philosophies

MDPI

Article

From Subjectivism and Pure Objectivism to Conditional Objectivism: A Criticism and Revision of Richard Arneson's Theory of Welfare

Chenju Xian [1] and Xinggui Mao [1,2,*]

[1] Department of Philosophy, College of State Governance, Southwest University, Chongqing 400715, China; xianchenju306@gmail.com
[2] Institute of Logic and Intelligence, Southwest University, Chongqing 400715, China
* Correspondence: justice@swu.edu.cn

Abstract

Richard J. Arneson's theory of welfare underwent a significant transformation from subjectivism to objectivism. Three difficulties—adaptive preferences, false beliefs, and non-prudential desires—demonstrate that a welfare theory grounded in subjective attitudes is untenable in principle, driving Arneson toward an objective list theory. Through his rejection of the endorsement constraint, he established a purely objectivist position: subjective attitudes are "neither necessary nor sufficient" for well-being. This position generates significant justificatory pressure toward hard paternalism. Arneson confronted this consequence by arguing that hard paternalism is defensible in principle, while contending that paternalistic intervention must remain restrained in practice on grounds of the intrinsic value of autonomy, the limitations of state capacity, and the costs of stigmatization. However, the reasons Arneson offers for restraint cannot be adequately supported by the objective list alone; they face explanatory pressure at the level of political application. Conditional objectivism can better fill this explanatory gap: for an item on the objective list to count as contributing to a particular individual's well-being, a negative condition must be satisfied—namely, that the individual would not rationally reject the item under conditions of reflective deliberation.

Keywords: welfarism; objective list theory; paternalism; endorsement constraint; distributive justice

Check for updates

Academic Editor: Thomas M. Besch

Received: 25 March 2026
Revised: 28 May 2026
Accepted: 5 June 2026
Published: 10 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

In contemporary Anglophone political philosophy on distributive justice, "Equality of What" has remained a central point of contention among philosophers. Since Amartya Sen raised this classic question, debates over the "currency of equality"—whether what ought to be equalized is resources, capabilities, or welfare—have dominated the theoretical landscape of political philosophy for the past half-century. Among defenders of welfarism, Richard J. Arneson stands as one of the most influential. He sought to develop an egalitarian approach that both respects personal responsibility and takes welfare as its ultimate concern, distinct from the resource egalitarianism of Ronald Dworkin and the capability egalitarianism of Sen—namely, "equality of opportunity for welfare".

In this theory, Arneson defined "welfare" as "preference satisfaction" and upheld a position he termed "distributive subjectivism": how well an individual's life goes is determined by the degree to which her desires or preferences are satisfied, and no external

standard of value should override the individual's own subjective evaluation. However, from the late 1990s onward, under the persistent challenges posed by adaptive preferences, false beliefs, and non-prudential desires, Arneson gradually abandoned this subjectivist stance and embraced an objective list theory, holding that an individual's subjective attitudes are "neither necessary nor sufficient" for something to intrinsically enhance her well-being. If welfare is purely objective, then an individual's judgment about whether something promotes her own well-being may be "dead wrong". Following this logic, Arneson faced significant pressure to defend hard paternalism: a state committed to promoting the welfare of its citizens need not, in principle, treat individuals' subjective judgments about their own well-being as an inviolable boundary. As for the limits of such paternalism, these are supposedly guaranteed by the "objective list" itself—autonomy, as a good on the list, constrains the scope of intervention. This paper aims to argue that when Arneson suggests paternalistic intervention must nonetheless remain somewhat restrained, he relies precisely on an element he has systematically rejected at the level of value theory—namely, some constitutive role for individual subjective attitudes in well-being. In other words, Arneson's "objectification" of the definition of welfare at the value-theoretic level carries a theoretical cost at the level of political philosophy: it makes it difficult for his theory to adequately account for the principled limits of paternalism by relying solely on the objective list. In order to address this difficulty, we offer a revision of his theory based on the endorsement constraint, which can be called conditional objectivism.

Section 2 of this paper introduces why Arneson's theory of welfare moved toward objectivism, analyzing the logic of his turn from subjectivism to pure objectivism. Section 3 presents the radical extent of Arneson's objectivist position, particularly examining his systematic rejection of various versions of the "endorsement constraint". Section 4 surveys the problems confronting Arneson's objectivist theory of welfare and how he responds to them. Section 5 argues that Arneson's strategy for limiting paternalism faces significant explanatory pressure given his commitments in value theory. Section 6 offers a possible revision, and Section 7 briefly discusses the implications of this revision for Arneson's distributive justice project.

2. Subjectivist Theory of Welfare and Its Difficulties

In his celebrated article "Equality and Equal Opportunity for Welfare" in 1989, Arneson explicitly defined "welfare" as preference satisfaction: "I take welfare to be preference satisfaction. The more an individual's preferences are satisfied, as weighted by their importance to that very individual, the higher her welfare." [1] (p. 82). In Arneson's view, the relevant "preferences" are "self-interested preferences"—what the individual desires with respect to her own advantage, rather than moral ideals or impersonal pursuits in general. To avoid the problem of "expensive tastes," Arneson further restricted these self-interested preferences to "ideally considered preferences"—preferences the individual would hold under conditions of thorough deliberation: "My ideally considered preferences are those I would have if I were to engage in thoroughgoing deliberation about my preferences with full pertinent information, in a calm mood, while thinking clearly and making no reasoning errors."[1] [1] (p. 83). This welfare theory centered on preference satisfaction was termed "distributive subjectivism" by Arneson. Its basic claim is that, in the face of fundamental disagreements about what constitutes the good life, a theory of justice should respect each individual's evaluation of her own life, rather than having philosophers or the state adjudicate what is "truly" good.

However, three fundamental difficulties confronting subjectivist welfare theory drove Arneson to gradually abandon this position. The first difficulty arises from the challenge of "adaptive preferences." In articulating his theory of equality of opportunity for welfare,

https://doi.org/10.3390/philosophies11030095

Arneson emphasized that "preferences" are "reasonable preferences" formed after careful deliberation, shaped by full consideration of external circumstances. Yet the rational "adaptiveness" that operates under adverse conditions can lead individuals to unconsciously lower their expectations and aspirations to fit their narrowly constrained reality[2]. Under extreme poverty or oppression (such as slavery), people suppress their desires to survive: "The hopeless destitute desiring merely to survive, the landless laborer concentrating his efforts on securing the next meal, the round-the-clock domestic servant seeking a few hours of respite, the subjugated housewife struggling for a little individuality, may all have learned to keep their desires in line with their respective predicaments." [5] (p. 191). If this is the case, a slave whose minimal wishes are satisfied may, on the welfarist view, count as having been "treated justly". If such preferences are accepted as the foundation of justice, justice degenerates into a tool for legitimating the status quo.

To address the challenge of adaptive preferences, Arneson attempted to invoke "the hypothetical ideal deliberation test". The presupposition of this test is that the root cause of adaptive preferences lies in the cognitive impairment suffered by the oppressed—people do not desire freedom or education mainly because they cannot think clearly about their situation. If this cognitive damage were removed and they could examine their preferences with full information and in a calm mood, they would identify desires distorted by unjust environments and restore their aspirations for freedom, education, or dignity [6] (p. 169). But Arneson immediately acknowledged a more intractable scenario: what if the damage to preference formation suffered by the oppressed is irreversible? A person who has been completely devastated psychologically by a "horrible preference-affecting injustice" may still, even under conditions of full information and cognitively unimpaired state, wholeheartedly endorse their distorted desires. In such a case, the hypothetical ideal deliberation test no longer has any corrective function. This difficulty forced Arneson to concede that, under a subjectivist welfare theory, if a person in unfortunate circumstances forms limited, unambitious desires that would be endorsed by an ideal advisor, then that person is living a good, choice-worthy life. By contrast, an objectivist welfare theory would deliver a verdict far more consonant with our intuitions. This was one of the key reasons that ultimately drove Arneson toward the objective list theory.

The second difficulty confronting subjectivist welfare theory concerns the problem of false beliefs. Rawls once posed the following thought experiment: if there is a person of extraordinary talent whose only desire is to count blades of grass, then merely satisfying his grass-counting preference would mean he is leading the best possible life for himself, no matter how much his talents are wasted [7] (pp. 379–380). Intuitively, this conclusion is unacceptable. Arneson drew from this a conceptual distinction of far-reaching significance for his theoretical development: that an individual has sovereignty over her own life and that an individual's actual choices genuinely make her life go better are two entirely distinct propositions. A person can exercise her legitimate right of self-determination while simultaneously committing serious errors in her factual judgments about her own well-being. It was precisely the recognition that judgments about one's own good lack first-person infallibility that constituted a key impetus for Arneson's abandonment of subjectivism.

The third difficulty concerns the extensional demarcation of the concept of welfare, namely the problem of non-prudential desires. Subjectivism equates welfare with the satisfaction of desires, but human beings possess desires that extend far beyond their own interests, and the fulfillment of some of these desires bears no relation to an individual's well-being. For instance, upon seeing a televised appeal for famine relief, a person may develop a strong desire to help distant starving strangers, yet the satisfaction of this desire does not intuitively constitute an enhancement of that person's own well-being [8] (p. 124). Derek Parfit also refutes this view with an example: "Suppose that I meet a

stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire-Fulfilment Theory, this event is good for me, and makes my life go better. This is not plausible." [9] (p. 494). But if we attempt to restrict welfare-relevant desires to those "about one's own life," the criterion for this restriction is itself deeply contested. A desire to sacrifice one's life for others is obviously "about one's own life" (since it involves giving up one's life), yet its satisfaction does not intuitively enhance the agent's welfare. Arneson further realized that allowing individuals to determine for themselves which of their desires are welfare-relevant would render the theory viciously circular, since making this determination already presupposes a concept of welfare independent of desire satisfaction. Notably, Arneson acknowledged that even a shift to the objective list theory cannot fully dissolve this difficulty—if "satisfaction of one's important basic desires" is itself an entry on the objective list (as seems reasonable), then the problem of screening which desires count as "welfare-relevant" persists within that entry. However, he maintained that the objective list theory at least narrows the scope of the problem: it no longer requires grounding all of welfare on desire satisfaction, but need only address this technical difficulty within a single entry.

It was the pressure of these three difficulties that drove Arneson to make a decisive turn in the late 1990s. In Arneson's view, the failure of subjectivist welfare theory is not technical but principled: the theoretical approach of taking individual subjective attitudes as the constitutive basis of well-being is itself unviable. The question, however, is how far Arneson's objectivist turn extends. To what degree does he reject the role of subjective elements? Answering these questions requires a closer examination of his treatment of the key stance of the "endorsement constraint".

3. The Rejection of the Endorsement Constraint

Having abandoned subjectivism, Arneson turned to the objective list theory—a purely objectivist theory of welfare[3]. "Objectivism" is itself a broad label that encompasses different theoretical approaches. A moderate objectivism might acknowledge that welfare has objective content while still treating individual subjective attitudes as a necessary condition for something to enhance well-being. But Arneson opted for a more radical and pure objectivist position, categorically denying any constitutive role for subjective attitudes.

On this objectivist position, Arneson affirmed "realism about prudential value"—the claim that there are objective facts about "what is good for an individual", facts that obtain independently of whether the individual endorses them or holds positive attitudes toward them. Just as facts about the physical world do not depend on whether we believe them, facts about prudential value (i.e., what constitutes an individual's well-being) likewise do not depend on an individual's subjective beliefs or preferences. This constitutes the philosophical foundation of the objective list theory.

Against this theoretical position, Sumner raised an influential objection: in principle, subjectivism and objectivism can generate similar lists of goods (such as knowledge, friendship, achievement, etc.), and merely providing a list does not constitute an independent welfare theory, since the real disagreement lies in the more fundamental question of "what makes something valuable for a person" [10] (pp. 45–47). Arneson's response was that even if the two lists are similar in content, the status of the items on each list differs between the two theories. For the subjectivist, the list is provisional and revisable—if people under appropriate conditions no longer hold positive attitudes toward an item, that item should be removed. For the objectivist, by contrast, items are on the list not because people happen to hold certain attitudes toward them, but because there is the strongest reason to regard them as genuinely valuable [8] (pp. 118–119). An objective list theory can perfectly well maintain

that "one good thing for an individual is that her important life aims be satisfied, with importance determined by her own subjective ranking" [8] (p. 117). Subjective attitudes may thus appear within the content of the list, but they do not function as a precondition for any item on the list to have prudential value.

However, Arneson's denial of any universal filtering function for subjective attitudes within his objective list theory has drawn considerable criticism. These critics, though differing in their theoretical starting points, commonly insist that individual subjective attitudes should play a stronger role in the constitution of well-being than Arneson allows. This insistence has been captured under the heading of the "endorsement constraint", whose core requirement is that for any objective good (or the object of an idealized desire) to enhance a particular individual's well-being, it must be supported by some positive attitude on the part of that individual. For Arneson's objectivist position to stand, he must first reject the endorsement constraint.

It is worth noting that the most canonical contemporary formulation of the endorsement constraint comes from Will Kymlicka, who explicitly argued: "No life goes better by being led from the outside according to values the person doesn't endorse. My life only goes better if I'm leading it from the inside, according to my beliefs about value." [11] (p. 12). This "leading life from the inside" argument constitutes the original version of the endorsement constraint, which holds that an individual's subjective endorsement of a good is a necessary condition for that good to possess welfare value. Thomas Hurka subjected Kymlicka's argument to systematic critical examination, distinguishing between a "strong version" and a "weak version" of the endorsement constraint: the strong version claims that endorsement is necessary for an activity to have any value at all (i.e., without endorsement, value is zero); while the weak version claims only that endorsement is necessary for an activity to have its highest value (i.e., without endorsement, value is diminished but not eliminated). Hurka argued that only the strong version can support the conclusion of state neutrality, but the strong version is implausible when confronted with concrete cases [12] (pp. 40–42). Arneson's rejection of the endorsement constraint can largely be understood as a continuation and deepening of Hurka's critique.

Arneson examined three versions of the endorsement constraint, arranged in decreasing order of strength. The first version comes from Dworkin. In analyzing how to evaluate the goodness of a life, Dworkin drew a distinction between two fundamentally different approaches: the "additive view" and the "constitutive view" [13] (pp. 80–83). The additive view holds that the various components of a person's life—achievements, experiences, relationships, and so on—possess objective value in themselves, and the individual's endorsement of these components merely adds an extra layer on top of this objective value. In other words, the individual's endorsement has a supplementary value-enhancing effect, but even without endorsement, the value of the other components remains independently intact. On the additive view, coercive paternalism is in principle defensible: if someone is forced to live a life whose objective components are superior, her well-being is enhanced regardless of whether she endorses that life. In contrast, Dworkin himself favored the "constitutive view", according to which no component of a life can make a non-instrumental contribution to its value unless it is endorsed by the person living it [13] (p. 81). In other words, endorsement is not an added bonus to value but a necessary condition for value to be constituted. Dworkin illustrated this with the example of the "misanthrope": if a misanthrope is deeply loved by others but regards that love as worthless, then the love of others does not make his life better for him. From the constitutive view, one can naturally derive a constraint on all prudential goods: anything that intrinsically enhances an individual's well-being must be accompanied by some positive attitude on the part of that individual. Under this understanding, the endorsement constraint can be considered as an

internal constraint on the objective list theory: the goods on the list are objective, but their translation into a particular individual's well-being requires passing through the subjective gate of endorsement.

The second version examined by Arneson comes from James Griffin. Unlike Dworkin's approach, Griffin's version arises from a more subjectivist theoretical context-informed-desire theories [14] (p. 11)[4]. This theory holds that what is truly good for a person is what she would desire upon full acquaintance with the relevant information and rational reflection. The theory determines the content of an individual's well-being by constructing an "idealized self" or "ideal advisor"—not what the individual actually desires at present, but what she would desire under epistemically ideal conditions. The problem, however, is that hypothetical desires under idealized conditions may diverge radically from an individual's actual present desires. For example, an idealized, fully informed version of me might desire that I study quantum physics, but the actual me has no interest whatsoever. In such a case, if social arrangements cause me to study quantum physics, thereby "satisfying" that idealized desire, does this count as enhancing my well-being? Griffin's answer is that for a hypothetical idealized desire to contribute to an individual's well-being upon being satisfied, that desire must have become an actual desire held by the individual at the time of its satisfaction. That is, merely "the idealized me would desire X" is not enough—the individual herself must actually desire X. The idealized desire must be endorsed and internalized as her actual desire before the attainment of X can count as enhancing her well-being [14] (p. 11).

Arneson employed the same rebuttal strategy against both Dworkin's and Griffin's versions of the endorsement constraint. He argued that although the two versions differ in their starting points, they share a core claim in their logical structure: if an individual does not hold a positive endorsing attitude toward a given good, then that good cannot count as enhancing her well-being. This claim grants the individual's subjective attitude a veto power—no matter how objectively valuable a good may be, the absence of the individual's subjective endorsement suffices to nullify its welfare contribution. Arneson's rebuttal targets precisely this point: people's reasons for refusing to endorse a given good can be weak, confused, or even nonexistent, and in such cases, allowing a refusal based on poor reasons to negate the welfare value of a good is unreasonable. Arneson illustrated this with the following example: imagine that Samantha has written an excellent poem, but she denies that this achievement has any value or in any way enhances her life. She makes this denial on the basis of "a shallow and silly aesthetic theory" that she has accepted without reflection. In these circumstances, her failure to endorse her achievement does not negate its value for her [8] (p. 136). Arneson conceded that if Samantha were able to endorse her achievement, her well-being would be higher (consistent with the additive view), since a well-grounded sense of accomplishment is itself a considerable good. But endorsement is by no means a precondition for well-being.

Arneson further considered a possible response: perhaps the reason Samantha's achievement still has value is only qua "impersonal achievement", not in terms of its value for her personally. Arneson rejected this distinction. He argued that if we compare two lives under otherwise identical conditions, one containing excellent poetry and the other not, most people would judge the former to be a better life for the person living it, even if that person does not endorse the achievement. The key point is not to evaluate the achievement in the abstract, but to affirm that it genuinely enhances the welfare of the person concerned [8] (pp. 136–137). Arneson also pointed out an often-overlooked consideration: the attractiveness of the endorsement constraint partly derives from a confusion between an empirical generalization about human psychology and a conceptual philosophical claim. A thing not recognized as valuable typically has a harder time producing a sense of value

for us, but recognition of value is not itself identical to value. People tend to accept the endorsement constraint because they misread the empirical proposition "subjective endorsement typically helps one to obtain objective goods" as the conceptual proposition "without subjective endorsement it is impossible to obtain objective goods".

The refutation of the first two, stronger versions might be thought to leave room for a weak endorsement constraint. After all, even if Samantha does not endorse her poetic achievement, she might still hold positive attitudes toward some aspects of her life. Arneson recognized this and therefore targeted the "weak endorsement constraint" as an independent objective. He constructed a strengthened version of the Samantha case: Samantha has not only written excellent poetry but also experiences genuine reciprocal love—she deeply loves another person who loves her in return. But Samantha subscribes to a Stoic philosophy of life according to which emotional attachment to others diminishes the quality of one's life, and so "She neither endorses this loving relationship nor actually desires it." [8] (p. 140). Moreover, she maintains the same detachment toward all other objectively valuable achievements in her life. Yet she is not unhappy; on the contrary, she possesses ample pleasure, satisfaction, and related forms of subjective contentment—but these subjective satisfactions are not connected to her objective achievements in the right way, and thus fail to meet the requirements of the weak endorsement constraint. Faced with this case, the weak endorsement constraint is forced to deliver what Arneson regards as a deeply implausible verdict: even though we acknowledge that having a loving relationship is a great human good, and even though Samantha indeed has such a relationship, this attainment has "no value whatsoever" for her welfare, solely because she lacks a positive attitude toward it.

Arneson also anticipated a defense of the weak endorsement constraint: if a person lives a good life according to the objective list but fails to satisfy the weak endorsement constraint, would she not be unhappy throughout her life, with all important desires unfulfilled and completely lacking in subjective satisfaction? If so, such a life would indeed be poor in prudential value. Arneson responded that if the objective list itself has the correct internal structure, this worry does not arise. He proposed two structural features: (1) the objective list includes certain types of subjective satisfaction among its entries; (2) the objective list has a threshold structure, requiring that an individual must achieve a certain level of subjective satisfaction—regardless of how high she scores on other dimensions of a good life—in order to count as having attained a satisfactory overall level of well-being [8] (p. 140). These two features are plausible, and a correct objective list theory would likely include them. But the crucial point is that a life satisfying both features can still fail to meet the weak endorsement constraint. For the sources of subjective satisfaction required to fulfill (1) and (2) need not include any positive attitude toward those elements of one's life that are entries on the objective list and that intrinsically enhance one's well-being [8] (p. 140). In other words, a person can derive ample subjective satisfaction from everyday leisure and simple pleasures while remaining entirely indifferent to her objective attainments in love, friendship, and creative achievement.

Having rejected all three versions of the endorsement constraint, Arneson stated his final position: "There is no attitude that an individual must have toward an element in her life if that element is to qualify as intrinsically augmenting her well-being according to the objective-list theory. To count as such, the goods in an individual's life need not be tinged with enjoyment nor colored by desire. In other words, for something to qualify as intrinsically enhancing an individual's well-being, it (1) is neither necessary nor sufficient that the individual actually desire that thing, and (2) is neither necessary nor sufficient that the individual actually enjoy (or have some other sort of positive experience of) that thing." [8] (p. 142). This conclusion pushed Arneson's position to the extreme end of

objectivist welfare theory: subjective attitudes constitute in principle neither a threshold nor a filter for well-being. This stance of categorically excluding any constitutive role for subjective attitudes at the level of value theory laid the groundwork for Arneson's arguments at the level of political philosophy—especially regarding paternalism—but also sowed the seeds of theoretical difficulties yet to be examined.

4. The Objective Theory of Welfare and the Limits of Paternalistic Intervention

The analysis of the preceding section showed that Arneson's objective list theory denies in principle any constitutive role for subjective attitudes in well-being: whether an individual desires something or enjoys it is not, and should not be, a precondition for that thing to enhance her welfare. Once this value-theoretic position is established, it produces a direct consequence for political philosophy: if an individual's judgment about her own good has no constitutive standing, then a person's saying "I don't want this" or "I don't consider this valuable for me" no longer constitutes a principled reason against others (including the state) making welfare-relevant decisions on her behalf. Under a subjectivist framework, individual preferences are themselves the content of welfare, so to interfere with preferences is to damage welfare, and the anti-paternalist position thereby gains a solid foundation. Under an objectivist framework, however, individual preferences can deviate from true well-being, and the state can in principle be a better judge of what is good for her, thereby intervening in the name of promoting her objective well-being. Arneson himself recognized this: "Hard paternalism naturally consorts with a conception of our ultimate striving as not to satisfy whatever desires and aims we now happen to embrace—these could all be failing to track correct values—but to achieve genuinely worthwhile values over the course of one's life." [16] (p. 267). Here, "genuinely worthwhile values" are precisely the goods on the objective list. Arneson further noted that under such a conception of the good, "we cannot rule out the possibility that restriction of an agent's liberty against her will for her own good may turn her from a track along which she will have a grim life to a track along which her life will be genuinely a life that is good for her." [16] (p. 267). In other words, the objectivist theory of welfare not only logically permits hard paternalism but provides positive reasons for it.

The implication of paternalism is not an oversight on Arneson's part but rather a theoretical challenge he must confront. Arneson's response can be divided into two levels: on the one hand, he argues that hard paternalism is defensible in principle, forthrightly acknowledging and accepting this political-philosophical consequence; on the other hand, he also maintains that paternalistic intervention should not be unlimited. The following examines each level in turn.

Arneson's first step in defending paternalism is to criticize the doctrine of personal sovereignty. Arneson's objective list theory denies "agent sovereignty"—the claim that the individual is the final arbiter of her own good, holding that an individual's judgment about what is good for her can be mistaken. But denying agent sovereignty at the level of value theory is not yet sufficient to justify paternalism, for there remains a distinct political-philosophical question: does the state have the right to intervene in an individual's self-regarding decisions on the grounds that "you are wrong"? The liberal tradition represented by Mill and Feinberg answers negatively. Mill declared in On Liberty: "The only part of the conduct of any one, for which he is amenable to society, is that which concerns others. . . Over himself, over his own body and mind, the individual is sovereign" [17] (pp. 80–81)[5]. Joel Feinberg inherited this intellectual legacy, distinguishing between "soft paternalism" and "hard paternalism": the former permits intervention only when an individual's choice is insufficiently voluntary; while the latter holds that even fully voluntary choices may be

overridden for the individual's own good. Feinberg accepted soft paternalism but firmly rejected hard paternalism [18] (p. 12).

Arneson responded to the doctrine of personal sovereignty from two directions. First, any absolutist claim about a moral right is implausible: a law requiring drivers to wear seatbelts and a theocratic government requiring citizens to live strictly according to the dictates of a particular religion are both paternalistic, yet they differ vastly in the degree to which they infringe personal sovereignty. Second, Feinberg's soft paternalist position rests all its theoretical force on the concept of "voluntary choice", but voluntariness and reasonableness are two different things—a point Feinberg himself concedes [18] (p. 112). A person who is fully alert and fully informed, knowing that not wearing a seatbelt significantly increases the risk of death in a car accident, may still choose not to wear one. This choice scores very high on Feinberg's voluntariness scale, yet from a prudential standpoint it is clearly unreasonable. In such cases, the soft paternalist cannot intervene (because the choice is voluntary), but the hard paternalist can (because the choice damages the individual's objective welfare). Treating voluntary choice as an irrevocable shield is a form of fetishism about voluntary choice.

Arneson's most powerful challenge to Feinberg comes from his constructed case of the "Pouting Young Adult" [16] (pp. 272–273). A young man named Tom, upset over some trivial disappointment—perhaps losing a job competition, being broken up with by a partner, or having a pet rabbit he wanted adopted by someone else—feels dejected and wants to express his displeasure by committing suicide. He is not deceived; he fully knows that if he goes on living, he will forget this disappointment and continue with a normal life, but he simply does not want that future right now. He has no mental illness and no lack of agency; he simply holds unusually willful and immature preferences. Under Feinberg's doctrine of personal sovereignty, Tom's choice is fully voluntary, and therefore he has a moral right to commit suicide, which others and the state should not interfere with. Arneson regards this conclusion as unacceptable.

In response to this case, Arneson proposed the "secular sanctity of life doctrine". The core idea of this doctrine is that under most normal circumstances, being alive and having a life to lead constitutes a precious opportunity, and making good use of this opportunity so that one's life has value for oneself and others constitutes a moral duty. The content of this duty is vague and admits of extensive gray areas, but in cases like Tom's, the individual has utterly failed to fulfill this duty, and hard paternalistic intervention is therefore morally defensible [16] (pp. 273–274). Arneson emphasized that this doctrine can distinguish the Pouting Young Adult's suicide from a terminally ill patient's choice of euthanasia to escape suffering: for the former, life still constitutes a precious opportunity; for the latter, the value of that opportunity has been all but extinguished by disease and pain. By contrast, Feinberg's doctrine of personal sovereignty cannot make this distinction, since both suicides are voluntary and enjoy equal standing under the protection of personal sovereignty.

Arneson's second path to defending hard paternalism draws on the perspective of distributive justice. He observed that people differ enormously in their capacity for prudent choice—some are skilled at gathering information, evaluating options, making sound decisions, and following through ("good choosers"), while others are severely deficient in these respects ("bad choosers"). Arneson did not assume that responsibility for this ability gap lies entirely with the individual; on the contrary, he acknowledged that genetic endowment and early socialization play significant roles. Good choosers would gain little from paternalistic policies (since they always make the best choice from available options), while bad choosers might benefit substantially from reasonable paternalistic restrictions. At the same time, bad choosers tend over the course of their lives to fare worse than good choosers [19] (p. 197). Arneson thus advanced a distributive justice argument: if we hold

some form of prioritarianism—the view that improving the welfare of the worse-off has greater moral value—then we have additional reason to support paternalistic policies. Since bad choosers tend to cluster among the worst-off, and paternalistic policies benefit bad choosers disproportionately, this argument further makes the anti-paternalist stance look, at the level of distributive justice, like "an ideology of the good choosers"—a doctrine that favors the already better-off at the expense of the more vulnerable and needy.

It is important to clarify that while Arneson argues that hard paternalism is defensible in principle, he does not claim the state may intervene without limit in the name of objective welfare. He also identified several reasons why paternalistic intervention must remain restrained in practice.

The first reason comes from within the objective list theory itself. Arneson holds that autonomy is itself an important good on the objective list. If this is correct, then paternalistic intervention, while potentially enhancing welfare on other dimensions, simultaneously inflicts a loss of welfare on the dimension of autonomy. In some cases, the loss in autonomy will outweigh the gains on other dimensions, making intervention detrimental to the individual on the whole. Arneson writes: "If autonomy is rightly regarded as a significant component of an individual's welfare... then restricting her freedom for her own good at the expense of her autonomy will sometimes be wrong simply because the loss in autonomy she would suffer would outweigh the gains in other components of her welfare the contemplated restriction will bring about." [19] (p. 199). In other words, goods on the objective list can conflict with one another, and autonomy as one such good naturally imposes an internal constraint on the scope of paternalistic intervention.

The second reason concerns the limited capacity of the state and the crudeness of its policy instruments. Arneson acknowledges that laws and public policies are inevitably blunt instruments that cannot finely distinguish each individual's specific circumstances. Even if, in principle, paternalistic intervention would enhance a particular individual's welfare, in practice, translating this judgment into universally applicable policy may result in numerous erroneous interventions—for instance, subjecting good choosers who need no protection to restrictions, while administrative corruption or implementation errors may cause unintended harm to bad choosers.

The third reason involves a special type of welfare damage—stigma costs. Publicly distinguishing bad choosers from good choosers and treating them differently inevitably brands the former as "inferior" in some sense. Arneson concedes that such branding carries a moral cost: it causes those branded to feel negatively about themselves and diminishes their sense of personal efficacy. This means that in some cases, the consideration of avoiding stigmatization will take priority over the goal of precisely tracking each individual's objective welfare, thereby limiting the granularity and scope of paternalistic policies.

In sum, Arneson honestly confronted the justificatory pressure that pure objectivism generates toward hard paternalism, arguing through his critique of the theoretical foundations of anti-paternalism and his appeal to the additional argument from distributive justice that hard paternalism is defensible in principle. On the other hand, he argued from the intrinsic value of autonomy, the limitations of state capacity, and the costs of stigmatization that paternalistic intervention must remain restrained in practice. Whether these reasons for restraint can find adequate explanatory support from within Arneson's objectivism, however, is a question that demands further scrutiny.

5. The Coherence Problem of Objectivist Welfare Theory

Richard Arneson argues that paternalism must remain restrained, and the reasons he offers for this view initially appear plausible. However, we will argue in this section that, although these reasons can to some extent be reinterpreted within an objectivist

framework, they nonetheless reveal that Arneson cannot draw persuasive principled limits on paternalistic intervention by appealing solely to an objective list that is independent of the subject's perspective. Rather, he must introduce some form of subject-perspectival, context-sensitive evaluative element. Specifically, these reasons either covertly rely on subjective elements he has already rejected in the value theory, or they are merely pragmatic expedients rather than principled boundaries derivable from within the objective list theory. If this argument succeeds, then the cost of Arneson's objectivist welfare theory is not merely the logical possibility of paternalism, but a deeper difficulty: insufficient explanatory resources at the level of political application.

The first incoherence concerns the status of autonomy in Arneson's theory. Arneson maintained that autonomy can demand restraint from paternalism from within the objectivist welfare theory: if autonomy is itself a good on the objective list, then paternalistic intervention, while enhancing welfare on other dimensions, simultaneously inflicts a welfare loss on the dimension of autonomy, potentially making the intervention counterproductive on the whole. This argument appears to grow organically from within the objective list theory—autonomy demands restraint not because the individual possesses some moral right independent of welfare, but because autonomy is itself a component of welfare. But the question is: what is the substantive content of "autonomy" as a good on the objective list? By Arneson's own definition, autonomy means one "pursues and achieves goals that she adopts for herself after reflective scrutiny". The core element of this definition is the individual's reflective endorsement of her own goals. A person is autonomous not because she happens to be living an objectively good life, but because she has reflectively endorsed her life goals and acts accordingly. In other words, the essence of autonomy as a good lies precisely in the individual's subjective attitudes—the specific relationship between her reflection, judgment, and endorsement on the one hand, and her way of life on the other.

It is instructive to compare Arneson's position with Joseph Raz's perfectionist liberalism. Raz likewise treats autonomy as an objective good and develops a perfectionist position in which the state has an obligation to promote citizens' good lives, with autonomy as a core element. But Raz's concept of autonomy differs fundamentally from Arneson's. Raz's autonomy centers on "appropriate mental abilities", "an adequate range of options", and "independence" [20] (p. 372)—all external conditions or general capacities rather than the individual's reflective attitudes toward specific life goals. Raz can therefore value autonomy while endorsing certain paternalistic measures, provided these measures are "compatible with respect for autonomy" and "confined to the creation of the conditions of autonomy" [20] (p. 423). Raz's case demonstrates that an objectivist can indeed value autonomy without reintroducing an endorsement constraint—on the condition that the concept of autonomy invoked centers on external conditions rather than subjective attitudes. But Arneson's situation is different. Arneson defines autonomy as "pursuing goals adopted after reflective scrutiny", in which "after reflective scrutiny" weaves reflective endorsement into the very internal structure of autonomy. Unlike knowledge or health, whose attainment does not conceptually require reflective attitudes, autonomy as Arneson defines it is a good whose realization internally contains a specific subjective attitude. Therefore, when Arneson places this reflective-scrutiny-linked autonomy on the list and assigns it sufficient weight to restrain paternalism, he is not introducing a good structurally identical to other objective goods; he is introducing a good whose conceptual structure itself requires reference to reflective attitudes. This makes autonomy functionally closer to a constitutive condition than to a defeasible item standing alongside others on the list.

This, then, generates a theoretical dilemma. If Arneson takes seriously the standing of autonomy as an objective good, then the state has principled reasons to attend to

an individual's reflective judgments about her own life when considering paternalistic intervention—not as an external pragmatic consideration, but as a constitutive dimension of welfare itself. When a person, after reflection, sincerely judges "I do not consider X valuable for me", forcing her to accept X is not merely a practical difficulty; it damages, in principle, her autonomy qua objective good. Yet this position is substantively very close to the endorsement constraint that Arneson has systematically rejected. Arneson may of course respond that autonomy is merely one good among many on the list, overridable by greater gains in other goods, whereas the endorsement constraint is a universal precondition spanning all goods—a structural difference. But this formal distinction fails to reach the substance of the problem. As the foregoing analysis of Raz has shown, the key issue is not whether autonomy is "just one item on the list", but whether the internal structure of the specific autonomy Arneson invokes makes it unlike other items on the list. The attainment of goods such as knowledge, friendship, and health does not conceptually require the individual's reflective attitudes as a constitutive element: a person can acquire knowledge without being aware of it, and can possess health without having reflected on it. But the very definition of autonomy contains the subjective-attitude element of "goals adopted after reflective scrutiny". This means that autonomy is not a good that merely happens empirically to be accompanied by subjective attitudes (as friendship typically accompanies joy); it is a good that conceptually requires a specific subjective attitude as a constitutive element. Placing such a good on the list and assigning it sufficient weight to restrain paternalism effectively elevates the individual's reflective endorsement to an unavoidable consideration in welfare assessment, creating a deep contradiction with the pure objectivist commitment that "subjective attitudes are neither necessary nor sufficient". Arneson rejected the principled standing of the endorsement constraint at the value theory level, yet at the political-philosophical level he is compelled to restore its practical effect through the objective good of autonomy—a good that internally contains reflective endorsement. This detour shows that the "internal limitation" that pure objectivism provides for paternalism substantively depends on what it sought to eliminate at the value theory level—namely, a certain irreducible weight of individual subjective attitudes.

The second incoherence concerns the evaluative standard on which Arneson's "secular sanctity of life doctrine" relies when distinguishing cases where intervention is appropriate from those where it is not. As noted in the previous section, Arneson used this doctrine to respond to the Pouting Young Adult case, holding that life under normal circumstances constitutes a precious opportunity, that making good use of this opportunity is a moral duty, and that the Pouting Young Adult has utterly failed in this duty, rendering hard paternalistic intervention justified. Arneson also stressed that this doctrine can distinguish the Pouting Young Adult's suicide from a terminally ill patient's choice of euthanasia, where the opportunity value of life has been all but extinguished by disease and suffering, and the state ought not to force the individual to continue enduring.

The question is whether a purely objectivist theory of welfare possesses the theoretical resources to make this distinction. The objective list theory tells us that life is a good, but this alone cannot distinguish the two cases Arneson describes. For in both the Pouting Young Adult case and the terminally ill patient's case, the objective list itself has not changed (life, health, rational capacity, etc., remain entries on the list); what has changed is the specific individual's possibilities and costs for realizing these goods in her particular circumstances. To judge whether a person's life in her current circumstances still constitutes a precious opportunity, one must introduce an evaluative perspective—from the standpoint of a reasonable agent situated in those particular circumstances, weighing the goods achievable by continuing to live against the costs to be borne, and then judging whether this life still has meaning for a person in such a situation. This evaluation is not purely objective, because

it does not merely ask "is life on the objective list?" but requires an intermediate step that takes into account the particular individual's circumstances, and renders a judgment about the significance of this good for this person from the standpoint of a reasonable subject.

This issue reveals a structural difficulty of pure objectivism. The objective list tells us which things are goods, but it does not by itself answer a different question: how much prudential value a good has for a particular individual in particular circumstances. Life is a good, but life's prudential value for the Pouting Young Adult and for the terminally ill patient is vastly different, and this difference cannot be derived from the bare fact that 'life is on the list'. It requires an intermediate judgment: in the relation to this person, do the goods achievable by continuing to live, weighed against the costs to be borne, still constitute a meaningful trade-off? Such a judgment is inherently perspectival; it requires the evaluator to assess the significance of a good from the standpoint of a reasonable agent situated in those particular circumstances. Pure objectivism lacks the theoretical resources to generate such judgments, because its core commitment is precisely that the value of goods is independent of the subject's perspective. This difficulty points toward a more general conclusion: if the prudential value of a good for a particular individual is context-dependent, then there must be some mediating link between the objective list and the individual's welfare—an evaluation conducted from the standpoint of a reasonable subject that takes into account the specific circumstances. The 'negative endorsement constraint' proposed in the following section can be understood as a formalization of this mediating link: it does not replace the objective list in determining what is good (that remains the list's work), but sets a checkpoint between the list and the individual, requiring that the prudential value assessment not entirely bypass the individual's rational reflective judgment in her specific circumstances.

The third source of explanatory pressure concerns Arneson's treatment of stigma costs. Arneson acknowledges that publicly branding individuals as "bad choosers" constitutes a morally relevant harm: such sorting "brands the less competent as inferior in a public way, which makes the branded feel badly about themselves and also diminishes their sense of personal efficacy" [19] (p. 198). An objectivist can respond that personal efficacy is itself an objective capacity, and that stigma instrumentally compromises many other goods on the list—opportunities, relationships, social standing. This response has some force. But even if we accept this reinterpretation, it at least shows that pure objectivism, when dealing with stigmatization, must take into account how individuals perceive themselves—a dimension that, however it is redescribed as an objective capacity, inescapably depends on some irreducible weight of subjective attitudes at the factual level. Compared with the first two sources of explanatory pressure, this argument carries relatively less force, but it provides supplementary evidence that pure objectivism faces difficulties in excluding subjective attitudes across multiple dimensions.

6. Toward a Conditional Objectivism

The analysis of the previous section demonstrated that Arneson's pure objectivism cannot rely solely on its own theoretical resources to delineate principled boundaries for paternalistic intervention. If this diagnosis is correct, then a structural adjustment to Arneson's theory is needed. This section offers a possible alternative: one that preserves the core objectivist critique of subjectivism while finding an appropriate place for the individual's rational reflective attitudes in the constitution of well-being.

To define the content and boundaries of this proposal precisely, it is necessary first to distinguish the different types of subjective attitudes an individual may hold toward a given good. Broadly speaking, an individual's attitude can be classified into four types:

Active endorsement: The individual recognizes upon rational reflection that the good is valuable for her and actually desires to obtain it.

Passive endorsement: The individual recognizes upon rational reflection that the good is valuable for her, but does not particularly desire to obtain it—perhaps due to inertia, distraction, or other non-cognitive reasons.

Neutral attitude: The individual thinks that the good would bring neither obvious benefit nor obvious harm to her, and is indifferent about whether she obtains it.

Rejection: After reflection, the individual explicitly judges that the good has no value for her and does not believe that obtaining it would enhance her well-being.

This four-fold distinction reveals an important fact: between "active endorsement" and "explicit rejection" lies a wide intermediate territory. The strong version of the endorsement constraint (such as Dworkin's constitutive view) requires that an individual must hold a positive endorsing attitude toward a good for it to count as enhancing her welfare. This means that even in the second attitude-type—where the individual recognizes upon rational reflection that a good has value but does not particularly desire it due to inertia or other non-cognitive reasons—the strong endorsement constraint would deny the good's welfare contribution for lack of a sufficiently positive attitude. In the third attitude-type, where the individual holds a neutral, indifferent stance toward a good, the strong endorsement constraint flatly rules that the good has no welfare value for this person. Arneson's Samantha case targets precisely this requirement. Samantha has written an excellent poem, but on the basis of an unreflectively accepted aesthetic theory she holds no positive attitude toward this achievement—her attitude falls between the second and third types. On the logic of the strong endorsement constraint, this poetic achievement contributes nothing to Samantha's welfare. But Arneson convincingly pointed out that Samantha's non-endorsement stems from a poor reason, and allowing such a reason to veto the welfare value of an objective achievement is unreasonable. On this point, Arneson's critique succeeds.

On the other hand, however, Arneson's pure objectivism goes too far. It not only denies the constitutive standing of subjective attitudes in the first three attitude-types (which is reasonable), but equally denies any principled weight for subjective attitudes in the fourth type. On the logic of pure objectivism, even if an individual, after thorough reflection, explicitly judges that a given good has no value for her, the state may still justify intervention on the sole basis that "this good is on the objective list"—because subjective attitudes are "neither necessary nor sufficient". It is precisely this overextension that traps pure objectivism in the dilemma analyzed in the previous section.

On the basis of this analysis, we can revise Arneson's purely objectivist list theory into a 'conditional objectivism': the objective list of goods still holds independently of an individual's actual preferences, but for an objective good to count as contributing to a particular individual's welfare, an additional negative condition must be met—the individual, once cognitive distortions, informational deficits, and irrational psychological mechanisms are set aside, would not rationally deny that the good has value for her life.

Conditional objectivism is connected, at the level of its strategy for limiting paternalistic intervention, to a noteworthy distinction in Sarah Conly's defense of coercive paternalism. Conly explicitly distinguished paternalism from perfectionism: paternalistic intervention must be grounded in the individual's own goals and values, helping them more effectively realize their own pursuits; perfectionism, by contrast, imposes ways of life that the government or culture deems objectively valuable, regardless of whether the individual endorses them [21] (pp. 102–112). Conly's key limiting condition is that "paternalism is not perfectionism"—intervention must be based on the individual's own values rather than externally determined ones. Conditional objectivism can be understood as a theoretical advance beyond Conly's core concern. Conly still employs a purely subjectivist

welfare standard—she defines welfare as the maximal satisfaction of subjective desires, and paternalistic intervention merely helps people realize their already-held goals more effectively. Conditional objectivism, by contrast, retains the existence of objective goods (enabling it to address problems like adaptive preferences that subjectivism cannot handle), while requiring that the state not override the individual's rational reflective rejection of those goods. In other words, conditional objectivism absorbs Conly's core concern—respect for the individual's deep judgments—within an objectivist framework, without thereby sacrificing the theoretical advantages of the objective list theory.

More precisely, what the negative endorsement constraint excludes is only a specific subset of the fourth attitude-type—cases where the individual not only holds a rejecting attitude, but where this rejection is rational. In the first three attitude-types, whether the individual actively endorses, passively endorses, or holds a neutral stance, she has not rationally rejected the good, and so from the perspective of the negative condition it may be regarded as enhancing her welfare. Even within the fourth type, if the reasons for rejection are irrational—as in Samantha's rejection based on 'a shallow and silly aesthetic theory,' or a Stoic Samantha's rejection grounded in a mistaken philosophical belief—the negative condition is likewise satisfied, and the good still enhances her welfare. Only when a person, having fully acquainted herself with a given objective good, through sincere reflection still judges that it has no significance for her life—not out of ignorance, error, or psychological distortion, but because her deep understanding of what constitutes a meaningful life is incompatible with that good—does the negative condition fail to be met, and the state loses its principled justification for intervening on the sole ground that 'this is on the objective list'.

Against such a proposal, Arneson might object that in practice the negative endorsement constraint yields virtually the same verdicts as pure objectivism. For in the vast majority of cases, rational reflection will lead a person to recognize the value of objective goods—health, knowledge, friendship, meaningful achievement, and so on. A rational person is unlikely, upon full acquaintance, to regard any of these as valueless for her. If the two theories yield highly convergent practical judgments, the proposal amounts to a formal rather than a substantive revision. This objection has some force in the majority of everyday cases. For goods such as health, basic education, and personal safety, no one will reject them after rational reflection, and the negative endorsement constraint would not modify the policy recommendations of pure objectivism in these domains. But the substantive difference between the two theories emerges precisely in cases involving deep value disagreement. Consider the following scenario: suppose "meaningful intimate relationships" is a good on the objective list (which most objectivists would accept), and a state, on the basis of this entry, promotes a specific model of intimate relationship—for instance, a nuclear family structure policy that imposes institutional disadvantages on citizens who choose to live alone, in non-marital partnerships, or in other relationship models. On pure objectivism, if the nuclear family structure is indeed the best way to realize "meaningful intimate relationships", then state intervention is defensible in principle, and the individual's rejection presents no principled obstacle. On conditional objectivism, however, if a person, having fully acquainted herself with the content and potential benefits of various intimate relationship models, through sincere reflection still judges that the nuclear family structure has no significance for her—not out of ignorance or prejudice, but because her deep understanding of what constitutes meaningful intimacy is incompatible with the model promoted by the state—then the state cannot justify intervention on the sole ground that "meaningful intimate relationships are on the objective list." Meaningful intimate relationships is indeed on the list, but its modes of realization are plural, and choosing among these plural modes is precisely a domain where the individual's rational

reflective judgment should not be bypassed. Similar situations may arise with specific aesthetic practices and reproductive choices. In these cases involving an individual's deep understanding of what makes a life meaningful, the two theories yield genuinely different verdicts, and the verdict of conditional objectivism better accords with our moral intuitions.

From a broader theoretical perspective, conditional objectivism occupies an intermediate position between pure objectivism and the strong endorsement constraint. It shares with pure objectivism the core premise that there exist objective goods independent of individual preferences and that individuals can err in their judgments about their own good; but it rejects the overextension of pure objectivism, holding that subjective attitudes retain principled weight in some cases. It shares with the endorsement constraint the core concern that whether an objective good contributes to an individual's well-being depends, to some extent, on that individual's own standpoint; but it rejects the excessive demands of the strong endorsement constraint, which would require active endorsement and desire before a good can count as enhancing one's welfare. The negative endorsement constraint thus occupies a position between two extremes. It holds that whether a good enhances a given individual's well-being does not depend on their positive endorsement of it, but rather on whether they would reject it under conditions of rational reflection. Such an intermediate position not only effectively avoids the threat that paternalism poses to individual freedom, but also secures individuals' important interests, without falling into the difficulties that Richard Arneson's theory faces.

Of course, this proposal itself faces questions requiring further answers. The most important is how to delineate the boundary between "rational rejection" and "irrational rejection". Samantha's rejection based on "a shallow and silly aesthetic theory" is clearly irrational; a person's rejection grounded in a deep worldview held after sincere reflection is clearly rational; but between these two extremes lies a vast gray area. Since the entire theoretical efficacy of the negative endorsement constraint depends on this distinction, it is necessary to provide at least a preliminary conceptual framework.

We propose that the "rationality" in the negative endorsement constraint should be understood as procedural rationality rather than substantive rationality. Procedural rationality requires only that the individual's rejection is not based on informational errors, cognitive distortions, or irrational psychological mechanisms (such as adaptive preferences, severe psychological defense mechanisms, or unreflectively accepted prejudices), but is a judgment made through sincere reflection after these interfering factors have been removed. Substantive rationality, by contrast, would further require that the individual's rejection be compatible with some specific class of reasons—for instance, requiring that the rejection be grounded in "correct" values or a "reasonable" philosophy of life. The negative endorsement constraint adopts a procedural rather than substantive standard because a substantive standard would require an independent theory of "what constitutes correct reasons for rejection", which would effectively return to the path of the objective list theory unilaterally determining the individual's good, thereby dissolving the independent theoretical function of the negative endorsement constraint. The procedural standard avoids this difficulty: it does not judge whether the individual's reasons for rejection are "correct", only whether those reasons are cognitively lucid, informationally adequate, and psychologically undistorted.

This procedural rationality standard bears structural similarities to the ideal-advisor models in contemporary welfare philosophy, but also exhibits a crucial difference. Peter Railton proposed that an individual's good consists in "what he would want himself to want, or to pursue, were he to contemplate his present situation from a standpoint fully and vividly informed about himself and his circumstances, and entirely free of cognitive error or lapses of instrumental rationality" [22] (p. 16). Richard Brandt proposed a "cognitive

psychotherapy" standard, holding that rational desires are those that survive correction by full information [15] (chs. 6–7). Both models employ procedural rationality, which they share with the negative endorsement constraint. But the crucial difference lies in functional direction: Railton's and Brandt's models are positive and constructive—they attempt to identify what is good for the individual by appealing to idealized desire or to the conditions that would ground such desire. The negative endorsement constraint is negative and filtering: it does not attempt to determine what is good (that is the objective list's work), but merely sets a checkpoint of rational rejection between the objective list and the individual. This difference in functional direction means that the negative endorsement constraint's demand on rationality is more restrained than that of ideal-advisor models: it does not need to construct a complete "idealized self" to determine all of an individual's goods, but only needs to judge whether the individual's rejection of a specific good withstands cognitive-level scrutiny.

Of course, even with a procedural rationality standard, the indeterminacy of borderline cases persists. For instance, if a person, on the basis of a non-mainstream but not obviously absurd philosophical tradition (such as thoroughgoing hedonism or a form of nihilism), rejects the value of friendship, achievement, or knowledge, whether this rejection withstands the scrutiny of procedural rationality depends on how we judge whether the philosophical tradition itself contains elements sufficient to constitute "cognitive distortion". But this indeterminacy is not unique to the negative endorsement constraint. Brandt's "cognitive psychotherapy" standard and Railton's "full information" standard face analogous boundary problems: how "full" must "full information" be? These models are still regarded as useful theoretical tools despite their inability to provide precise verdicts in borderline cases. The negative endorsement constraint is similar: it provides clear guidance in core cases (Samantha's irrational rejection does not constitute an obstacle; a reflective person's rational rejection does constitute one), while leaving room for context-specific judgment in the gray area.

7. Conclusions

This paper has traced Arneson's turn from subjectivism to pure objectivism in his theory of welfare and analyzed the theoretical consequences of this turn for the problem of paternalism. Arneson abandoned subjectivist welfare theory due to the problems of adaptive preferences, false beliefs, and non-prudential desires, and adopted an objective list theory instead. His systematic rejection of all versions of the endorsement constraint pushed this objectivist turn to its extreme. Through the analysis of this paper, we can see that this objectivism, so thoroughgoing at the level of value theory, makes concessions and compromises at the level of political philosophy. The reasons Arneson adduces for why paternalistic intervention must remain restrained—the intrinsic value of autonomy, context-sensitive evaluation from the standpoint of a reasonable agent, and the irreducible subjective dimension of dignity within stigma costs—each in substance reintroduces subjective attitudes into the consideration of welfare, demonstrating that these reasonable political-philosophical positions cannot be coherently derived from a purely objectivist value theory that rejects all forms of the endorsement constraint. Only by adjusting pure objectivism into a "conditional objectivism"—acknowledging that rational reflective endorsement occupies an indispensable place in the process by which objective goods become a particular individual's welfare, and establishing a negative endorsement constraint—can the theory achieve coherence.

Finally, it is worth briefly discussing the implications of conditional objectivism for Arneson's distributive justice project. If the negative endorsement constraint is adopted to revise the definition of "welfare", the core change is that when assessing whether an

individual has equal opportunity for welfare, one cannot merely look at whether the goods on the objective list are accessible to her; one must also consider whether she would rationally reject those goods under conditions of reflective deliberation. A person who, after thorough reflection, rationally rejects a certain model of intimate relationship and a person who misses out on meaningful intimate relationships due to ignorance should be treated differently in the assessment of distributive justice: the former is not thereby at a welfare-opportunity disadvantage (because the good in question does not constitute welfare for her), while the latter is indeed at a disadvantage (because her loss stems from insufficient cognitive conditions rather than rational rejection). This revision enables Arneson's distributive justice project to maintain its welfarist character while avoiding undue neglect of reasonable value pluralism.

Author Contributions: Conceptualization, C.X. and X.M.; methodology, C.X.; formal analysis, C.X.; investigation, C.X.; resources, X.M.; writing—original draft preparation, C.X.; writing—review and editing, C.X. and X.M.; supervision, X.M.; project administration, X.M.; funding acquisition, X.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Innovation Research 2035 Pilot Plan of Southwest University, grant number SWUPilotPlan018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

Conflicts of Interest: The authors declare no conflicts of interest.

Notes

1 Dworkin's "expensive tastes" argument imagines a society that has achieved equality of welfare, in which someone deliberately cultivates an expensive taste he did not previously have—say, a taste for plovers' eggs and pre-phylloxera claret. Once cultivated, his welfare level will fall below its former level unless he obtains more wealth. If justice requires equality of welfare, society must give this person extra resources to compensate for the welfare decline caused by his expensive taste. But this is intuitively absurd—why should someone who chooses to cultivate a luxury taste demand that others foot the bill? Dworkin uses this case to show that equality of welfare is mistaken [2] (pp. 48–59).

2 This concept was first proposed by Jon Elster in his essay "Sour Grapes: Utilitarianism and the Genesis of Wants" (in Utilitarianism and Beyond, Amartya Sen and Bernard Williams eds., Cambridge University Press, 1982, p. 219), and later systematically elaborated in his monograph Sour Grapes: Studies in the Subversion of Rationality (Cambridge University Press, 1983) [3,4]. The core idea is that, just as the fox in Aesop's fable declares the grapes sour because he cannot reach them, people in straitened circumstances often alleviate the frustration of unfulfilled desires by reducing their aspirations.

3 In contemporary welfare philosophy, Derek Parfit proposed a classic tripartite taxonomy of welfare theories, dividing them broadly into hedonism, desire-fulfillment theories, and objective list theories. This classification provided the theoretical coordinates for Arneson's turn [9] (p. 493).

4 For further discussion of such views, see also Richard B. Brandt, A Theory of the Good and the Right, Oxford: Clarendon Press, 1979, chs. 6 and 7 [15].

5 It is worth noting, however, that Mill's own position was not as absolute as this declaration suggests. In Chapter V of the same work, he discussed a case involving a dangerous bridge: if a bridge has been ascertained to be unsafe, forcibly stopping a person from crossing it does not constitute a real infringement of his liberty, because "liberty consists in doing what one desires, and he does not desire to fall into the river"; but when there is not a certainty but only a danger of mischief, then the person "ought, I conceive, to be only warned of the danger; not forcibly prevented from exposing himself to it" [17] (p. 158).

References

1. Arneson, R.J. Equality and Equal Opportunity for Welfare. *Philos. Stud.* **1989**, *56*, 77–93. [CrossRef]
2. Dworkin, R. *Sovereign Virtue: The Theory and Practice of Equality*; Harvard University Press: Cambridge, MA, USA, 2000.

3. Elster, J. Sour Grapes: Utilitarianism and the Genesis of Wants. In *Utilitarianism and Beyond*; Sen, A., Williams, B., Eds.; Cambridge University Press: Cambridge, UK, 1982; pp. 219–238.
4. Elster, J. *Sour Grapes: Studies in the Subversion of Rationality*; Cambridge University Press: Cambridge, UK, 1983.
5. Sen, A. Well-Being, Agency and Freedom: The Dewey Lectures 1984. *J. Philos.* **1985**, *82*, 169–221. [CrossRef]
6. Arneson, R.J. Liberalism, Distributive Subjectivism, and Equal Opportunity for Welfare. *Philos. Public Aff.* **1990**, *19*, 158–194.
7. Rawls, J. *A Theory of Justice, Revised Edition*; Harvard University Press: Cambridge, MA, USA, 1999.
8. Arneson, R.J. Human Flourishing versus Desire Satisfaction. *Soc. Philos. Policy* **1999**, *16*, 113–142. [CrossRef]
9. Parfit, D. *Reasons and Persons*; Oxford University Press: Oxford, UK, 1984.
10. Sumner, L.W. *Welfare, Happiness, and Ethics*; Oxford University Press: Oxford, UK, 1996.
11. Kymlicka, W. *Liberalism, Community, and Culture*; Clarendon Press: Oxford, UK, 1989.
12. Hurka, T. Indirect Perfectionism: Kymlicka on Liberal Neutrality. *J. Political Philos.* **1995**, *3*, 36–57. [CrossRef]
13. Dworkin, R. Foundations of Liberal Equality. In *The Tanner Lectures on Human Values*; Peterson, G.B., Ed.; University of Utah Press: Salt Lake City, UT, USA, 1990; Volume 11, pp. 1–119.
14. Griffin, J. *Well-Being: Its Meaning, Measurement, and Moral Importance*; Oxford University Press: Oxford, UK, 1986.
15. Brandt, R.B. *A Theory of the Good and the Right*; Clarendon Press: Oxford, UK, 1979.
16. Arneson, R.J. Joel Feinberg and the Justification of Hard Paternalism. *Leg. Theory* **2005**, *11*, 259–284. [CrossRef]
17. Mill, J.S. *On Liberty*; Bromwich, D., Kateb, G., Eds.; Yale University Press: New Haven, CT, USA, 2003.
18. Feinberg, J. *Harm to Self (The Moral Limits of the Criminal Law)*; Oxford University Press: Oxford, UK, 1986; Volume 3.
19. Arneson, R.J. Egalitarian Perspectives on Paternalism. In *The Routledge Handbook of the Philosophy of Paternalism*; Grill, K., Hanna, J., Eds.; Routledge: London, UK, 2018; pp. 194–205.
20. Raz, J. *The Morality of Freedom*; Clarendon Press: Oxford, UK, 1986.
21. Conly, S. *Against Autonomy: Justifying Coercive Paternalism*; Cambridge University Press: Cambridge, UK, 2013.
22. Railton, P. Facts and Values. *Philos. Top.* **1986**, *14*, 5–31. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.