British Academy, 48,* 1-25.

Wolf, S. (1990). *Freedom Within Reason.* Oxford University Press.