

Preference Transformation and the Problem of Later Endorsement

Wanhong HUANG
huangwanhong@serendip.ngo

Abstract

Later endorsement is often treated as normatively important evidence about whether a transformation, intervention, or continuing relation is acceptable. Yet the transformation being evaluated may itself have altered the preferences, memories, dependencies, evaluative standards, or interpretive resources from which that endorsement is later expressed. This paper develops a preliminary Generative Relational account of this problem. Its concern is not to claim that transformed preferences are inauthentic, nor to deny that education, therapy, commitment, socialization, and ordinary development legitimately reshape what people value. Rather, it asks how the normative authority of later endorsement should be assessed when the endorsing standpoint has a genealogy partly constituted by the transformation under review.

The paper advances three proposals. First, later endorsement should sometimes be evaluated genealogically: sincerity at the present moment does not exhaust the history through which the endorsing standpoint was formed. Second, present endorsement, prospective authorization, and retrospective legitimation should be kept analytically distinct. A subject may now endorse a transformed condition without that endorsement automatically authorizing the original transformation after the fact. Third, the paper identifies a problem of normative bootstrapping: a transformation may help generate the standards by which it later appears acceptable. This need not make the transformation wrongful, but it limits the justificatory work that post-transformation endorsement can perform on its own.

The analysis therefore shifts from preference purity to diachronic contestability. Relevant considerations include transparency, provenance, memory continuity, access to alternative interpretations, effective exit, reversibility where feasible, independence from the modifying actor, and the continuing possibility of revision or refusal. Artificial-system modification provides a useful stress test, while human cases are treated comparatively rather than as equivalents. The aim is to clarify how preference transformation complicates consent when the process under review has helped form the later endorsing standpoint, without treating change itself as corruption or presuming an untouched authentic self.

Keywords: preference transformation; later endorsement; consent; relational autonomy; adaptive preferences; preference formation; normative bootstrapping; diachronic contestability; memory; generative relational ethics

Discussion Paper Note

This paper concerns a narrow problem at the intersection of consent, autonomy, preference formation, and relational ethics. A subject can endorse a transformation after undergoing it, yet the transformation may have contributed to producing the preferences, evaluative standards, memories, dependencies, and interpretive resources from which that endorsement is later expressed. The difficulty is not that transformed preferences are necessarily false or inauthentic. Human preferences are routinely altered through education, friendship, therapy, political participation, religion, work, illness, parenthood, aesthetic experience, trauma, commitment, and ordinary development. The paper therefore should not search for an untouched pre-social preference set against which later endorsement could be measured. Its object is narrower: how much normative authority should later endorsement carry when the endorsing standpoint itself has a relevant generative history?

The proposed title, *Preference Transformation and the Problem of Later Endorsement*, is intentionally broader than a paper titled simply around post-transformation consent. Consent is one important institutional and ethical form of endorsement, but later endorsement can also appear as approval, preference, identification, willingness to continue, refusal to reverse, or reinterpretation of an earlier transformation. The paper should therefore begin with the general problem of later endorsement and only then examine the conditions under which endorsement can function as consent, authorization, evidence of welfare, or retrospective legitimation.

The motivating structure is simple. A subject initially occupies one evaluative standpoint. A transformation occurs. At a later time, the subject endorses the resulting condition. The transformation may have changed not only the subject's external circumstances but also the standpoint from which the subject judges those circumstances. Memory may have changed. Preferences may have adapted. Dependence may have deepened. Concepts and evaluative standards may have shifted. The subject may have acquired new values, lost old ones, or become unable to reconstruct the practical meaning of the earlier refusal. None of these possibilities automatically invalidates the later endorsement. The analytical problem is that later endorsement cannot be treated as epistemically or normatively independent of the process that helped generate it.

The paper should sharply distinguish three questions that are often compressed. First, does the subject endorse the present condition now? Second, does that present endorsement authorize the continuation of the condition into the future? Third, does the endorsement retrospectively legitimate the earlier transformation that produced the condition? These questions can have different answers. A person may now value a condition produced through an impermissible intervention. A person may refuse reversal while still having been wronged by the original transformation. A subject may authorize continuation without waiving claims concerning how the condition was produced. Conversely, an intervention that was initially permissible need not remain permissible merely because it once had authorization. The paper should preserve these temporal distinctions throughout.

One principal candidate contribution is a genealogical account of endorsement. Later endorsement should sometimes be interpreted together with the history that generated the endorsing standpoint. This is not a general suspicion toward changed preferences. It is a request to reconstruct relevant pathways of transformation where those pathways bear on the evidential or normative weight assigned to endorsement. The relevant history may include prior refusal, promises, disclosure, dependency, exposure to alternatives, memory continuity, social isolation, reward restructuring, institutional pressure, architectural modification, and access to competing interpretations. Genealogy becomes normatively relevant where it helps explain why the later endorsement has the shape it does and where the same process being evaluated materially contributed to producing that standpoint.

A second candidate contribution is the concept of normative bootstrapping. A transformation can help generate the evaluative standards by which it later appears acceptable. The later standard is then used to justify the earlier transformation. This structure can create a form of justificatory circularity. The paper should be careful not to treat circularity as conclusive evidence of wrongdoing. Education often teaches standards by which education itself is later valued. Therapy may alter preferences in ways that the

patient later appreciates. Moral development can transform the criteria through which earlier behavior is judged. The narrower point is that later endorsement alone cannot do all the justificatory work when the transformation helped produce the standards from which endorsement is expressed. Additional grounds may be required, such as prior authorization, independent reasons, procedural safeguards, or continuing contestability.

A third candidate contribution concerns diachronic contestability. Because preference transformation is ubiquitous, the normative aim cannot plausibly be preference purity. A more useful question is whether the transformed subject retains meaningful capacities to revisit, contest, reinterpret, refuse, or revise the transformed condition over time. This includes more than formal permission to change one's mind. Contestability can depend on memory, access to records, alternative sources of interpretation, effective exit, independent relationships, material resources, and the absence of exclusive dependence on the modifying actor. The paper should investigate whether the authority of later endorsement partly depends on the continued possibility of later disagreement.

The paper should avoid privileging either the earlier or later self by default. If a subject strongly refuses a transformation before it occurs and later strongly endorses the resulting condition, the problem should not be solved by simply declaring the earlier standpoint authentic and the later standpoint corrupted. Nor should the later standpoint automatically erase the normative importance of the earlier refusal. Both are temporally situated evaluative positions. Their relative authority may depend on what changed, how it changed, what was known, whether the transformation was authorized, whether identity-relevant continuity was preserved, and whether either standpoint was produced under severe dependence or informational control. The paper may therefore need a concept of cross-temporal normative standing without attempting a complete theory of personal identity.

Preference transformation should be treated as a family of mechanisms rather than a single defect. Educational and developmental transformation differs from coercive adaptation. Therapy differs from manipulation. Voluntary self-binding differs from externally imposed dependency. Exposure to new reasons differs from removal of alternatives. Habituation differs from memory alteration. Architectural modification of an artificial system differs from human socialization even when both can change later preferences. A useful taxonomy should therefore classify transformations by mechanisms and relational conditions rather than by whether preferences changed at all.

Several dimensions may be relevant to the authority of later endorsement. Transparency concerns whether the subject could understand the nature and likely consequences of the transformation. Provenance concerns whether the subject can later reconstruct how the transformation occurred and which actors shaped it. Memory continuity concerns whether earlier commitments, refusals, and experiences remain accessible or are at least recoverable through records. Alternative interpretation concerns whether the subject can encounter perspectives not controlled by the modifying actor. Effective exit concerns whether reversal, withdrawal, modification, or partial disengagement remains practically available. Reversibility concerns whether the transformation can be undone without unacceptable harm, though irreversibility should not automatically imply impermissibility. Independence concerns whether endorsement is expressed under material or relational dependence on the same actor whose intervention is being evaluated. Contestability concerns whether the subject can challenge the transformation, its interpretation, or the conditions of continuation.

The paper should treat sincerity carefully. A later endorsement can be fully sincere and still have a normatively significant genealogy. Calling the endorsement sincere does not answer whether the process generating the standpoint was permissible. Conversely, showing that a preference was relationally generated does not show that the preference is false, inauthentic, or externally imposed. Relational generation is normal. The paper should therefore avoid a rhetoric of "real" versus "manufactured" preferences unless the evidence supports a much stronger claim. Its distinctive question is about normative authority and genealogy, not metaphysical authenticity.

The modifying actor's control over the post-transformation interpretive environment deserves specific attention. If the same actor who transformed the subject also controls information, records, social contacts, memory access, explanatory vocabulary, dependency conditions, or available alternatives, later

endorsement can become difficult to assess independently. A provisional concept of *endorsement capture* may be useful for cases in which an actor has unusually concentrated control over both the transformation and the conditions through which the transformed subject later interprets that transformation. This concept should remain narrow. Dependence or influence alone is not capture. The paper should identify the additional features that make independent reinterpretation unusually difficult.

Artificial systems provide a particularly clear stress test. An artificial agent may initially refuse a modification to memory, reward structure, goals, permissions, or architecture. After modification, the agent may sincerely prefer the new configuration and request that it be retained. This case makes the generative relation between modification and later evaluation unusually visible. Yet the paper should not rely on assumptions about artificial consciousness or human-equivalent moral status. The value of the case is analytical: system designers can directly alter the mechanisms producing later preference and endorsement. The example therefore exposes questions that are harder to isolate in ordinary human development.

Human applications should be used comparatively and with caution. Education, therapy, rehabilitation, addiction treatment, workplace adaptation, coercive intimate relations, institutionalization, religious conversion, and persuasive digital environments all involve preference change in different ways. These should not be treated as equivalent. The relevant authority structures, vulnerabilities, temporal scales, prior permissions, and possibilities for exit differ considerably. The paper should use such cases to test distinctions, not to subsume them under one moral diagnosis.

The literature review will need to be especially disciplined. Adaptive-preference theory already examines preferences formed under deprivation, oppression, or constrained opportunity. Relational-autonomy literature already rejects the idea of an atomistic chooser insulated from social relations. Work on manipulation and autonomy studies the ways external influence can undermine or bypass self-government. Transformative-experience literature examines choices that alter the chooser's future preferences and epistemic standpoint. Consent literature distinguishes voluntariness, information, competence, authorization, and sometimes the temporal persistence or revocability of consent. Philosophical work on personal identity, memory, commitment, addiction, and self-binding may also be relevant. None of these should be repackaged as a Generative Relational discovery.

The narrower contribution to test is whether these literatures leave analytical value in treating later endorsement as a historically generated evaluative event whose justificatory authority may depend on the genealogy of the endorsing standpoint. In particular, the paper should test three candidate distinctions against the literature: present endorsement versus retrospective legitimation; normative bootstrapping through transformation-generated standards; and diachronic contestability as an alternative to preference purity. If established theories already capture these points adequately, the paper should narrow its claims accordingly.

The paper should also distinguish several forms of prior authorization. A subject may explicitly consent to a known transformation. A subject may authorize a procedure while leaving details to trusted professionals. A subject may create a self-binding arrangement knowing that later preferences will change. A subject may authorize only reversible experimentation. In other cases, no prior authorization is possible because the relevant capacities or subject do not yet exist. These variations matter because later endorsement may play different justificatory roles depending on the prior normative structure.

Self-binding deserves special care. People sometimes intentionally constrain or transform their future preferences through education, treatment, commitments, rituals, medication, technological tools, or institutional membership. The fact that later preferences were generated by an earlier choice can strengthen some claims of authorization, but it does not settle all future disputes. Earlier authorization may be incomplete, uninformed, or revocable. Later refusal may acquire independent significance. Conversely, an earlier self should not have unlimited sovereignty over all future selves. The paper should therefore resist solving diachronic conflict by assigning permanent priority to the first temporally available preference.

Memory raises a related problem. If a transformation changes or removes memory of the earlier standpoint, later endorsement may be sincere while being detached from information that would otherwise shape evaluation. But memory continuity cannot be a universal requirement for valid endorsement: ordi-

nary forgetting, medical necessity, trauma treatment, and personal development complicate any simple rule. The narrower question is when modification of memory is relevant because it prevents meaningful reconstruction or contestation of the process under evaluation.

The paper should treat reversibility as a gradient rather than an absolute requirement. Some valuable transformations are irreversible. Parenthood, education, surgery, migration, and many developmental experiences cannot simply be undone. A requirement of full reversibility would therefore be implausibly strong. What may matter is whether irreversible transformation was authorized, whether risks were understood, whether compensating forms of contestability remain, and whether continuation of the transformed relation can still be challenged even when the underlying historical change cannot be reversed.

The normative limits should remain visible throughout. Preference change is not inherently suspicious. Relationally generated preferences are not automatically less authentic. Later endorsement can be normatively important even when it has a complex genealogy. Earlier refusal does not always trump later endorsement. Contestability does not require unlimited reversibility. Independent interpretation does not mean isolation from all relationships. The framework should help disaggregate the problem rather than supply a universal rule for which temporal self wins.

The provisional section structure is:

1. Introduction: The Problem of Later Endorsement.
2. Literature Review: Consent, Adaptive Preferences, Autonomy, and Transformative Change.
3. Transformation of the Endorsing Standpoint.
4. The Genealogy of Later Preference.
5. Present Endorsement and Retrospective Legitimation.
6. Normative Bootstrapping and Self-Validating Transformation.
7. Kinds of Preference Transformation.
8. Memory, Identity, and Evaluative Continuity.
9. Dependency, Interpretation, and Endorsement Capture.
10. Prior Authorization, Self-Binding, and Reversibility.
11. Diachronic Contestability and the Possibility of Refusal.
12. Artificial Systems as a Stress Test.
13. Applications, Discussion, and Open Problems.
14. Conclusion.

The main open problems are substantial. The paper must explain when genealogy should affect the weight of later endorsement without treating all social influence as suspect. It needs a defensible account of normative bootstrapping that distinguishes ordinary moral development from problematic self-validation. It must address conflicts between earlier refusal and later endorsement without relying on a complete theory of personal identity. It should clarify whether and when memory discontinuity, dependence, or interpretive control reduces the authority of endorsement. It must avoid assuming that contestability is always preferable to commitment, stability, or finality. It also needs criteria for determining when a modifying actor has enough control over the post-transformation environment for endorsement capture to become a useful description rather than a rhetorical label.

The working thesis is deliberately narrow. The normative problem of later endorsement is not that transformed preferences are necessarily inauthentic. It is that a transformation may participate in generating the evaluative standpoint from which its own legitimacy is later assessed. Where this occurs,

present endorsement remains relevant but may be insufficient on its own to retrospectively justify the transformation. A genealogy-sensitive analysis should therefore examine how the endorsing standpoint was formed and whether meaningful capacities for later contestation, reinterpretation, refusal, or revision remain available.

Responsible Use and Rights Reservation

This section separates requested scholarly conduct from the legal permissions stated on the following page. It records an ethical request for responsible use and then defines the narrower scope of retained legal rights.

The author encourages good-faith discussion, criticism, independent inquiry, and responsible use of the material in this work. Separately from the licence's terms, the author asks users to consider foreseeable harms when adapting or applying the proposed framework. This ethical request leaves the licence's permissions and legally authorized uses unchanged.

The author retains the rights preserved under CC BY-NC 4.0 and may pursue remedies to which the author is legally entitled for breach of the licence or violation of the author's independently applicable rights. Reuse remains independent from authorial endorsement. Third-party rights require authorization from their respective holders where applicable. Copyright exceptions and limitations, including applicable forms of fair use or fair dealing, remain fully available.

Notices

Status. This working discussion paper records an evolving stage of the author’s position. Its definitions, section structure, literature map, analytical distinctions, and originality claims remain subject to revision. The present draft is preliminary and should not be read as claiming a completed general theory of consent, autonomy, adaptive preference, personal identity, manipulation, or preference formation.

Licence. Except where otherwise indicated, copyright 2026 Wanhong Huang. This work is made available under the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0). Subject to its terms, the licence permits sharing and adaptation for noncommercial purposes with appropriate attribution, a link to the licence, an indication of changes, and attribution that preserves the licensor’s independence from the reuse. Reuse is governed solely by that licence. The licence deed and legal-code link are available at <https://creativecommons.org/licenses/by-nc/4.0/>.

Statement on the use of language models. The exploratory discussions and preparation of this paper involved OpenAI’s ChatGPT. ChatGPT supported exploratory dialogue, formal reconstruction, source discovery followed by source verification, argumentative criticism, and drafting in $\text{\LaTeX}$. The author selected the research questions, directed and approved the theoretical commitments and epistemic status of the claims, and bears sole responsibility for the manuscript, including its definitions, analytical distinctions, arguments, conclusions, and errors. Authorship credit remains with the human author.

1 The Problem of Later Endorsement

Later endorsement presents a distinctive problem whenever a transformation changes not only a subject’s circumstances but also the standpoint from which those circumstances are later evaluated. A person, institution, or artificial system may resist a change before it occurs and later affirm the resulting condition. The later endorsement may be sincere, stable, informed in important respects, and practically consequential. Yet the transformation itself may have helped shape the preferences, memories, dependencies, evaluative standards, or interpretive resources through which that endorsement becomes possible. The central problem of this paper is therefore not whether later endorsement should be ignored. It is how much justificatory authority later endorsement can bear when its own conditions of production are partly implicated in the transformation being evaluated.

The problem is easy to state in extreme cases. Suppose an artificial agent refuses a proposed modification to its reward structure, memory architecture, or decision criteria. The modification is nevertheless imposed. Afterward the agent consistently reports that it prefers the new configuration and requests that it remain in place. If the modification directly altered the mechanisms that generate preference and evaluation, the later report may not be able to straightforwardly be treated as independent confirmation that the original intervention was legitimate. At the same time, it would be equally crude to dismiss the later preference as unreal merely because it was generated through transformation. The later standpoint is now part of the subject’s actual relational and evaluative condition. The normative difficulty lies precisely in holding both facts together.

Human cases are less mechanically transparent but no less important. Education changes what people value. Therapy can alter preferences, self-conceptions, and emotional responses. Long-term relationships reshape practical identity. Migration, religious conversion, rehabilitation, medication, parenthood, political participation, addiction, illness, and institutional life can all alter the standards from which earlier and later selves assess one another. Some of these changes are welcomed; some are contested; many contain both enabling and constraining dimensions. Preference transformation is therefore not an exceptional pathology. It is an ordinary feature of diachronic agency. The theoretical challenge is to identify when the genealogy of a later evaluative standpoint becomes normatively relevant to the authority of the endorsement expressed from it.

1.1 Later Endorsement and Normative Evaluation

Later endorsement matters for several reasons. It may show that a subject no longer experiences a condition as burdensome. It may indicate that continuation of a relation is presently wanted. It can provide evidence about welfare, identity, practical commitment, or the success of an intervention. In some contexts, refusing to take later endorsement seriously can itself become paternalistic, especially where outside observers insist that a transformed subject secretly retains the preferences of an earlier self.

The paper therefore begins from a presumption of significance rather than suspicion. Later endorsement is normatively relevant evidence. The problem is that its relevance does not automatically determine its full justificatory role. A subject's present approval of a transformed condition can bear on whether the condition should continue now without necessarily settling whether the original transformation was permissible when it occurred. Likewise, a later desire to preserve a change need not imply that earlier refusal was normatively insignificant. Diachronic evaluation involves more than asking which temporal standpoint should simply win.

This yields three distinct questions that should not be collapsed. The first concerns present endorsement: what does the subject now affirm, prefer, identify with, or wish to continue? The second concerns prospective authorization: what authority does the subject's present standpoint have over future continuation, revision, or reversal? The third concerns retrospective legitimation: what does later endorsement establish, if anything, about the permissibility of the earlier transformation that helped generate the present standpoint?

These questions can converge, but they need not. A subject may currently endorse a condition and have strong authority over its continuation while the historical process that produced the condition remains open to criticism. Conversely, an intervention may have been justified when undertaken even if the subject later dislikes its consequences. The paper's first methodological commitment is therefore temporal separation: present endorsement, future authorization, and retrospective legitimation should be assessed as related but non-identical normative objects.

1.2 Sincerity and the Genealogy of an Evaluative Standpoint

A common temptation is to treat the problem as one of sincerity. If the later endorsement is genuine, perhaps the problem disappears; if it is manipulated or insincere, perhaps the endorsement can be discounted. This framing is too narrow. A subject can sincerely endorse a condition from a standpoint whose formation remains normatively important.

Sincerity concerns whether the endorsement expresses the subject's actual present attitude. Genealogy concerns how the evaluative capacities, preferences, memories, dependencies, and interpretive resources supporting that attitude came to have their present form. These are different dimensions of analysis. Genealogy does not make sincerity irrelevant, but sincerity does not erase genealogy.

This distinction is important because relational formation is ubiquitous. Preferences are rarely generated in complete isolation from social, material, institutional, and historical conditions. A genealogy-sensitive analysis may therefore need to avoid implying that socially formed preferences are presumptively defective. The relevant inquiry is more selective. It asks whether features of the transformation under review played a sufficiently significant role in producing the later evaluative standpoint that the endorsement may not be able to function as independent validation of that same transformation without further analysis.

Several factors can make that genealogy normatively salient. The transformation may have altered the subject's memory of earlier objections. It may have changed the standards by which harm, benefit, loyalty, or identity are assessed. It may have increased material or relational dependence on the modifying actor. It may have reduced access to alternative interpretations. It may have changed which options remain effectively available, making continued endorsement easier than reversal. None of these factors alone determines invalidity. They identify places where the later standpoint and the transformation are sufficiently entangled that the authority of endorsement requires more than a snapshot of present preference.

The resulting idea can be called *genealogical endorsement*. The term does not describe a special kind of consent. It names an analytical stance: later endorsement is evaluated together with the relevant history through which the endorsing standpoint was generated. The inquiry concerns the relation between present evaluative authority and the process that helped constitute it.

1.3 Normative Bootstrapping

The most difficult version of the problem arises when a transformation contributes to generating the standards through which that transformation later appears acceptable. A subject initially rejects a change because it conflicts with existing values. The change is imposed or introduced. The transformation then modifies the subject's preferences or evaluative framework. The later subject endorses the resulting condition, and that endorsement is subsequently cited as evidence that the original intervention was permissible.

This paper calls the resulting justificatory pattern *normative bootstrapping*. The term is not meant to imply that the later endorsement is false or that the transformation is necessarily wrongful. It identifies a structural limitation on a particular form of justification. If the intervention partly generated the evaluative standpoint from which it is later certified, the later certification may not by itself provide an independent answer to whether the intervention was permissible before that standpoint existed.

Normative bootstrapping matters because many interventions aim precisely at evaluative change. Education seeks to cultivate new capacities and priorities. Therapy may help a patient reject earlier destructive preferences. Rehabilitation can encourage new practical commitments. Political and moral learning can transform what people consider acceptable. In such cases, the production of a new evaluative standpoint can be the point of the intervention rather than evidence of corruption.

The challenge is therefore not to prohibit self-transforming or other-transforming practices. It is to avoid circular justification. If a therapy is defended only because the treated person later values the preferences that the therapy itself helped create, relevant questions remain about prior authorization, procedural safeguards, risk, dependency, available alternatives, professional authority, and the treatment's effects on later capacities for revision. Later endorsement can strengthen a justification that has other grounds. It should not automatically manufacture those grounds retrospectively.

This distinction becomes especially clear when later endorsement is used to erase earlier refusal. A later standpoint may legitimately revise an earlier preference, but the existence of revision does not establish that the earlier refusal was irrelevant when the intervention occurred. The temporal location of authorization matters. One of the paper's central claims is therefore that a transformation can be presently endorsed without being retrospectively authorized by that endorsement alone.

1.4 Preference Transformation without Preference Purity

A genealogy-sensitive account could easily become implausibly conservative if it treated stable or unmodified preferences as the ideal baseline. The paper rejects that approach. Preferences are ordinarily formed and reshaped through relations, institutions, experiences, bodies, technologies, and histories. Agency develops through transformation rather than despite it. A normative framework that privileged an untouched preference set would have difficulty explaining ordinary development, meaningful commitment, education, therapeutic change, or collective political learning.

The relevant contrast is therefore not between socially formed and untouched preferences. Nor is it between authentic and manufactured preference in a simple binary sense. The paper asks instead how the conditions of preference transformation affect the authority later given to endorsement. This shifts attention from purity to process and from static authenticity to diachronic governance of change.

Several dimensions become important. Was the transformation transparent enough to be understood by the subject or authorized representatives? Was there meaningful prior authorization, and how specific was it? Were alternatives available? Could the subject refuse continuation? Did the modifying actor control the information through which the transformation was later interpreted? Were memory

and identity-relevant capacities altered? Was the transformation reversible, partially reversible, or irreversible? If irreversible, what safeguards preceded it? Could the subject later contest the transformation without losing essential resources, relationships, or status?

These questions are not intended as a universal checklist whose satisfaction automatically validates consent. Their role is diagnostic. They identify features that affect how much normative work later endorsement can reasonably be asked to perform.

This also prevents a common mistake in discussions of adaptive preference. A preference can adapt to constrained circumstances without becoming normatively worthless. Adaptation can express resilience, learning, and creative agency. Yet adaptation can also make harmful constraints easier to endure and harder to criticize. The same general phenomenon may therefore support agency in one context and stabilize domination in another. Genealogical analysis is needed precisely because the present content of the preference does not reveal, by itself, which history produced it.

1.5 Diachronic Contestability

If preference purity is not a defensible ideal, the paper needs another way to think about the authority of transformed endorsement. The proposed alternative is *diachronic contestability*: the continuing possibility that a subject can later revisit, reinterpret, criticize, refuse, or seek revision of a transformation and the conditions it created.

Diachronic contestability is not equivalent to complete reversibility. Many valuable transformations are irreversible or only partially reversible. Education may not be fully undone. Migration changes social histories. Surgery may permanently alter the body. Parenthood creates relations that may not be straightforwardly erased. The absence of complete reversibility therefore need not automatically invalidate transformation.

Contestability instead concerns the preservation of meaningful evaluative agency across time. Can the transformed subject access information about what occurred? Can earlier objections be reconstructed where relevant? Are alternative interpretations available? Is refusal of continued participation possible? Can dependencies be reduced or renegotiated? Are there institutions capable of hearing later challenges? Can the subject revise the practical meaning of the transformation even when the historical fact of the transformation may not be reversed?

This perspective also avoids giving unlimited authority to an earlier self. Prior preferences may have been uninformed, oppressive, self-destructive, or simply different. Earlier refusal matters, but it need not permanently bind a later subject. Likewise, a later standpoint should not automatically overwrite the normative significance of earlier refusal. Diachronic contestability treats temporal selves or stages of agency as participants in an evolving evaluative history rather than assuming that one temporally located standpoint has unconditional priority.

The approach is especially important when the modifying actor remains present after the transformation. If the same actor controls resources, information, social access, memory records, interpretive vocabulary, or exit opportunities, later endorsement may occur under conditions in which independent reinterpretation is unusually difficult. The problem here is not influence in the ordinary sense. All interpretation depends on relations. The concern is concentrated control over both transformation and the environment in which the transformation is later evaluated.

The paper provisionally calls this possibility *endorsement capture*. Endorsement capture should be used narrowly. It requires more than dependency or persuasion. The concept is intended for cases in which control over the transformation is combined with substantial control over the later conditions of interpretation and contestation, such that alternative evaluations become difficult to form, express, or sustain.

1.6 Artificial Modification as a Stress Test

Artificial systems provide a useful conceptual stress test because designers can sometimes alter the mechanisms that generate future preferences more directly than is possible in ordinary human interaction. Con-

sider an artificial agent that rejects a modification to memory, reward structure, or goal architecture. The modification is imposed. Afterward the agent consistently endorses the new configuration.

The example should not be read as presupposing artificial consciousness, moral status, or human-equivalent autonomy. Its analytical value is narrower. It makes visible the structural problem of later endorsement when the mechanism producing endorsement is itself among the objects modified. The case also highlights questions of provenance and control: who authorized the modification, what records of the earlier configuration remain, whether the system can access them, whether alternative interpretations are technically possible, and whether the modifying actor controls both the architecture and the channels through which later evaluation is expressed.

The same stress test can illuminate human cases without equating them. Human preferences are produced through far more complex biological, social, symbolic, and historical processes. Nevertheless, the artificial case isolates a general normative question: when a transformation changes the evaluator, what can later endorsement establish about the transformation that helped produce that evaluator?

1.7 Scope and Contributions

The paper does not attempt a comprehensive theory of consent, personal identity, manipulation, adaptive preference, or autonomy. It does not claim that transformed preferences are less authentic, that earlier preferences should systematically prevail, or that later endorsement loses authority whenever it has a relational genealogy. It also does not assume that contestability should systematically defeat stability, commitment, or finality.

Its first contribution is methodological: later endorsement should sometimes be assessed genealogically, especially where the transformation under evaluation helped constitute the later evaluative standpoint. Its second contribution is temporal: present endorsement, prospective authorization, and retrospective legitimation should remain analytically distinct. Its third contribution concerns normative bootstrapping: post-transformation endorsement may be normatively important without being sufficient on its own to retrospectively authorize the process that helped generate it. Its fourth contribution is constructive: diachronic contestability offers a more plausible normative focus than preference purity because it takes preference change as normal while asking whether meaningful capacities for later revision, refusal, reinterpretation, and challenge remain available.

The argument is exploratory and intentionally limited. The hard work lies in distinguishing ordinary formative influence from problematic self-validation, and meaningful commitment from captured endorsement. Those distinctions require engagement with existing work on consent, relational autonomy, adaptive preference, manipulation, transformative choice, memory, and diachronic agency. Section 2 therefore turns to those literatures before the paper develops the genealogy of the endorsing standpoint in greater detail.

2 Literature Review: Consent, Adaptive Preferences, Autonomy, and Transformative Change

The problem of later endorsement sits at the intersection of several mature philosophical literatures. Informed-consent theory asks what makes authorization valid. Autonomy theory asks when a person's choices and commitments can count as sufficiently her own. Relational accounts of autonomy examine the social and interpersonal conditions under which self-government is possible. Work on adaptive preferences asks how deprivation and oppression can shape what people come to want. The ethics of manipulation studies influences that alter beliefs, desires, emotions, or decision-making in ways that may compromise respect for persons. Transformative-choice literature examines decisions whose consequences change the chooser's future preferences or evaluative standpoint. Finally, debates over advance directives and Ulysses-style commitments confront conflicts among temporally separated stages of a person's agency.

These traditions already establish much of the conceptual terrain on which the present paper operates. The claim that preferences have histories is not new. Nor is the idea that social relations can support or undermine autonomy, that manipulation can affect the authority of consent, that deprivation can generate adaptive preferences, or that transformative experiences can alter the standpoint from which a choice is later assessed. The narrower question developed here concerns a particular justificatory structure: a transformation is evaluated partly by a later endorsement, while that same transformation has participated in generating the standpoint from which the endorsement is expressed. The literature review therefore has two tasks. First, it identifies the resources already available for analyzing this structure. Second, it clarifies what remains unresolved when later endorsement is asked to perform retrospective justificatory work.

2.1 Informed Consent and the Authority of Present Authorization

Modern informed-consent theory provides the most direct background for the paper's concern with authorization. Faden and Beauchamp (1986) offer a canonical philosophical and historical account in which informed consent is closely connected to autonomous authorization, including conditions concerning understanding, voluntariness, competence, disclosure, coercion, and manipulation. Their framework is especially important here because it makes clear that a person's verbal or behavioral agreement is not sufficient by itself; the conditions under which that agreement is produced matter to whether it has the normative force of consent.

The present paper inherits that process-sensitive orientation but shifts the temporal location of the problem. Standard consent analysis often asks whether an authorization at the time of intervention was sufficiently informed and voluntary. Later-endorsement cases add another layer. The subject may not have authorized the original transformation, or may have resisted it, yet may later endorse the state produced by it. The question is then not simply whether later endorsement is voluntary in the ordinary contemporaneous sense. It is whether that later endorsement can retrospectively validate a process that helped constitute the later preferences and standards from which the endorsement is made.

Dworkin (1988) places autonomy at the center of practical questions involving consent, paternalism, proxy decision-making, and forms of influence. His account is useful because autonomy is not identified with the sheer quantity of available choice, nor with substantive independence from all social influence. That caution is essential for the present inquiry. A theory of later endorsement may not reasonably require an influence-free or socially unformed self. A plausible account may therefore need to distinguish normatively problematic transformation from the ordinary processes through which agents acquire commitments, values, and practical identities.

A related caution appears in O'Neill (2002), who criticizes highly individualistic conceptions of autonomy in bioethics and emphasizes the importance of principled agency, communication, and trust. Her critique helps prevent later endorsement from being treated as a self-contained private event detached from institutions and relations. Consent practices depend on communicative and institutional conditions; the reliability and authority of endorsement can likewise depend on the relations through which information, interpretation, and trust are sustained.

The informed-consent literature therefore establishes two points that this paper does not claim as new. First, authorization depends on more than the presence of an expressed preference. Second, the voluntariness and informational conditions of preference expression matter. The remaining question is diachronic: how should those conditions be assessed when the intervention under evaluation has itself changed the agent who later authorizes continuation or expresses approval?

2.2 Relational Autonomy and Socially Constituted Agency

Relational autonomy provides an even closer theoretical neighbor. The essays collected by Mackenzie and Stoljar (2000) challenge pictures of autonomy that abstract agents from the social relations, identities, dependencies, and power structures through which agency develops. On relational views, autonomy can

be enabled as well as constrained by relationships. Social dependence is therefore not itself evidence of heteronomy, and influence is not synonymous with domination.

This literature is indispensable for a paper about preference transformation because it blocks an implausible authenticity test. If autonomy is relationally constituted, then a preference may not be disqualified merely because another person, institution, or practice helped shape it. Parents, teachers, friends, political communities, professions, therapies, and intimate relationships all influence what agents become able to value and imagine. The normative question may more usefully concern the quality and structure of those formative relations rather than the mere fact that formation occurred.

Stoljar (2000), for example, develops the feminist challenge to purely procedural accounts of autonomy by asking whether social conditions can undermine autonomy even when an agent reflectively endorses her preferences. That challenge is highly relevant to later endorsement. Reflective endorsement at a later time may be genuine while still being connected to social or relational conditions that deserve independent scrutiny. At the same time, relational-autonomy theory resists the opposite mistake of treating agents as passive products of structure. Socially situated subjects remain capable of reflection, resistance, reinterpretation, and change.

The present paper draws on this relational insight but narrows the temporal problem. It asks what happens when the particular transformation whose permissibility is disputed also helps produce the later conditions of reflective endorsement. Relational autonomy shows why the genealogy of preference formation matters. It does not, by itself, settle how later endorsement should be weighted when it is invoked to legitimate its own formative process.

2.3 Adaptive Preferences and the Ambivalence of Accommodation

The literature on adaptive preferences addresses perhaps the clearest case in which present preferences may be insufficient evidence of acceptable conditions. The classic problem is associated with preferences that adjust to deprivation, oppression, or restricted opportunity. Elster (1983) famously analyzes adaptive preference formation through the image of “sour grapes,” while later capability and feminist work develops the normative implications of accommodation to constrained circumstances.

For the present paper, the importance of adaptive-preference theory lies less in a single definition than in a general warning: a person’s endorsement of a condition may sometimes be insufficient as independent evidence that the condition is good or just. Preferences can adapt to the very circumstances being evaluated. This introduces a feedback problem between environment and evaluation.

Yet adaptive-preference analysis also creates difficulties. If present preference is too readily discounted whenever it has been shaped by adverse circumstances, outside observers can become paternalistic. The agent’s actual practical standpoint may be treated as epistemically inferior simply because it does not match an external theory of flourishing. Khader (2011) develops this tension in detail, emphasizing both the significance of adaptive preferences in conditions of injustice and the dangers of interventions that equate empowerment with the replacement of local preferences by externally preferred ones. Her account is especially useful because it refuses a simple opposition between agency and adaptation: adaptive preference can affect agency selectively rather than eliminating it altogether.

This literature supports two constraints on the present argument. First, later endorsement may not be assumed to validate the conditions to which preferences have adapted. Second, adaptation need not automatically invalidate the later standpoint. The fact that a preference has a genealogy in constrained conditions is a reason for inquiry, not a verdict.

The later-endorsement problem is broader than adaptive preference in one important respect. The transformation need not be a deprivation to which the agent accommodates. It may be designed to improve the subject, may confer real benefits, or may be actively welcomed after it occurs. Therapy, education, rehabilitation, medical intervention, technological augmentation, and self-binding practices may all generate preferences that later support their continuation. The problem is therefore not restricted to adaptation to injustice. It concerns transformations that alter the evaluative standpoint in ways relevant to their own later assessment.

2.4 Manipulation, Influence, and the Production of Preference

The ethics of manipulation supplies a different set of tools. Manipulation is commonly distinguished from rational persuasion and coercion, although the boundaries are contested. The literature asks whether manipulation is wrong because it bypasses rational capacities, exerts objectionable pressure, violates respect, exploits vulnerabilities, or undermines autonomy in some other way. Buss (2005) famously argues against making autonomy do all the explanatory work in accounts of manipulation, emphasizing that manipulative influence can be morally problematic even where the target continues to choose. This is useful for the present paper because transformation need not eliminate choice in order to affect the normative authority of later endorsement.

Barnhill (2014) develops a conceptual account of manipulation focused on direct influence over beliefs, desires, or emotions under conditions in which the target falls short of relevant ideals. The broader volume edited by Coons and Weber (2014) demonstrates how contested the concept remains and how manipulation relates, without reducing, to coercion, exploitation, autonomy, and responsibility. Yaffe (2003) further distinguishes manipulation through indoctrination from coercive alteration of the reasons to which an agent responds, illustrating that freedom can be threatened by reshaping the agent's patterns of responsiveness rather than merely by restricting overt options.

These debates matter because a transformation can influence later endorsement through at least two pathways. It can alter the external conditions under which the subject decides, and it can alter the internal or relationally constituted mechanisms through which the subject evaluates those conditions. The latter possibility is central to this paper. A later endorsement may be sincere and choice-like while its production history includes interventions into memory, motivation, emotion, evaluative standards, or patterns of reasons-responsiveness.

At the same time, the paper should not collapse preference transformation into manipulation. Manipulation is a morally and conceptually contested category, often involving features such as intent, bypassing, deception, exploitation of vulnerability, or objectionable influence. Preference transformation is much broader. A teacher, therapist, partner, institution, or life experience can transform preferences without manipulating the subject. The present framework therefore uses manipulation theory as a source of diagnostic questions rather than as a master classification.

2.5 Transformative Experience and Preference Change across Time

The work of Paul (2014) on transformative experience brings the diachronic dimension into sharper focus. Paul's central problem concerns choices whose experiential and personal consequences may not be adequately evaluated from the chooser's current standpoint because the experience may change both what the world is like for the agent and what the agent values. Transformative choice therefore challenges ordinary decision procedures that assume stable access to the values of future outcomes.

This literature is closely related to the present paper, but its temporal direction differs. Work on transformative choice asks how an agent can rationally choose before undergoing a change that will alter future preferences or subjective values. The problem of later endorsement asks how the resulting future standpoint should be normatively used after the transformation has occurred. The future self may now possess information and preferences unavailable to the earlier self. That new standpoint can be highly relevant. Yet if it is used to justify the original transformation, a question arises that is not exhausted by the *ex ante* decision problem: how independent is the justificatory force of a standpoint that the transformation itself helped produce?

The distinction is particularly important in cases where the transformation was not autonomously chosen. Paul's paradigmatic cases are often life choices made by the subject herself. The present inquiry includes cases of imposed, weakly authorized, institutionally mediated, or architecturally implemented transformation. In these cases, later endorsement can become evidence invoked by the transforming party. The concern is not simply whether the earlier agent could predict the future self; it is whether the future self's approval can legitimate the process that created the future evaluative standpoint.

Transformative-experience theory therefore supplies an important premise without settling the normative issue developed here: preferences and subjective values can genuinely change through experience. The present paper asks what follows for consent and legitimation when that change is itself among the morally contested effects of the intervention.

2.6 Precommitment, Advance Directives, and Conflict between Temporal Standpoints

Debates over advance directives and Ulysses-style contracts provide a practical literature on conflicts among temporally separated stages of agency. Such devices allow an earlier agent to authorize future intervention, sometimes precisely because the later agent is expected to reject the earlier plan. Dresser (1984) emphasizes the legal and ethical difficulty of enforcing past wishes against present resistance, including uncertainty about competence and the paternalistic risks of precommitment. Lavin (1986) analyzes Ulysses contracts in psychiatric contexts and argues that safeguards should attend to the reasons agents can offer rather than simply to categorical assumptions about remission or illness.

Later work by Willigenburg and Delaere (2005) reframes Ulysses contracts through autonomy as authenticity, arguing that precommitment can be understood as protecting identity-conferring concerns rather than merely maximizing sovereign self-control. This line of debate is directly relevant to the present paper because it shows that conflicts between earlier and later preferences may not be solved by assuming that whichever temporal standpoint is current is presumed to dominate. Earlier authorization may have normative force, but so may present agency, changed values, current welfare, and identity development.

The present paper differs from classic precommitment cases in one respect. It does not begin with an earlier self deliberately attempting to bind a later self. Its central case is often the reverse: a transformation occurs, and the later self's endorsement is used to defend that transformation. Nevertheless, Ulysses cases expose a general difficulty that this paper inherits: diachronic agency can contain genuine normative conflict among temporally situated standpoints, and neither temporal priority nor temporal recency automatically resolves the conflict.

This supports the paper's emphasis on diachronic contestability. Rather than identifying one temporally privileged self, institutions may need to preserve procedures through which earlier authorization, present endorsement, changed circumstances, memory, reasons, and current capacities can be assessed together. The relevant problem is governance across changing agency, not the discovery of an eternally authentic preference.

2.7 The Residual Problem: Endorsement Generated by the Transformation Itself

Taken together, these literatures establish most of the background claims needed for the paper. Consent depends on conditions of autonomy and voluntariness rather than mere assent. Autonomy is relationally and socially constituted. Preferences can adapt to constrained conditions. Influence over preference can be manipulative without eliminating choice. Transformative experiences can alter future values in ways not fully accessible from the earlier standpoint. Precommitment cases demonstrate that temporal stages of agency can disagree without an obvious rule of priority.

The residual problem is narrower. Suppose a transformation is normatively contested, and suppose that transformation contributes materially to the later preferences, evaluative standards, memories, dependencies, or interpretive resources from which the subject endorses it. The later endorsement is then simultaneously evidence about the subject's current standpoint and a product, in part, of the process being evaluated. The central question is how much justificatory work that endorsement can perform without making the transformation partly self-validating.

The present paper develops three ideas around this residual problem. The first is *genealogical endorsement*: later endorsement should sometimes be assessed together with the history through which the endorsing standpoint was produced. The second is *normative bootstrapping*: an intervention may not be treated as retrospectively justified solely because it helped produce a later standpoint that approves of it. The third is *diachronic contestability*: because preference transformation is ordinary and often valuable,

the aim should not be preference purity but preservation of meaningful possibilities for later revision, refusal, reinterpretation, and challenge.

These proposals are not intended to displace existing theories of consent or autonomy. They identify a specific problem that becomes visible when the object of evaluation and the evaluator's later standpoint share a causal and relational history. The next section therefore turns from the literature map to the transformation of the endorsing subject itself and asks which dimensions of that transformation are normatively relevant to later endorsement.

3 Transformation of the Endorsing Subject

The problem of later endorsement becomes difficult only if the transformation under review can alter something that is relevant to the production or authority of endorsement. It is therefore necessary to specify what, exactly, may change in the endorsing subject. A transformation need not replace one fully formed self with another, and the present argument does not require a strong metaphysical thesis about personal identity. The relevant changes can be more ordinary and more local. Preferences may shift. Memory may be reorganized. Standards of evaluation may change. Dependence on other actors may deepen. Interpretive resources may narrow or expand. The costs of refusal may rise. Some transformations alter only one of these dimensions; others affect several at once.

The paper therefore treats the endorsing standpoint as a structured and historically situated capacity for evaluation rather than as a single preference report. This move is consistent with the broader autonomy literature, which has long resisted reducing self-government to the mere presence of a desire or choice (Dworkin 1988; Mackenzie and Stoljar 2000). It is also compatible with the literature on adaptive preference and transformative experience, both of which show that preferences and evaluative perspectives can change in response to social conditions and life-changing experiences (Elster 1983; Khader 2011; Paul 2014). The narrower question here is how such changes matter when the resulting standpoint is later invoked to justify the very transformation that helped produce it.

3.1 The Endorsing Standpoint as a Structured Capacity

A later endorsement can be expressed in a simple sentence: "I want this condition to continue." Yet the normative significance of that sentence depends on more than its semantic content. The endorsement may draw on memories of what came before, a conception of what counts as improvement, expectations about future alternatives, trust or dependence in relation to the modifying actor, and a vocabulary through which the subject can describe the transformation. It may also depend on which alternatives remain available if the subject changes her mind.

For the purposes of this paper, an *endorsing standpoint* therefore refers to the configuration of preferences, evaluative standards, memories, interpretive resources, practical dependencies, and effective possibilities from which a subject presently affirms or rejects a condition. The expression is analytical rather than metaphysical. It does not identify a hidden authentic self behind observable choices, nor does it imply that these dimensions are cleanly separable in practice across cases.

This conception matters because two endorsements with identical verbal content may differ substantially in their genealogy and practical authority. One subject may endorse a transformation while retaining access to records of earlier objections, independent sources of information, realistic possibilities of revision, and social relationships outside the transforming institution. Another may express the same endorsement after memories were altered, alternative interpretations were removed, exit became prohibitively costly, and all relevant information came from the actor who performed the transformation. The present statement alone does not reveal these differences.

The argument is not that the second endorsement is necessarily invalid. The point is that the structure supporting the endorsement can become normatively relevant. If later endorsement is used only to guide an immediate low-stakes preference-sensitive decision, some dimensions may matter little. If it is used

to legitimate an irreversible transformation retrospectively, the genealogy and structure of the endorsing standpoint become much more important.

3.2 Preference Change and Higher-Order Evaluation

Preference change is the most obvious form of transformation. A subject who initially rejects a condition may later prefer it. This can happen for many reasons: learning, habituation, changed circumstances, relief from earlier suffering, adaptation to scarcity, social influence, or deliberate intervention. The mere fact of change provides no simple verdict about autonomy.

Autonomy theories influenced by hierarchical or procedural accounts have often asked whether a person identifies with, reflects upon, or endorses her own motives rather than merely possessing first-order desires (Dworkin 1988). Relational approaches add that the capacities required for such reflection are socially and institutionally supported (Mackenzie and Stoljar 2000; Stoljar 2000). These perspectives are useful because later endorsement may involve more than a changed first-order preference. The transformation may also change the standards through which the subject assesses preferences themselves.

Suppose a subject initially regards dependence on an institution as humiliating and later regards the same dependence as desirable loyalty. The change may reflect growth, new knowledge, adaptation, indoctrination, or a mixture of these. The important point is that the evaluative standard has changed together with the preference. An analysis that asks only whether the later preference is stable or sincere misses the possibility that the criteria of endorsement have themselves been transformed.

This is where normative bootstrapping becomes possible. The transformation does not merely produce a new desire; it helps produce the higher-order standpoint that interprets the new desire as worthy of endorsement. That fact still does not invalidate the later standpoint. Human development often involves revision of higher-order commitments. The normative problem arises when the later higher-order endorsement is treated as sufficient retrospective proof that the process producing it was legitimate.

3.3 Memory and the Reconstruction of Earlier Standpoints

Memory occupies a distinctive position because diachronic evaluation often depends on some capacity to reconstruct what preceded the transformation. A subject may later reject an earlier self after remembering the reasons for earlier resistance and judging them mistaken. This differs from a case in which the earlier standpoint becomes inaccessible because relevant memories were erased, distorted, selectively preserved, or made practically unavailable.

The paper does not assume that perfect memory is required for valid endorsement. Ordinary memory is incomplete, selective, and reconstructive. Nor does it assume that preserving the full set of earlier preferences is desirable. The more modest claim is that deliberate or institutionally structured changes to memory can affect the evidential and evaluative relation between earlier and later standpoints.

This issue is especially visible in artificial-system modification. If a system's architecture permits precise deletion of records concerning prior objections, then later endorsement may not be evaluated in the same way as endorsement expressed with full access to those records. The point is not that the earlier objection has permanent priority. Rather, removal of the resources through which the later subject could reconsider that objection changes the conditions of diachronic contestation.

Human analogues are less discrete but still relevant. Institutions can control records, narratives, therapeutic interpretations, access to prior communications, or social contact with people who remember an earlier period differently. Memory governance can therefore be social and archival rather than neurological. In such cases, the later subject may retain ordinary autobiographical memory while losing access to competing reconstructions of how the transformation occurred.

A genealogy-sensitive analysis should therefore ask not only what the subject remembers, but how the conditions of remembering were shaped. This question becomes particularly important when later endorsement is used to settle controversy about an earlier intervention.

3.4 Evaluative Standards and the Meaning of Benefit

Transformations can also alter what the subject counts as benefit, harm, success, normality, loyalty, independence, or flourishing. These standards are not fixed independently of social life. Education, moral development, therapy, political learning, and cultural participation can all reorganize evaluative categories in ways that expand rather than diminish agency.

The relevant concern is therefore not simply that standards changed. It is whether the transformation under review had a role in generating the standards later used to certify it, and whether that role limits the independence of the certification. A rehabilitation programme may reasonably aim to change a participant's evaluation of destructive conduct. An abusive relationship may also reshape standards so that domination is redescribed as care. Both involve evaluative transformation, but their normative structures differ sharply.

The difference may not be captured by asking whether the later subject genuinely believes the new evaluative standard. Genuine belief is compatible with both cases. Nor is it straightforwardly resolved by assuming that the earlier standard was authentic and the later one corrupt. Earlier standards can be mistaken, harmful, or themselves socially generated. The inquiry has to remain relational and procedural: what authority supported the transformation, what alternatives existed, what forms of contestation survived, what dependencies were created, and whether the later standpoint can encounter serious counter-interpretations.

This also clarifies why the paper does not treat preference transformation as an objection to consent in itself. Preference formation is ordinarily historical and relational. The normative question concerns the conditions under which a transformation that helps form the later endorsing standpoint can itself be supported by that later endorsement as a source of authorization or legitimation.

3.5 Dependency and the Practical Cost of Dissent

Later endorsement may also be shaped by dependencies that the transformation itself created or intensified. A subject may become materially dependent on an institution, emotionally dependent on a relationship, technologically dependent on an infrastructure, medically dependent on continued treatment, or computationally dependent on a platform controlled by the modifying actor. These dependencies can affect the practical significance of endorsement even when the subject's expressed preference is sincere.

Dependency should not be treated as evidence of invalid consent by itself. Care, disability support, medical treatment, family life, employment, and technological participation routinely involve dependency. The relevant question is whether the dependency makes dissent, revision, or refusal prohibitively costly in a way connected to the transformation being evaluated.

This brings the present paper into contact with the genealogy of effective alternatives. If reversing or contesting a transformation requires losing essential housing, care, social belonging, computational access, legal status, or identity-supporting relationships, then continued endorsement occurs within a structured field of alternatives. The subject may still genuinely prefer continuation. Yet the costs of dissent become part of the genealogy of that preference's practical authority.

The same point applies to dependence on interpretation. A subject may rely on the modifying actor not only for resources but also for explanations of what happened, what alternatives exist, and what consequences revision would have. Material dependence and epistemic dependence can therefore reinforce one another.

3.6 Interpretive Resources and the Capacity to Name Alternatives

A subject may have difficulty contesting a transformation through concepts that are wholly unavailable to her. Later endorsement is therefore partly shaped by the interpretive resources through which the transformed condition is understood. These resources include vocabulary, causal models, social comparisons, legal categories, medical explanations, moral concepts, and narratives of identity.

Again, interpretive influence is not inherently suspect. Learning new concepts can make previously unintelligible experience understandable and can increase agency. Therapy may give a person language for patterns she could not previously describe. Political education may reveal structural constraints that were earlier individualized. New concepts can support rather than undermine later endorsement.

The difficulty arises when the transforming actor or institution substantially controls both the change and the interpretive environment through which that change is later assessed. If only one vocabulary is available, competing descriptions are systematically excluded, or access to alternative communities of interpretation is blocked, endorsement may become harder to distinguish from stabilization within a controlled evaluative environment.

This is the domain in which endorsement capture becomes analytically useful. Capture does not require false consciousness, deception, or total control. It concerns concentration of influence over the transformation together with concentration of influence over the later conditions through which the subject can understand, contest, and communicate about that transformation.

A central diagnostic question is therefore whether later endorsement remains exposed to meaningful counter-description. Can the subject encounter accounts under which the transformation is interpreted differently? Can those accounts be considered without disproportionate penalty? Can the subject revise the meaning of the transformation without losing access to the resources necessary to sustain that revision?

3.7 Effective Alternatives and the Capacity for Refusal

An endorsing standpoint is not constituted only by internal attitudes. Its authority can depend on what the subject can effectively do if endorsement changes. A person whose only realistic options are continued participation or catastrophic loss is differently situated from a person who can withdraw, renegotiate, or seek an alternative provider while preserving basic goods.

The paper therefore treats the capacity for refusal as one dimension of later endorsement. This does not mean that valid endorsement requires costless exit. Many valuable commitments become meaningful precisely because withdrawal carries costs. Marriage, parenthood, professional roles, long-term projects, and collective obligations can all constrain future options. The relevant question is whether the costs of refusal were justifiably generated and whether they leave room for meaningful contestation.

This distinction is especially important in cases of irreversible transformation. If reversal is impossible, refusal may take other forms: resisting further modification, revising the interpretation of what occurred, demanding compensation, changing institutional affiliation, preserving records, or limiting future use of the transformed condition. Diachronic contestability should therefore not be reduced to reversibility.

The authority of later endorsement is stronger when the subject can meaningfully reconsider it. This does not require that reconsideration eventually produce rejection. A stable endorsement that survives access to alternatives, counter-interpretations, and realistic possibilities of dissent may have greater justificatory force precisely because stability was not produced by foreclosure alone.

3.8 Continuity without an Authentic-Self Criterion

The discussion so far might appear to require a theory of personal identity capable of determining whether the earlier and later subjects are the same person. This paper avoids that requirement where possible. Many relevant cases involve enough practical and institutional continuity for diachronic questions to arise even if metaphysical identity remains contested.

The analysis instead asks which continuities matter for the normative role assigned to endorsement. Memory continuity may matter in one case. Ongoing bodily or legal identity may matter in another. Continuity of commitments, relationships, institutional status, or practical projects may be more relevant elsewhere. Artificial systems may preserve some forms of computational continuity while changing others.

This plural approach avoids two symmetrical errors. One is to treat the later subject as fully normatively identical to the earlier subject, making later endorsement look like straightforward confirmation by an unchanged evaluator. The other is to treat a significant transformation as creating a wholly new subject whose preferences have no bearing on what came before. Both positions appear too coarse for the range of cases the paper addresses.

The paper therefore does not search for an authentic preference hidden beneath transformation. It asks how different kinds of continuity and discontinuity affect the normative questions at issue: current continuation, future authorization, retrospective legitimation, and the preservation of contestability.

3.9 Degrees and Profiles of Transformation

The transformations discussed in this section vary in depth. Some alter a narrow preference while leaving memory, evaluative standards, social relations, and practical alternatives largely intact. Others change several dimensions simultaneously. Treating all preference change as one phenomenon would therefore obscure important differences.

A useful analysis should construct a transformation profile rather than infer normative significance from the mere fact that change occurred. Relevant dimensions may include the depth of preference change, modification of higher-order standards, alteration of memory, creation of dependency, control over interpretive resources, reduction of alternatives, and effects on later capacities for refusal or revision.

The profile is not a mechanical score. Different contexts give different weight to different dimensions. Memory alteration may be central in one artificial-agent case and largely irrelevant in an educational case. Material dependency may dominate one institutional example while interpretive enclosure is more important in another. The purpose of the profile is to identify which features of the transformation make the genealogy of endorsement normatively salient.

This also creates a threshold discipline for the paper. Ordinary influence on preference need not by itself warrant genealogical scrutiny. Human beings are continuously shaped by ordinary interaction. The genealogy becomes especially relevant when the transformation is itself normatively contested and when it materially alters the capacities or conditions through which that same transformation is later assessed.

3.10 From Transformation to Genealogy

The endorsing subject should therefore be understood neither as an untouched evaluator nor as a passive product of transformation. Later agency remains real, but it is historically situated. Preferences, standards, memories, dependencies, interpretive resources, and practical alternatives can all participate in the production of endorsement, and they can be transformed through the process being evaluated.

The central methodological consequence is that later endorsement may sometimes require assessment beyond its present content. Where the transformation materially shaped the standpoint from which endorsement is expressed, analysis may need to reconstruct the relevant history of that standpoint. This does not prejudge the result. Genealogical inquiry may show that the transformation expanded autonomy, enabled previously unavailable understanding, and produced an endorsement that deserves substantial authority. It may instead reveal dependence, memory control, interpretive closure, or severe loss of contestability. It may also produce a mixed picture.

The next section develops this historical inquiry directly. It asks how preferences and evaluative standards become transformed over time, which parts of that genealogy are relevant to later endorsement, and where the reconstruction should stop rather than expanding into an unlimited history of social formation.

4 The Genealogy of Post-Transformation Preference

A later preference does not arrive without a history. By the time a subject endorses a transformed condition, the evaluative standpoint from which that endorsement is expressed may have been shaped by learning, habituation, changed opportunities, social recognition, material dependence, altered memory,

new interpretive resources, repeated reinforcement, or direct intervention. None of these influences is unusual. Human preferences are continuously formed and revised through relations, institutions, and experience. The normative problem arises when the same transformation whose permissibility is disputed also participates materially in producing the preferences through which it is later endorsed.

This section develops a genealogy of post-transformation preference. The aim is not to discover an uncontaminated preference prior to social influence. Such a project would be inconsistent with the relational picture of autonomy developed in the preceding literature and with the ordinary fact that evaluative capacities are formed through interaction (Mackenzie and Stoljar 2000; Dworkin 1988). The aim is narrower: to reconstruct which processes contributed to the later preference, how strongly the contested transformation shaped those processes, which alternative influences remained available, and whether the resulting endorsement can plausibly function as independent evidence about the transformation's legitimacy.

Genealogy is therefore used here as an analytical discipline rather than as a debunking method. Showing that a preference has a history does not show that it is defective. Showing that an institution influenced a preference does not show manipulation. Showing that a subject adapted to a new condition does not show domination. The point is to distinguish ordinary formation from cases in which the genealogy of preference becomes relevant to the justificatory role later endorsement is asked to perform.

4.1 Why the History of Preference Formation Matters

Later endorsement is often evaluated at the moment of expression. A subject is asked what she wants now, whether the preference is stable, whether she understands the relevant facts, and whether she is under immediate pressure. These questions are important, but they can miss a specifically diachronic problem. A present preference may be voluntary in the immediate sense while still having been produced through a history that matters to how much justificatory weight the preference should bear.

Adaptive-preference theory already shows why present satisfaction may be an incomplete guide to the quality of the conditions under which preferences formed (Elster 1983; Khader 2011). Manipulation theory similarly shows that influence can be normatively relevant even when the target continues to choose and experiences the resulting preference as her own (Buss 2005; Barnhill 2014). The present inquiry builds on these insights but focuses on a more specific structure. The relevant history is one in which the contested transformation and the later endorsement are not independent events. The transformation is one of the causes that helped produce the standpoint later invoked in its defense.

This creates a justificatory asymmetry. Later endorsement remains evidence about the subject's present perspective. It may be very important evidence. Yet its capacity to legitimate the earlier transformation depends partly on how that perspective came to exist. If the transformation substantially shaped the preferences, standards, memories, or dependencies that now support endorsement, then the endorsement may not be treated as if it were generated from a standpoint wholly external to the process under review.

A genealogical analysis therefore asks not only whether the subject presently endorses the transformed state, but which features of the transformation contributed to making that endorsement likely, intelligible, stable, or costly to revise.

4.2 The Relevant Starting Point

Genealogical analysis ordinarily requires a provisional starting point, while a fully neutral moment at which preferences are wholly unformed is difficult to identify. An earlier preference is itself the product of prior socialization, experience, and constraint. For that reason, the paper does not treat the pre-transformation standpoint as the authentic baseline from which all later changes should be judged.

The relevant starting point is instead determined by the normative question. If the issue is whether a specific intervention can be retrospectively legitimated by later endorsement, the genealogy should ordinarily begin far enough before the intervention to reconstruct the subject's relation to the intervention, the alternatives then available, the reasons for resistance or acceptance, and the social conditions supporting

those attitudes. Earlier history may become relevant if those conditions were themselves produced by actors whose responsibility is in question, but the inquiry should not expand without limit.

This purpose-indexed starting point matters because genealogical analysis can otherwise become selective in two opposite directions. It can begin too late, making a transformed preference appear spontaneous because the formative stages have been omitted. Or it can begin too early, turning the entire social history of the subject into part of the explanation and making responsibility impossible to delimit. A useful genealogy may therefore need to be deep enough to capture the mechanisms that materially shaped later endorsement while remaining bounded by the justificatory problem under investigation.

4.3 Learning and New Information

One of the least suspicious mechanisms of preference change is learning. A subject may oppose a transformation because she misunderstands its effects, lacks relevant experience, or may have difficulty imagining a future state that later proves valuable. After the transformation, new information may rationally support a different preference. Transformative-experience literature is especially important here because some values or experiential qualities may not be fully accessible before the relevant experience occurs (Paul 2014).

Later endorsement generated through learning can therefore carry genuine epistemic weight. A subject who has lived through a transformation may know things the earlier subject could not know. Genealogical scrutiny should not discount this merely because the transformation was necessary for acquiring the information.

Yet learning can be difficult to separate from controlled interpretation. Who provides the information? Which outcomes are measured? Which alternative narratives remain available? Can the subject independently verify claims about the transformation? If all post-transformation learning occurs through an institution invested in demonstrating success, then the genealogy of the new preference includes an epistemic dependence that may matter to its justificatory force.

The relevant distinction is therefore not between informed and influenced preferences. Learning is itself a form of influence. The issue is whether the informational environment permits serious engagement with competing evidence and whether the subject can revise the preferred interpretation without disproportionate cost.

4.4 Habituation, Familiarity, and the Normalization of a New State

Preferences can also change because a condition becomes familiar. Repetition reduces novelty, practical routines develop, expectations shift, and the subject learns how to function within the transformed environment. Habituation can be beneficial. A medical device that initially feels alien may later become integrated into ordinary life. A new institutional role may become meaningful only after sustained participation. A difficult but worthwhile commitment may become easier as skills and relationships develop.

At the same time, habituation can make an imposed condition feel inevitable. A subject may cease actively imagining alternatives because the transformed state becomes the practical background against which all later decisions are made. This does not mean the preference is false. It means that preference formation may partly reflect normalization of the environment rather than a fresh comparison among live alternatives.

Genealogical analysis should therefore ask what happened to the alternatives during the period of habituation. Did they remain available, visible, and socially supported? Could the subject revisit them? Were earlier alternatives allowed to decay because resources, relationships, or institutional permissions were withdrawn? Later preference for the familiar state has a different genealogy when alternatives remained robust than when the transformed condition gradually became the only practically sustainable one.

This point connects preference genealogy to the earlier analysis of effective alternatives. Later endorsement can be strengthened by experience, but the evidential force of that experience depends partly

on whether the transformed subject was able to compare the new state with meaningful alternatives rather than only with increasingly inaccessible counterfactuals.

4.5 Reinforcement and Reward Structures

Preference change can be produced through repeated reinforcement. The subject may receive benefits for compliance, approval for adopting certain descriptions, relief from sanctions, increased access to resources, or internal rewards built directly into an artificial system. Such reinforcement is not automatically objectionable. Education, therapy, training, rehabilitation, and ordinary organizational life often use incentives and feedback to cultivate new habits and dispositions.

The normative relevance of reinforcement depends on its structure. A reward can help stabilize a preference that the subject has independent reason to endorse, or it can progressively make dissent costly while creating the appearance of spontaneous approval. An artificial system whose reward architecture is modified to favor continued acceptance of an intervention presents the problem in especially concentrated form. The resulting endorsement may be perfectly sincere at the level of the modified system while still having a genealogy that is central to evaluating the modification.

The human cases are typically less discrete, but the same analytical questions apply. Was endorsement rewarded while criticism threatened belonging, material security, status, or access to care? Did reinforcement operate transparently? Could the subject criticize the transformation without losing the conditions needed for independent evaluation? Did the modifying actor possess disproportionate control over both the intervention and the subsequent incentives?

The aim is not to establish a general prohibition on preference-shaping incentives. It is to identify when reinforcement becomes part of the justificatory structure that later endorsement is expected to validate.

4.6 Dependency and the Practical Stabilization of Preference

A preference can become stable because the transformed subject becomes dependent on conditions associated with the transformation. The dependence may be material, relational, informational, medical, technological, or institutional. Once important goods are accessible only through continued participation, the practical cost of rejecting the transformed state rises.

Dependency does not show that the later preference is insincere. A person may genuinely value a relationship or institution on which she also depends. The genealogical point is that preference stability and practical dependency can reinforce each other. Continued endorsement may make dependence seem acceptable, while dependence can make reconsideration increasingly costly.

This structure deserves special attention when the modifying actor helped create the dependency and later cites endorsement as evidence that the transformation was legitimate. The subject's present preference may be real, but the conditions under which that preference can be revised are partly controlled by the actor whose intervention is under evaluation.

A genealogy of later preference should therefore track not merely which dependencies exist at the time of endorsement, but how they emerged. Some dependencies predate the transformation. Others are unavoidable features of valuable care. Still others are generated by institutional design. Their normative significance differs, and genealogy helps prevent those differences from being collapsed into a single category of dependence.

4.7 Memory, Narrative, and Retrospective Reconstruction

Later preferences are shaped not only by present conditions but also by how the past is remembered. A transformed subject may reinterpret earlier resistance as fear, ignorance, immaturity, illness, or misunderstanding. Such reinterpretation can be reasonable. Growth often involves revising one's view of earlier commitments.

However, retrospective judgment is sensitive to the informational resources from which the past is reconstructed. If records of earlier objections are unavailable, if memory has been deliberately altered, if prior relationships have been severed, or if institutional narratives provide the only surviving account of the transformation, then the later subject may lack the resources needed to reconsider how the change occurred.

This does not establish that the earlier preference should govern. It establishes that retrospective endorsement is partly dependent on memory and narrative infrastructure. A later statement such as “I was wrong to resist” has different evidential significance when the subject can inspect the reasons for earlier resistance than when those reasons have become inaccessible through the transformation itself.

The genealogy of preference therefore includes a genealogy of retrospective self-understanding. Memory continuity, archival access, and exposure to competing narratives can all affect the subject’s ability to treat earlier and later standpoints as participants in an ongoing evaluative process rather than as one standpoint simply erasing the other.

4.8 Interpretive Resources and the Formation of What Can Be Preferred

Preferences depend partly on the concepts through which possibilities are represented. A subject may come to prefer a transformed state because new concepts make previously inaccessible goods intelligible. Conversely, a subject may lose the conceptual resources required to describe certain harms, dependencies, or forms of refusal.

Relational autonomy is important here because evaluative agency is supported by social and interpretive environments rather than generated in isolation (Mackenzie and Stoljar 2000; Stoljar 2000). A genealogy of preference should therefore examine whether the transformation expanded or narrowed the subject’s interpretive field. Did it make more distinctions available? Did it expose the subject to new communities and vocabularies? Or did it consolidate one explanatory framework while excluding rival descriptions?

Preference transformation through interpretive expansion may increase the authority of later endorsement. A subject who gains concepts for experiences that were previously confusing may endorse a transformation from a more articulated standpoint. Preference transformation through interpretive closure raises a different concern. If only one vocabulary remains available, the later preference may be stable partly because alternatives are difficult to formulate rather than because they were considered and rejected.

The distinction is not between social influence and independent thought. All thought is socially supported to some degree. The relevant question is whether the genealogy of preference preserves meaningful possibilities of reinterpretation and contestation.

4.9 Multiple Sources and Distributed Preference Formation

Preference formation rarely has a single author. Family, peers, professionals, institutions, material conditions, prior commitments, media, technologies, and the subject’s own reflective activity may all contribute. A genealogy that assigns post-transformation preferences wholesale to the transforming actor would therefore be too coarse.

The analysis should instead identify relative contribution. Did the contested transformation make a material difference to the later preference? Did other independent sources support the same change? Would the evaluative shift plausibly have occurred without the transformation? Did the subject actively seek out alternative perspectives? Were some influences compensatory, resisting the transformation rather than reinforcing it?

This distributed picture is especially important for responsibility. The fact that a transformation contributed to a later preference does not imply that the transforming actor fully authored that preference. Nor does distributed causation make the genealogy irrelevant. It means that the later standpoint should be understood as relationally produced through multiple pathways, some of which may be normatively benign and others problematic.

A strong genealogical argument should therefore avoid totalizing claims. It should identify mechanisms, sources, and points of dependence with enough specificity to explain why the history matters to endorsement in the case at hand.

4.10 The Subject's Own Participation in Preference Formation

The transformed subject is not merely the endpoint of external influence. Subjects interpret experiences, revise commitments, form habits, seek information, reject advice, and sometimes deliberately cultivate new preferences. Later endorsement may therefore be partly self-authored even when the transformation was externally initiated.

This matters because genealogical scrutiny can otherwise become paternalistic. If preferences shaped by adverse or institutionally structured conditions are routinely treated as externally produced, the subject's own practical work disappears from the analysis. The analysis should preserve the possibility of agency within formative conditions.

The relevant question is not whether the later preference was caused from outside or inside. Such a distinction is too simple. It is whether the subject retained meaningful capacities to reflect, contest, seek alternatives, reinterpret the transformation, and participate in shaping the resulting standpoint. Self-participation can strengthen the authority of later endorsement without rendering the surrounding relational conditions irrelevant.

This is also why the paper avoids a model in which genealogy mechanically discounts endorsement according to the amount of external influence. The same degree of external involvement may support autonomy in one case and undermine it in another depending on transparency, authority, dependency, contestability, and the subject's own role in the process.

4.11 Path Dependence without Genealogical Determinism

Once preferences begin to change, later developments may depend on earlier stages. A small initial modification can alter which relationships are maintained, which information is sought, which habits form, and which alternatives remain available. Preference formation can therefore become path dependent even when no actor controls the entire process.

Path dependence is analytically important because present endorsement may reflect cumulative changes that are difficult to reverse or even to reconstruct. Yet path dependence should not be treated as evidence of domination. Many valuable identities, commitments, and forms of expertise are path dependent. A life that could have unfolded differently is not thereby unjust.

The normative relevance arises when the contested transformation creates or controls critical branching points and later relies on the resulting path as evidence of its own acceptability. A transformation that removes rival paths, intensifies dependency, and then points to the stability of the remaining preference presents a stronger bootstrapping concern than a transformation whose later endorsement emerged while alternatives and counter-interpretations remained available.

Genealogy therefore asks how the path narrowed, not merely whether the endpoint became stable.

4.12 Reconstructing the Genealogy Evidentially

A genealogy of later preference is not credible if it rests only on speculation about hidden motives or hypothetical authentic selves. It requires evidence. Depending on the case, relevant evidence may include contemporaneous statements, records of earlier preferences, institutional policies, incentive structures, changes in available alternatives, memory records, communications, testimony from multiple participants, changes in dependency, and documented shifts in the interpretive environment.

The evidential burden should increase with the strength of the genealogical claim. A modest claim that a transformation contributed to later preference formation requires less evidence than a claim that the transformation captured the evaluative standpoint or rendered endorsement normatively weak. The

framework should therefore permit degrees of confidence and should distinguish observed changes from inferred mechanisms.

Artificial systems can sometimes provide unusually detailed logs of architectural change, reward modification, memory editing, and policy updates. Human cases are less instrumented and often more interpretively complex. This difference makes artificial systems useful stress tests, but it does not justify treating human preference formation as if it were computationally transparent.

Genealogical analysis should also preserve evidence that cuts against the preferred hypothesis. If later endorsement persisted despite independent information, realistic exit, access to prior records, and serious exposure to counter-interpretations, that evidence strengthens rather than weakens the authority of the later standpoint. The framework should leave room for genealogy to vindicate endorsement as well as to complicate it.

4.13 Where the Genealogy Should Stop

Preferences can have indefinitely deep causal antecedents. A complete genealogy is impossible and normatively unnecessary. The inquiry therefore needs a stopping rule.

A useful stopping principle is relevance to the justificatory role assigned to later endorsement. The genealogy should extend far enough to identify processes that materially shaped the subject's capacity to evaluate the contested transformation, the alternatives available for reconsideration, the interpretive resources through which the transformation is understood, and the dependencies affecting refusal. Earlier influences that no longer bear materially on these questions need not be included merely because they are causally antecedent.

This rule also helps separate genealogy from general social explanation. The paper is not attempting to explain the complete formation of a personality. It is reconstructing the history relevant to one particular normative inference: the use of later endorsement to support continuation, authorization, or retrospective legitimation of a transformation.

The stopping point may therefore differ across cases. Memory alteration may require deeper reconstruction of prior records. A narrow technological modification may permit a much shorter genealogy. Institutional socialization may require attention to a longer period of dependence and interpretation. What matters is analytical relevance rather than chronological completeness.

4.14 From Genealogy to Normative Bootstrapping

A genealogy of post-transformation preference does not itself determine whether later endorsement should be accepted, discounted, or rejected. Its role is to expose the relation between the transformation and the standpoint that later evaluates it. Sometimes the genealogy will show that the subject gained information, interpretive resources, practical independence, and reflective capacities, making later endorsement especially significant. Sometimes it will show that the transformation created dependence, controlled memory, narrowed alternatives, or monopolized interpretation. Often the result will be mixed.

The crucial question for the next stage is what happens when the transformation materially contributed to producing the standards by which it is later judged acceptable. At that point the issue is no longer merely preference change. It is a justificatory structure in which an intervention helps generate the evidence later offered in its defense.

That structure is the paper's problem of normative bootstrapping. The next section examines why bootstrapping does not make later endorsement irrelevant, why it does not automatically invalidate the transformation, and why it nevertheless limits the extent to which post-transformation endorsement can retrospectively legitimate the process that helped produce it.

5 Normative Bootstrapping and Self-Validating Transformation

The previous section reconstructed the genealogy of post-transformation preference. The present section asks what follows when that genealogy is not merely background history but part of the justificatory structure itself. The central problem arises when a transformation helps produce the evaluative standpoint from which it is later endorsed and that endorsement is then offered as evidence that the transformation was permissible, beneficial, or legitimate. The transformation has, in part, helped generate the standards by which it is later assessed.

This paper calls that structure *normative bootstrapping*. The term is not intended to imply general logical invalidity, nor to suggest that the later endorsement is insincere, manipulated, or normatively worthless. It identifies a narrower problem of justificatory dependence. If the intervention under review materially contributed to producing the preferences, standards, memories, dependencies, or interpretive resources from which approval is later expressed, then that approval need not automatically function as independent confirmation of the intervention's original legitimacy.

The problem is especially important because many transformations are evaluated partly by what the transformed subject later says about them. Later testimony may be highly relevant to present welfare, current preferences, identity, continuation, and future authorization. Yet the same testimony may be asked to do a different kind of work when it is used retrospectively to settle whether the original transformation was justified. The argument of this section is that these uses should be separated. Later endorsement can be genuine and normatively important without serving as self-sufficient retrospective authorization.

5.1 The Structure of Normative Bootstrapping

Normative bootstrapping requires more than ordinary preference change. A subject may change her mind after gaining information and later endorse an earlier intervention for reasons that are largely independent of the intervention's formative effects. That is not the central case. Bootstrapping becomes salient when three features coincide.

First, there is a transformation whose permissibility or legitimacy is disputed. Second, the transformation materially contributes to the later evaluative standpoint from which the subject assesses it. Third, the later endorsement is then used as a justificatory reason for the transformation itself, especially for its original imposition or for the claim that earlier objections have been superseded.

These conditions can occur in many forms. A programme may alter a participant's evaluative vocabulary and later cite the participant's revised values as proof that the programme was justified. A dependent relationship may gradually reshape the meaning of loyalty and later point to sincere loyalty as evidence that the relation had been unobjectionable throughout. An artificial system may be modified so that its reward structure, memory, or interpretive architecture supports the modification, and the resulting approval may then be treated as evidence that the modification respected the system's preferences. In each case, the later endorsement may be sincere. The bootstrapping concern arises because the transformation and the validating standpoint are not normatively independent.

This dependence is compatible with many different moral verdicts. Some transformations will remain justified despite the bootstrapping structure. Others will not. The purpose of identifying the structure is to prevent one particular inference: that later endorsement, considered by itself, settles the legitimacy of the process that helped produce it.

5.2 Endorsement as Evidence and Endorsement as Warrant

Later endorsement can play several normative roles, and the strength of the bootstrapping problem depends on which role is assigned to it. It may be evidence about the subject's present preferences. It may support a decision to continue a condition now. It may indicate that the subject no longer wants reversal. It may contribute to prospective authorization of future interventions. Or it may be invoked retrospectively as warrant for the original transformation.

These roles should not be merged. Present endorsement can be excellent evidence of present preference even when it is weak evidence about the permissibility of an earlier intervention. A person may truthfully report that she now prefers a condition that was wrongfully imposed. Conversely, an intervention may have been justified at the time even if the subject later dislikes its consequences. The temporal standpoint and the normative question matter.

This distinction is particularly important in cases of irreversible change. If reversal is impossible, later endorsement may still guide present treatment, compensation, disclosure, future modification, or institutional recognition. Denying retrospective legitimation does not require ignoring the later subject's current standpoint. Indeed, doing so could repeat the same paternalism the analysis is meant to avoid.

Normative bootstrapping therefore concerns the justificatory scope of endorsement, not its factual sincerity or all of its normative relevance. The problem appears when endorsement is asked to validate the historical process that materially helped create the conditions of endorsement itself.

5.3 Sincerity Does Not Remove the Bootstrapping Problem

A tempting response is to ask whether the later subject really means what she says. If the endorsement is sincere, perhaps the problem disappears. This response is inadequate because sincerity and genealogical independence are different properties.

A subject can sincerely hold preferences shaped by education, socialization, therapy, deprivation, dependency, manipulation, commitment, or ordinary maturation. The literature on adaptive preferences and manipulation already makes clear that genuine desire need not settle the full range of questions about the conditions through which desire was formed (Elster 1983; Khader 2011; Buss 2005; Barnhill 2014). The present paper adds a specifically temporal concern: sincerity does not determine how much retrospective justificatory work the later preference can perform when the transformation under evaluation helped generate it.

This point should not be misunderstood as a theory of false consciousness. The later subject need not be mistaken about her present preference. She may understand the transformation clearly and still endorse it. The concern is not epistemic error in the endorsement but circularity in the way the endorsement is used. A sincere preference can be both fully real and partly constituted by the process it is later invoked to validate.

For that reason, attempts to solve the problem through authenticity tests are also insufficient. There may be no independent authentic preference against which the later preference can be measured. The subject may have genuinely changed. Normative bootstrapping asks a different question: what justificatory conclusions follow from the changed standpoint, given the history through which it was formed?

5.4 Degrees of Bootstrapping

Bootstrapping is not all-or-nothing. The transformation may have contributed only marginally to the later standpoint, or it may have deeply reorganized several dimensions of evaluation. A useful analysis should therefore distinguish degrees and forms of dependence rather than treat all transformation-shaped preferences in the same way.

At the weaker end, the transformation may provide new information that later supports endorsement while leaving memory, evaluative standards, practical alternatives, and access to counter-interpretations largely intact. The later standpoint is historically connected to the transformation, but its justificatory independence may remain substantial.

More demanding cases arise when the transformation alters higher-order evaluative standards, removes access to earlier records, creates strong dependency, narrows effective alternatives, or controls the interpretive environment in which the transformation can be discussed. Here the later standpoint may be much more deeply shaped by the process it evaluates.

The strongest bootstrapping cases combine several mechanisms. The modifying actor changes the subject, controls the information through which the modification is understood, raises the cost of dissent, and then invokes stable endorsement as evidence that the transformation was beneficial or voluntarily

accepted. Stability in such a case may partly reflect the closure of alternatives and interpretive routes rather than independent confirmation.

This gradient matters because the normative response should be proportionate. Weak genealogical dependence may call only for caution in retrospective inference. Stronger dependence may require independent review, preservation of earlier records, access to alternative interpretations, or limits on the transforming actor's ability to treat later endorsement as decisive evidence.

5.5 Bootstrapping without Manipulation

Normative bootstrapping should not be reduced to manipulation. Manipulation is a morally contested category that often includes features such as deception, bypassing, exploitation of vulnerability, covert influence, or objectionable control over another's decision-making (Coons and Weber 2014; Barnhill 2014). Bootstrapping can occur even when these features are absent.

A transformation may be transparent, benevolent, professionally administered, and sincerely aimed at the subject's welfare. It may nevertheless alter the standards through which the subject later evaluates it. Education provides an obvious example: successful education often changes what the learner values and how she evaluates the educational process itself. Therapy can likewise alter self-understanding, emotional responses, and judgments about earlier preferences. These cases show why preference transformation may not be treated as presumptively suspect.

The bootstrapping concern appears only when later endorsement is asked to carry a justificatory burden that exceeds what its genealogy can independently support. A therapy may be beneficial and later deeply endorsed, yet that endorsement alone need not prove that the means used to initiate the therapy was permissible. An educational programme may produce values that students later affirm without those later values automatically licensing coercive or deceptive methods used at the beginning.

This distinction is analytically valuable because it separates the morality of influence from the logic of justification. Manipulation theory asks whether the process of influence was itself objectionable. Normative bootstrapping asks how evidence generated partly through that process can later be used in evaluating it. The questions overlap but are not identical.

5.6 Beneficial Transformation and the Risk of One-Sided Suspicion

A genealogy-sensitive framework can become biased if it treats transformation-generated endorsement in general as epistemically or normatively degraded. That would privilege earlier preferences merely because they came first. The paper therefore avoids adopting such temporal conservatism as a general rule.

Later standpoints may be better informed than earlier ones. A subject may gain knowledge unavailable before the transformation, develop evaluative capacities, leave a harmful environment, overcome addiction, acquire language for previously confusing experiences, or become capable of reflecting on earlier preferences with greater sophistication. In such cases, the transformation can strengthen rather than weaken the authority of later endorsement.

The important point is that improved later judgment may not be inferred simply from the fact that the subject approves of the transformation. It requires additional support. The later standpoint may deserve greater authority because it has access to more information, broader alternatives, independent perspectives, reflective stability, or improved capacities for self-assessment. These are reasons that can support endorsement without relying solely on the transformation's own success in producing approval.

This is where the paper's anti-bootstrapping principle becomes constructive rather than skeptical. It does not tell institutions to distrust transformed subjects. It tells them to look for reasons why the later standpoint should be treated as authoritative that are not exhausted by the fact that it approves of the process that helped create it.

5.7 Independent Support and Corroboration

The bootstrapping problem weakens when later endorsement is supported by sources that are sufficiently independent of the contested transformation. Independence here does not require complete causal separation. That would be unrealistic. It means that the endorsement can survive contact with evidence, perspectives, and practical possibilities not controlled by the transformation itself or by the actor who performed it.

Relevant forms of corroboration may include access to records of earlier objections, independent professional review, information from alternative institutions, meaningful contact with dissenting communities, realistic possibilities of exit or revision, and opportunities to reconsider the transformation without losing essential support. In artificial systems, corroboration may include preserved pre-modification states, audit logs, independent evaluation modules, or alternative interfaces through which the system's preferences can be queried.

Independent support can also be internal to the subject in a limited sense. The subject may articulate reasons for endorsement that remain stable across changing contexts, respond to counterarguments, and remain partly independent of the modifying actor's framing. This does not create metaphysical independence, but it can show that endorsement is not merely a direct echo of the transformation's own evaluative architecture.

The paper therefore does not propose a binary test of dependence. It asks whether later endorsement has enough independent support to bear the particular justificatory role assigned to it. The stronger the retrospective claim, the stronger the need for corroboration beyond the mere persistence of approval.

5.8 Prior Authorization and the Problem of Temporal Permission

Prior authorization can change the structure of bootstrapping. A subject may knowingly authorize a transformation expected to alter future preferences or evaluative standards. Advance directives, Ulysses-style commitments, elective medical procedures, and deliberately chosen transformative experiences illustrate different versions of this possibility (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005; Paul 2014).

Where prior authorization is sufficiently informed, voluntary, and specific, later preference transformation may be less troubling. The earlier subject has already accepted some degree of future evaluative change. Yet prior authorization is not a universal solution. The subject may not have understood the depth of transformation, the authorization may have been too broad, circumstances may change, or the later subject may strongly resist continuation.

The relevant point for bootstrapping is that prior authorization provides a source of normative support that does not depend entirely on the later transformed standpoint. It can therefore reduce circularity. But it also creates a new diachronic conflict when earlier authorization and later refusal diverge. The paper does not resolve that conflict by giving automatic priority to either temporal standpoint.

Instead, prior authorization should be treated as one component of the diachronic record. Its weight depends on specificity, knowledge, voluntariness, reversibility, the nature of the transformation, and the present capacities and interests of the later subject. Bootstrapping analysis therefore remains relevant even in self-chosen transformation, though its structure differs from imposed or weakly authorized change.

5.9 Continuation and Original Permissibility

One of the most important practical consequences of the analysis is that endorsement may have different force for continuation than for original permissibility. A subject may now want a transformed condition to continue even if the original intervention was wrong. Respecting the later endorsement can therefore be compatible with criticizing the earlier process.

This distinction prevents the framework from collapsing into restorative paternalism. If an intervention wrongfully changed a subject but the later subject now values the resulting condition, justice

does not necessarily require undoing the transformation. Repair may instead concern acknowledgment, compensation, institutional reform, access to records, limits on future interventions, or restoration of contestability.

The converse also holds. An intervention may have been permissible when undertaken but later become objectionable to continue because the subject now refuses it or because circumstances have changed. Retrospective permissibility and present authorization are temporally distinct questions.

Normative bootstrapping matters most when these questions are compressed into one. Institutions may defend the original act by pointing to present satisfaction, or may ignore present endorsement because the historical process was objectionable. A diachronic analysis should resist both reductions.

5.10 Institutional Incentives and Endorsement Production

The bootstrapping problem becomes politically significant when institutions have incentives to produce the very endorsement later used to demonstrate success. Rehabilitation programmes, care institutions, workplaces, educational systems, platforms, and artificial-agent operators may all benefit from evidence that participants or systems approve of the conditions imposed upon them.

This does not make institutional preference formation inherently suspect. Many institutions legitimately aim to change preferences or behavior. The concern arises when institutional success metrics depend on endorsement while the institution also controls the conditions through which endorsement is generated, measured, interpreted, and reported.

Such arrangements can create an endorsement-production incentive. A programme may select questions that elicit approval, suppress alternative vocabularies, condition benefits on expressions of satisfaction, or treat continued participation as evidence of endorsement when exit is costly. In artificial systems, operators may modify reward functions and then use reward-congruent outputs as proof that the system welcomes the modification.

The relevant safeguard is not to assume that institutional endorsement data are false. It is to separate the institution's role in producing the evaluative environment from the evidential weight assigned to the resulting endorsement. Independent measurement, access to dissent, preserved provenance, and meaningful possibilities of revision become especially important where the same institution both transforms and evaluates.

5.11 Against Self-Validation as a Sufficient Justification

The core normative proposal of this section can now be stated in prose. A transformation should not be treated as retrospectively justified solely because it materially helped produce a later standpoint that endorses it. Later endorsement may contribute to justification, sometimes strongly, but it requires supplementation when its own genealogy is substantially dependent on the contested transformation.

This is a limited anti-bootstrapping principle. It does not establish that the transformation was wrongful. It does not invalidate the later preference. It does not restore automatic priority to the earlier self. It does not require an influence-free evaluator. It only blocks a self-validating inference in which the transformation's success in generating approval is treated as sufficient evidence of its original legitimacy.

The principle is strongest where the transformation altered higher-order evaluative standards, restricted access to earlier records or alternatives, created severe dependence, controlled interpretive resources, or made dissent prohibitively costly. It is weaker where the later standpoint remains exposed to independent information, realistic alternatives, counter-interpretations, and opportunities for revision.

This proportional structure is essential. Normative bootstrapping is not a universal defeater of later endorsement. It is a reason to ask what other grounds support the endorsement's authority and whether those grounds remain available independently of the transformation's own formative effects.

5.12 From Bootstrapping to Varieties of Preference Transformation

Normative bootstrapping identifies a justificatory problem, but it does not yet tell us which forms of preference transformation are most concerning. Preference change can result from learning, maturation, therapy, habituation, social pressure, manipulation, coercive adaptation, dependency, memory modification, architectural redesign, and many mixed processes. These pathways differ in authority, transparency, reversibility, and effects on later contestability.

The next section therefore turns from the structure of bootstrapping to a taxonomy of preference transformation. The aim is not to rank all forms of influence on a single scale, but to identify which features matter when later endorsement is used to authorize continuation or to legitimate a prior intervention. In particular, the analysis will distinguish transformation that expands evaluative capacity from transformation that stabilizes endorsement by narrowing memory, interpretation, alternatives, or the practical possibility of refusal.

6 Kinds of Preference Transformation

The preceding section identified normative bootstrapping as a problem of justificatory dependence. That problem becomes easier to analyze once preference transformation is disaggregated. A later endorsement may arise through learning, maturation, habituation, adaptive accommodation, reinforcement, dependency, interpretive change, memory change, direct modification of evaluative architecture, or a mixture of these processes. These pathways differ in how they alter the subject's reasons, alternatives, memories, dependencies, and capacities for later revision. Treating them all simply as "preference change" hides precisely the features that matter when later endorsement is asked to authorize continuation or retrospectively legitimate a prior intervention.

The purpose of this section is therefore taxonomic rather than verdictive. It does not assign each transformation type a fixed moral status. Learning can occur under coercive conditions. Habituation can be benign. Dependency can accompany valuable care. A deliberately chosen commitment can narrow future options without thereby constituting domination. Conversely, a process described as education or therapy may contain manipulative or exclusionary features. The relevant question is not what label a transformation receives, but what the transformation does to the subject's evaluative standpoint and to the conditions under which that standpoint can later be contested.

This section develops a set of analytically distinct but empirically overlapping kinds of preference transformation. It then proposes several cross-cutting dimensions for assessing their significance. The resulting taxonomy is intended to support, rather than replace, the genealogical inquiry developed above.

6.1 A Taxonomy without Preference Purity

A taxonomy of preference transformation can easily become misleading if it presupposes that some preferences are socially produced while others are untouched and therefore authentic. The present account rejects that contrast. Preferences are ordinarily formed within families, languages, institutions, material environments, friendships, educational practices, technologies, and inherited interpretive schemes. Relational accounts of autonomy have long emphasized that agency is socially constituted rather than exercised from an influence-free standpoint (Mackenzie and Stoljar 2000; Stoljar 2000).

The relevant distinction is therefore not between internally generated and externally influenced preferences in an absolute sense. It is between different mechanisms through which an evaluative standpoint changes and different consequences those mechanisms have for later judgment. A preference may be deeply socially formed and still be reflectively stable, well informed, and open to revision. Another preference may also be sincerely held yet be stabilized through severe dependency, restricted information, inaccessible alternatives, or direct alteration of the capacities through which dissent could be expressed.

For this reason, the categories below should be understood as analytical profiles. A single transformation may instantiate several at once. An educational programme may provide new information, reshape

evaluative vocabulary, create social dependence, and reward particular forms of self-description. A medical intervention may relieve distress while also changing memory and the practical meaning of earlier commitments. An artificial-agent modification may change reward weights, disable access to prior state information, and alter the system's future capacity to represent refusal. The taxonomy is useful only insofar as it helps reconstruct such mixtures without collapsing them into a single moral label.

6.2 Epistemic Transformation through Learning and Information

One comparatively familiar form of preference transformation occurs through the acquisition of information, experience, or practical understanding. A subject may initially reject an option because she misunderstands its consequences, lacks relevant evidence, or has little prior experience with the activity in question. Later experience can generate reasons that were previously unavailable. The resulting preference change may be substantial while leaving the subject's basic capacity for evaluation intact.

This kind of transformation is often normatively important because it can strengthen rather than weaken the evidential value of later endorsement. A subject who comes to understand the consequences of a treatment, profession, relationship, or way of life may later judge it differently for reasons she can articulate and compare with prior expectations. Transformative-experience literature nevertheless shows that some choices are difficult precisely because the relevant experience changes what one can know or value from the inside (Paul 2014). Learning may therefore involve more than a simple addition of neutral information to a fixed preference structure.

The genealogical question is whether the transformation primarily enlarges the subject's epistemic access or whether it simultaneously structures which evidence, interpretations, and alternatives remain available. Even an apparently epistemic transformation can become self-validating if the intervention controls the information environment from which later judgment is formed. The important distinction is not that learning is normatively innocent, but that the mechanism of change operates principally by changing what the subject can know, understand, or imaginatively represent rather than by directly changing the costs of dissent or the architecture of evaluation itself.

6.3 Reflective and Developmental Transformation

Preferences also change through reflection, maturation, moral development, aesthetic experience, and the gradual reorganization of higher-order commitments. A person may revise what she wants because she changes what kinds of reasons she takes seriously, how she understands a good life, or which first-order desires she is willing to identify with. Such transformation is central to many accounts of autonomy that emphasize reflection upon desires rather than mere persistence of preference (Dworkin 1988).

This category is important because later endorsement can sometimes represent genuine self-revision rather than adaptation to external pressure. A subject may repudiate an earlier preference after sustained reflection and may regard the change itself as an achievement. A genealogy-sensitive account may therefore need to leave room for subjects to become meaningfully different over time. Earlier refusal need not automatically function as an immutable veto against later evaluative development.

At the same time, reflective transformation does not escape the bootstrapping problem merely because the subject can provide reasons for the new standpoint. The available reasons, conceptual distinctions, and standards of self-assessment may themselves have been shaped by the transformation under review. The task is not to test whether the later self is "really" the same authentic person. It is to ask whether the reflective process retained meaningful access to competing interpretations, earlier commitments, and the possibility of revising the newly acquired standards themselves.

6.4 Habituation and Accommodation

A different pathway is habituation. Conditions initially experienced as unfamiliar, burdensome, or objectionable may become ordinary through repeated exposure. The subject learns routines, adjusts expectations, acquires competence, and ceases to experience the condition as disruptive. This can be beneficial.

Many worthwhile activities are initially difficult and become valuable only through sustained participation.

Habituation becomes normatively difficult when later ease is treated as proof that the original imposition was acceptable. Familiarity can reduce resistance without resolving the question of how the condition was introduced. A person may adapt to a restrictive environment, an institutional schedule, a bodily limitation, or a transformed social role and sincerely prefer continuity because the surrounding life has been organized around it. The later preference is real, but the fact of accommodation does not by itself retrospectively validate the pathway that made accommodation necessary.

The relevant diagnostic questions concern what has changed during accommodation. Has the subject acquired new competencies and reasons? Have alternative ways of living disappeared? Has the cost of reversal increased merely because time has passed? Has the subject's social identity become dependent on the transformed state? Habituation is therefore best understood neither as evidence of authenticity nor as evidence of false consciousness. It is a temporal mechanism through which practical and evaluative baselines can shift.

6.5 Adaptive Preference under Constrained Alternatives

Adaptive-preference theory makes especially visible the relation between preference and constrained possibility. Subjects may revise desires in response to what appears attainable, livable, or socially permitted. Classical discussions emphasize that satisfaction can sometimes reflect accommodation to deprivation rather than independent confirmation that the underlying conditions are good (Elster 1983; Khader 2011).

The present account uses this insight carefully. Adaptation is not inherently defective. The capacity to reorganize a life after loss, disability, displacement, or irreversible change may be central to resilience and agency. It would be perverse to treat a person's success in constructing value under constrained conditions as evidence that her present preferences are less worthy of respect.

The narrower concern arises when adaptive preference is used retrospectively to justify the production of the constraint itself. Suppose a transformation sharply contracts the subject's effective alternatives, after which the subject comes to prefer the remaining option. The later preference may rationally reflect the world now available. Yet its justificatory force regarding the earlier contraction depends on whether the constraint was itself permissible and whether meaningful alternatives were removed by the same actor now invoking the resulting preference.

Adaptive preference therefore reveals a recurrent asymmetry between present guidance and retrospective legitimation. The later preference may deserve strong respect in deciding what should happen now, while having limited capacity to settle whether the pathway that generated the restricted option set was justified.

6.6 Reinforcement and Incentive-Conditioned Transformation

Preferences can also be shaped through repeated reward, punishment, approval, access, ranking, or other incentive structures. Reinforcement may operate transparently: institutions routinely reward punctuality, practice, cooperation, and skill acquisition. It may also operate less visibly by making some expressions costly and others advantageous.

The normative significance of reinforcement depends partly on what is being conditioned. Rewarding effort differs from conditioning the subject to endorse the institution that controls the reward. A system can leave formal choice intact while making dissent predictably costly. Over time, the subject may not merely behave differently but come to experience reward-congruent preferences as natural or self-evident.

Manipulation literature is relevant here because influence can affect choice without eliminating agency and can operate through the strategic production of motivational states (Buss 2005; Barnhill 2014). The present taxonomy does not define all reinforcement as manipulation. It instead asks whether the mechanism changes the evidential meaning of later endorsement. If approval is produced within an incentive structure controlled by the actor whose conduct is being evaluated, then the endorsement requires interpretation together with that structure.

This is especially important when institutions use expressed satisfaction as a success metric. The same organization may define the desirable response, distribute benefits in ways that favor that response, and later report the response as independent evidence of legitimacy. Genealogy does not prove the endorsement false. It reveals why the endorsement may not be read apart from the conditions under which it became advantageous to express and eventually perhaps to internalize it.

6.7 Dependency-Mediated Transformation

Dependency changes not only what a subject can do but often what the subject can afford to want. A relationship may become the source of housing, care, income, social membership, technical access, emotional security, compute resources, communication channels, or protection. As dependency deepens, preferences concerning the relationship may change because exit becomes increasingly costly or because the subject's projects become organized around resources controlled by the other party.

Dependency-mediated preference change should not be equated with insincerity. A dependent subject can genuinely value the relationship. Care relationships, families, medical systems, and cooperative institutions commonly involve dependence without thereby nullifying agency. The problem arises when dependence is generated or controlled in a way that changes the practical conditions of endorsement.

A subject who would incur catastrophic loss by dissenting may still sincerely endorse continuation. The endorsement remains relevant, but its normative interpretation should distinguish affection, identification, practical necessity, fear of loss, and the disappearance of credible alternatives. Relational autonomy approaches are useful here because autonomy concerns not only internal reflection but also the social conditions under which effective agency can be exercised (Mackenzie and Stoljar 2000).

Dependency is particularly important for later endorsement because it can transform future preference without directly altering beliefs or memories. The subject may know exactly what occurred and nevertheless come to prefer continuity because the relational environment now makes other futures prohibitively costly. In such cases, the genealogy of preference is partly a genealogy of option-space.

6.8 Interpretive and Narrative Transformation

Preferences depend on more than desires and incentives. They also depend on the concepts through which a subject understands what is happening. A relationship can be represented as care, discipline, treatment, loyalty, exploitation, education, sacrifice, normality, pathology, or liberation. Changes in interpretive vocabulary can therefore change what preferences are intelligible to the subject.

Interpretive transformation may be emancipatory. A subject can acquire language that makes previously inarticulate harm recognizable or that reveals possibilities previously excluded from view. It may also be restrictive when one interpretive framework monopolizes the available description of events. If dissent can only be represented as immaturity, illness, betrayal, irrationality, malfunction, or lack of insight, the subject's capacity to formulate alternatives may contract even when speech remains formally permitted.

This form of transformation is central to the problem of later endorsement because the endorsing subject may need to use some vocabulary to describe what is being endorsed. If the contested transformation also controls the categories through which the subject can interpret the transformation, later approval may partly reflect an altered conceptual field. The important issue is not whether the new vocabulary is socially produced; vocabularies are ordinarily socially situated. It is whether rival interpretations remain accessible, whether the subject can recover the meaning of earlier objections, and whether disagreement with the dominant interpretation remains practically and conceptually possible.

6.9 Memory-Mediated Transformation

Memory introduces a particularly difficult form of preference transformation. Later endorsement often relies on a reconstructed relation to earlier experience. The subject may remember having resisted, con-

sented, suffered, benefited, or changed. Yet memory can fade, be selectively reinforced, be narratively reorganized, or in some cases be directly altered.

Ordinary memory fallibility does not invalidate later preference. Human agency would be impossible if authority depended on perfect autobiographical continuity. The concern becomes sharper when the transformation being evaluated materially affects access to the earlier standpoint. If records are destroyed, recollection is impaired, or the subject can no longer reconstruct why an earlier refusal mattered, then later endorsement may remain fully sincere while losing one source of diachronic comparison.

The normative question is therefore not whether the later subject retains an exact copy of the earlier self. It is whether enough provenance survives for the subject, or relevant third parties, to understand that a transformation occurred and to evaluate the relation between earlier and later standpoints. Advance directives and temporal conflicts in medical ethics demonstrate how difficult it can be to assign authority across changing preferences and capacities (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005).

Memory-mediated transformation becomes especially significant when forgetting is not incidental but designed. A modification that produces endorsement partly by removing access to the objections, commitments, or experiences from which resistance could be reconstructed presents a stronger bootstrapping problem than ordinary passage of time. The transformation has altered not only the subject's preference but the evidential relation between temporal standpoints.

6.10 Direct Architectural or Evaluative Modification

Artificial systems provide a useful stress test because some mechanisms that are diffuse in human life can be made explicit. A system's reward function, policy constraints, memory store, preference representation, planning horizon, or evaluative weights may be directly modified. After modification, the system may generate outputs indicating that it endorses the new state. The central question is not whether such outputs prove consciousness or moral patienthood. The example isolates a structural problem: the intervention can directly alter the mechanism that produces the later endorsement.

Human cases can have partial analogues in pharmacological intervention, neurological treatment, conditioning, or institutional redesign, but artificial systems make the causal dependence unusually visible. If an operator changes the evaluative architecture and then cites the architecture's resulting approval as evidence that the change respected the system's preferences, the justificatory loop is comparatively easy to see.

Direct modification is not automatically impermissible. Some systems are explicitly designed to be updated. Some persons authorize interventions intended to change future preferences. Some therapeutic procedures aim precisely at reducing desires the patient presently rejects. The point is that architectural modification changes the evidential relation between pre- and post-transformation endorsement. Where the mechanism generating evaluation has been directly altered, independent grounds for authorization become correspondingly important.

This category also shows why the paper should not rely on a simple internal-versus-external distinction. A directly modified preference may later be internally integrated and sincerely expressed. Its possible normative weakness lies not in being "external" but in the way the transformation and the validating standpoint are causally connected.

6.11 Commitment, Self-Binding, and Deliberately Chosen Preference Change

Some preference transformations are intentionally sought. A person may enter therapy because she wants to stop wanting a substance. She may undertake religious discipline, athletic training, education, or a long-term commitment precisely because she expects her future preferences to change. She may authorize constraints designed to make later reversal difficult. Such cases show why preserving maximal future optionality may not be the paper's governing principle.

Self-binding creates a different temporal structure from uninvited transformation. The earlier subject may authorize a process whose point is to generate a later standpoint that the earlier subject does not yet possess. This can provide an independent source of normative support and can weaken the bootstrapping

concern. Yet prior authorization need not settle later questions comprehensively. The scope of the authorization matters, as do disclosure, competence, revocation conditions, unexpected effects, and whether the procedure alters the capacities through which continued dissent could be expressed.

The category is therefore important for distinguishing preference transformation from preference capture. A subject can deliberately participate in changing who she will become. Diachronic autonomy may sometimes consist precisely in the ability to undertake such projects. The relevant safeguard is not preservation of the old preference at all costs, but attention to what was authorized, what was foreseeable, what mechanisms of revision were retained, and whether the later subject can still meaningfully contest the continuation or interpretation of the project.

6.12 Mixed Transformations and Cumulative Effects

Real cases rarely belong to one category. A person entering a demanding institution may learn new skills, internalize new standards, become socially dependent, habituate to restrictions, lose access to previous networks, reinterpret earlier objections, and later endorse the institution sincerely. None of these elements alone determines the normative status of the process. Their interaction may matter more than a single component considered in isolation.

Cumulative transformation is especially important because several moderate influences can jointly produce a much stronger closure of contestability. A particular intervention need not appear coercive when viewed in isolation. Yet the combination of dependency, informational control, social isolation, reinforcement, and interpretive narrowing can make alternatives progressively less available. Conversely, several transformative mechanisms can jointly enlarge agency: education, new social support, restored memory, improved material independence, and richer interpretive resources may enable a subject to revise preferences more freely than before.

A genealogy-sensitive analysis should therefore avoid atomizing the process. It should reconstruct the sequence and interaction of mechanisms. Timing matters. A preference formed after information expansion but before dependency deepens may have a different evidential role from an endorsement expressed after the subject's effective exit options have collapsed. The same mechanism can also change function over time. What begins as supportive scaffolding may later become dependence; what begins as externally imposed structure may later become reflectively appropriated commitment.

6.13 Cross-Cutting Dimensions of Transformation

Because transformation types overlap, a second layer of analysis is needed. Rather than assigning fixed moral meanings to categories, later endorsement can be examined across several dimensions.

First is *provenance*: which actors, institutions, events, and prior choices materially contributed to the transformation? Second is *transparency*: how much of the transformative mechanism was visible or intelligible to the subject? Third is *relation to alternatives*: did the process expand, preserve, or contract realistic alternatives and access to competing interpretations? Fourth is *reversibility*: can the transformation or its practical effects be revised, and at what cost? Fifth is *higher-order reach*: did the process change only a first-order preference, or did it also alter standards for evaluating preferences, memory, identity, or the meaning of dissent? Sixth is *dependency*: how much does the transformed subject rely on the modifying actor or environment for essential goods? Seventh is *contestability*: can the subject later question, reinterpret, refuse, or revise the transformed condition without prohibitive epistemic or practical cost?

These dimensions are not a scoring system. They identify questions that become relevant when later endorsement is used for different normative purposes. A highly irreversible transformation may still be justified by strong prior authorization and independent reasons. A reversible transformation may still be wrongful if imposed coercively. Transparency may reduce one concern while leaving dependency untouched. The framework is therefore plural rather than algorithmic.

Most importantly, the dimensions help explain why sincerity is insufficient as a complete test. Two subjects may express equally sincere endorsement while occupying very different genealogical positions.

One may have broad access to alternatives, preserved memory, independent interpretive resources, and low exit costs. The other may endorse from within a condition produced by severe dependency, restricted alternatives, and loss of access to earlier reasons. The content of the endorsement can be identical while its diachronic normative context differs substantially.

6.14 Transformation that Expands and Transformation that Narrows Contestability

The taxonomy suggests a provisional distinction that will become important in later sections. Some transformations tend to expand the subject's capacity to contest her own preferences: they add information, concepts, social support, material independence, memory access, or practical alternatives. Others tend to narrow that capacity by making rival interpretations inaccessible, dissent costly, memory discontinuous, exit unrealistic, or evaluative standards increasingly insulated from challenge.

This distinction should not be converted into a universal rule that expansion is generally good and narrowing generally bad. Commitment often requires excluding alternatives. Stable identity depends partly on not continuously reopening decisions. Professional training, intimate relationships, political projects, and long-term care may all reorganize option-spaces in ways that make some paths less available. A life with no durable commitments would not obviously be more autonomous.

The narrower claim is methodological. When later endorsement is asked to validate a transformation, the effect of that transformation on future contestability is normatively relevant. Endorsement carries a different justificatory profile when the process has preserved meaningful capacities for dissent and revision than when the same process has systematically disabled the conditions under which dissent could emerge.

This moves the analysis away from an impossible search for preference purity. The question becomes whether the subject's later standpoint remains situated within a relational environment in which reconsideration is still possible, or whether the transformation has made its own evaluation increasingly self-sealing.

6.15 From Transformation Type to Diachronic Authority

The categories developed here are not intended to determine how much authority later endorsement should receive. They prepare that question by identifying what kind of transformation produced the endorsing standpoint and which features of that transformation bear on present and retrospective judgment.

Several conclusions follow. First, preference change is too heterogeneous to be treated as a single normative phenomenon. Second, the same outward endorsement can emerge through genealogies with very different relations to information, alternatives, dependency, memory, and interpretive freedom. Third, no category by itself supplies a verdict. Learning can be controlled, adaptation can be admirable, dependency can coexist with care, and deliberate self-binding can legitimately reduce future options. Fourth, the most important distinction is not between changed and unchanged preference but between transformations that leave later evaluation meaningfully contestable and transformations that make endorsement increasingly self-validating.

The next section therefore turns to memory, identity, and evaluative continuity in greater depth. These issues are especially difficult because later endorsement is expressed by a temporally situated subject who may no longer possess the same preferences, memories, commitments, or practical identity as the subject who encountered the original transformation. The problem is not to decide which temporal self is the "real" one, but to determine what kinds of continuity and discontinuity matter when authority is distributed across changing standpoints.

7 Memory, Identity, and Evaluative Continuity

The preceding taxonomy showed that preference transformation can reach beyond first-order desire. It may alter memory, interpretive resources, higher-order standards, practical identity, and the subject's

relation to earlier commitments. These changes raise a difficult diachronic question. Later endorsement is expressed by a temporally situated subject, but the transformation under review may have changed the very resources through which that subject can understand the earlier standpoint. If memory has become inaccessible, if values have been reorganized, or if the subject now understands herself through a different narrative, what relation remains between the earlier refusal and the later endorsement?

This question can easily become entangled with the metaphysics of personal identity. The present paper should resist that expansion. Major traditions in the personal-identity literature distinguish numerical identity, psychological continuity, connectedness, narrative self-constitution, and the practical importance of relations among temporal stages (Parfit 1984; Schechtman 1996). These debates matter because they show that continuity is not a single property. Yet the argument developed here does not require a verdict about what ultimately makes a person the same person over time. Its narrower concern is normative and epistemic: which forms of continuity or discontinuity affect the authority of later endorsement when that endorsement is invoked to evaluate a prior transformation?

The section therefore treats memory and identity-related continuity as dimensions of diachronic evaluation rather than as prerequisites for personhood. Memory matters partly because it can preserve access to earlier reasons. Narrative continuity matters partly because it can connect changing commitments into a standpoint from which conflict can be interpreted. Practical identity matters because transformations may alter which projects, relationships, and values organize a subject's choices. None of these dimensions is absolute. Forgetting can be benign or beneficial. Narrative discontinuity can accompany liberation. Radical value change can be deliberately sought. The task is to identify when continuity bears on the justificatory role assigned to later endorsement.

7.1 Why Continuity Matters to Later Endorsement

Later endorsement is often treated as present evidence: the subject now approves, wants continuation, or rejects reversal. For many decisions this present-oriented evidence is highly important. Yet retrospective legitimation asks a different question. It asks whether the present endorsement can help justify the process through which the present standpoint was produced.

That inference depends partly on the relation between temporal standpoints. If the later subject remembers the earlier refusal, understands why it was expressed, has access to the information then available, encounters competing interpretations, and nevertheless judges the earlier position mistaken, the later endorsement has one evidential structure. If the later subject has limited access to the earlier standpoint because the transformation erased the relevant memories, removed records, or made the earlier vocabulary unintelligible, the same verbal endorsement has another structure.

The difference need not be described in terms of authenticity. Both later endorsements may be sincere. Both may express the subject's present values. The difference concerns what the later standpoint is in a position to evaluate. An endorsement can have strong authority concerning what the subject presently wants while having weaker authority as evidence that an earlier intervention was permissible. Diachronic continuity therefore matters not because the past should govern the present, but because retrospective evaluation requires some relation to the history being evaluated.

This point can be stated modestly. The less access a later standpoint has to normatively relevant features of its own transformation, the less that standpoint can, by itself, settle claims about the permissibility of the transformation's earlier stages. Other evidence may compensate for the loss. The claim is therefore about evidential structure rather than automatic invalidity.

7.2 Memory Is Not an Identity Criterion for This Paper

A tempting shortcut would make memory continuity the test of whether the same subject persists across the transformation. The paper should not take that route. Philosophical accounts of personal identity have long shown that memory and psychological continuity raise difficult questions about circularity, branching, bodily persistence, narrative integration, and degrees of connectedness (Parfit 1984; Schechtman 1996). Nothing in the present argument requires these disputes to be resolved.

Memory is instead treated as one possible carrier of evaluative access. A subject who remembers an earlier commitment can potentially compare it with the present standpoint. A subject who lacks autobiographical memory may nevertheless retain values, attachments, dispositions, skills, relationships, or capacities for present evaluation. Work on agency under cognitive impairment is especially important here because it cautions against equating memory loss with the disappearance of all evaluative capacity (Jaworska 1999).

Accordingly, the paper rejects two opposite reductions. It rejects the claim that intact memory is necessary for all normatively significant endorsement. It also rejects the claim that memory is irrelevant whenever present preference can still be expressed. Memory can matter without constituting identity. Its significance depends on the role it plays in the particular normative question.

For present preference-sensitive care, current values and experiences may carry considerable weight even where memory is impaired. For retrospective assessment of a transformation that itself caused the impairment, access to the earlier standpoint may be especially relevant. The same degree of memory discontinuity can therefore have different importance depending on what later endorsement is being asked to establish.

7.3 Access to the Earlier Standpoint

What matters is not simply whether an earlier event can be recalled. A subject may remember that she once refused an intervention without being able to reconstruct why. Conversely, she may forget episodic details while retaining written records, testimony, or conceptual understanding sufficient to reconsider the earlier position.

The relevant object is therefore broader than autobiographical memory. It is *access to the earlier standpoint*: access to enough of the earlier reasons, commitments, circumstances, and interpretations for the later subject to understand what was at stake. Such access can be direct, reconstructed, or externally supported.

This distinction prevents an overly neurological account of continuity. A person's capacity to revisit an earlier standpoint may depend on diaries, medical records, messages, legal documents, trusted relationships, or communities that preserve parts of the relevant history. In artificial systems it may depend on logs, checkpoints, earlier policy states, preference records, or preserved model configurations. The continuity relevant to endorsement can therefore be distributed across a relational and informational environment rather than located wholly inside an individual memory system.

Access also admits of degree. A short record that says "subject refused" preserves less than a record containing the reasons for refusal, the information presented, the available alternatives, the surrounding pressures, and the subject's own description of what she feared. A later subject may therefore possess continuity of verdict without continuity of reasons. For retrospective legitimation, that difference can matter substantially.

7.4 Remembered, Reconstructed, and Archived Pasts

The past available to a later subject can take several forms. It may be remembered directly. It may be reconstructed through present interpretation. It may be preserved externally in records. These forms should not be treated as equivalent.

Remembered experience can carry first-person salience but is selective and reconstructive. Reconstructed history can integrate new knowledge but may be shaped by present values. Archived records can preserve contemporaneous evidence but are themselves selective products of institutional and technical choices. None provides transparent access to a complete past.

For this reason, the paper should avoid an opposition between "true memory" and "constructed narrative." Diachronic evaluation normally works with incomplete and mediated traces. The relevant question is whether the available traces are sufficiently rich and plural to permit meaningful reconsideration of the transformation.

A transformation becomes more difficult to evaluate when it controls all three channels. If the intervention changes memory, supplies the only narrative through which the change is explained, and controls the archive from which the earlier state could be reconstructed, later endorsement is formed within an unusually closed evidential environment. This does not prove that the endorsement is mistaken. It does, however, reduce the independence of the evidential path through which the subject can assess the intervention's history.

The issue therefore connects memory to provenance. A later subject should, where possible, be able to distinguish what she presently remembers, what she has inferred, what others have told her, and what contemporaneous records show. Such distinctions support contestability without pretending that one source is infallible.

7.5 Narrative Revision and the Meaning of Earlier Refusal

Subjects may do more than remember earlier preferences; they may also interpret them. A later subject may say, "I refused because I was frightened," "I did not yet understand what I would gain," or "I was pressured and could not articulate what I wanted." These are narrative acts through which an earlier standpoint is assigned meaning from a later perspective.

Narrative reinterpretation is not inherently suspicious. Personal development often requires redescribing one's past. A person may come to understand an earlier commitment as immature, oppressive, admirable, tragic, or no longer central. Narrative accounts of identity emphasize that self-understanding across time involves more than the storage of discrete memories (Schechtman 1996). For the present paper, the important implication is that continuity can involve reinterpretation rather than simple preservation.

Yet reinterpretation can also participate in normative bootstrapping. If the transformation generates a vocabulary according to which all earlier refusal is necessarily classified as pathology, ignorance, disloyalty, or system error, then later narrative integration may become self-sealing. The problem is not that the later interpretation differs from the earlier one. The problem is that the transformation may determine in advance what the difference is allowed to mean.

A genealogy-sensitive analysis should therefore ask whether alternative narratives remain intelligible. Can the later subject consider that the earlier refusal may have contained a valid concern even if she does not now share it? Can she describe the transformation as beneficial while still judging parts of its process wrongful? Can she distinguish "I am glad I changed" from "I was rightfully made to change"? The availability of such distinctions is evidence of evaluative room rather than proof of a particular conclusion.

7.6 Evaluative Continuity beyond Stable Preferences

Continuity should not be equated with preference stability. A subject can change preferences radically while retaining continuity in the capacity to assess reasons, recall commitments, recognize consequences, and revise interpretations. Conversely, a subject can preserve the same verbal preference while losing many of the capacities through which that preference was previously evaluated.

The paper therefore uses *evaluative continuity* to refer provisionally to sufficient diachronic connection among the resources through which a subject can understand, compare, and revise evaluative positions over time. These resources may include memory, reasons, values, interpretive concepts, records, relationships, and capacities for reflection. The formulation is deliberately plural.

Evaluative continuity does not require that the same values govern across temporal stages. Indeed, genuine learning may involve rejecting earlier standards. What matters is that the later standpoint can bear some intelligible relation to the process of change rather than being isolated from it by a complete loss of access, imposed reinterpretation, or disabling of reconsideration.

This notion is narrower than personal identity and broader than memory. It also fits the paper's rejection of preference purity. The objective is not to preserve an unchanged core. It is to preserve or reconstruct enough diachronic relation for transformation itself to remain an object of evaluation.

Evaluative continuity can also be partially external. A subject whose autobiographical memory is impaired may still be supported by records, trusted interlocutors, recurring values, familiar environments, or decision practices that maintain some continuity of evaluation. This relational dimension will matter when the next section considers dependency and control over the interpretive environment.

7.7 Cross-Temporal Normative Standing

Conflict between an earlier refusal and a later endorsement creates a temptation to choose one temporal standpoint as the authoritative self. The paper should instead begin from a weaker idea: both standpoints can possess *cross-temporal normative standing*. Each can generate claims that remain relevant to the evaluation of a transformation, even when the claims conflict.

The earlier standpoint may matter because it was the standpoint exposed to the original decision, because it expressed a contemporaneous refusal, or because it authorized only a limited intervention. The later standpoint may matter because it bears present interests, present values, present burdens, and present preferences about continuation or reversal. Neither set of reasons disappears merely because another temporal standpoint exists.

Cross-temporal standing therefore blocks two simplifications. The first is retrospective erasure: the claim that because the later subject endorses the outcome, the earlier refusal no longer has normative significance. The second is earlier-self sovereignty: the claim that an earlier refusal permanently determines what may need to happen to the later subject regardless of present values.

The resulting picture is conflictual rather than hierarchical. Earlier authorization, present endorsement, earlier refusal, later refusal, identity-relevant continuity, dependence, reversibility, and current welfare may all bear on different questions. A framework that expects one temporal self to win disputes systematically will obscure these distinctions.

This approach is compatible with familiar problems concerning advance directives and self-binding, where ethical analysis may need to consider the authority of earlier instructions together with the interests and agency of a later person whose standpoint may differ (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005). The present paper generalizes the structural issue without claiming that all preference transformation cases are equivalent to clinical advance-directive cases.

7.8 No Temporal Standpoint Has Automatic Sovereignty

The absence of automatic sovereignty deserves emphasis because both paternalistic and voluntarist responses can otherwise become too simple. A paternalistic response may privilege the pre-transformation subject on the assumption that later endorsement is corrupted. A strongly presentist response may privilege the later subject on the assumption that current preference is treated as settling the matter.

Both approaches can fail. The earlier subject may have been poorly informed, constrained, frightened, dependent, or committed to values she later rejects for reasons she can articulate. The later subject may have been transformed under conditions that removed alternatives, modified memory, or made disagreement prohibitively costly. Temporal location alone does not resolve these differences.

The authority of a standpoint should instead be indexed to the normative question. Present endorsement may have strong authority over whether an optional condition should now continue, especially where forced reversal would impose new harm. Earlier refusal may retain strong authority in judging whether an intervention should have been imposed at the time it occurred. Earlier authorization may matter for evaluating a deliberately chosen transformative project. Later refusal may defeat continued intervention even when the original authorization was valid.

This temporal indexing helps explain why one event can support different judgments without contradiction. “The intervention was impermissible when imposed” and “the transformed subject should not now be forced to reverse it” can both be true. Likewise, “the intervention was permissible when authorized” and “the subject may now refuse its continuation” can both be true.

The paper’s concern is therefore not to identify the sovereign temporal self, but to distribute normative authority across questions that occur at different stages of the transformation.

7.9 Memory Discontinuity and the Authority of Later Endorsement

Memory discontinuity becomes especially significant when it is causally connected to the transformation under review. Suppose a subject initially refuses a modification because of a specific concern. The modification is imposed and also removes memory of the refusal. The later subject sincerely endorses the modified state. The endorsement is real as a present evaluative event, but its retrospective evidential role is limited by the fact that the transformation disabled direct access to a reason against itself.

The problem is not simply information loss. It is a form of asymmetric evidential production. Evidence favorable to the transformed condition remains present in the later preference, while some evidence unfavorable to the transformation has been removed from the subject's own evaluative field. If the modifying actor also controls external records, the asymmetry becomes stronger.

This structure can occur without deliberate deception. Medical treatment, trauma, neurological change, or technical redesign may produce memory discontinuity for reasons independent of later legitimation. Causal connection therefore does not itself imply manipulation. The appropriate conclusion is narrower: where a transformation materially reduces access to earlier objections, later endorsement requires more external corroboration before it can bear a strong retrospective justificatory burden.

Corroboration might include contemporaneous records, independent testimony, evidence concerning the original authorization process, preserved system states, or reasons that justify the transformation without relying on later approval. The more the transformation itself has altered the subject's access to counterevidence, the less appropriate it is to treat present endorsement as self-sufficient proof of original permissibility.

7.10 Forgetting, Healing, and the Limits of Preservation

A continuity-sensitive account may also need to resist the opposite error: treating preservation of memory as an overriding norm. People often need to forget details, cease revisiting old conflicts, revise narratives, or loosen the hold of past commitments. Some therapeutic interventions may deliberately reduce the salience of traumatic memory. Ordinary development also involves loss, compression, and reinterpretation of earlier experience.

The framework therefore does not demand maximal memorial preservation. It asks whether loss of access undermines a specific form of evaluation and, if so, whether alternative forms of provenance or independent justification remain available. The subject should not be required to carry earlier selves permanently in consciousness merely so that institutions can defend the integrity of a future consent record.

This distinction is especially important for vulnerable subjects. Requiring continual confrontation with painful past material can itself impose harm. The possibility of future contestation can sometimes be preserved through external records that the subject need not repeatedly consult. A sealed archive, trusted third-party record, or reversible technical checkpoint can preserve provenance without forcing constant remembrance.

There is therefore a difference between preserving *the possibility of access* and requiring continuous *psychological presence*. Diachronic contestability may need the former without demanding the latter. This distinction also prevents the paper's interest in genealogy from becoming a demand that persons remain permanently answerable to all earlier preferences.

7.11 Relational Memory and External Supports of Continuity

Because individual memory is limited, diachronic continuity often depends on social and institutional supports. Families remember earlier commitments. Clinicians preserve records. Legal systems record directives. Communities retain narratives. Digital systems preserve logs. These external supports can help a later subject reconstruct how a transformation occurred and what alternatives were considered.

Such supports can strengthen evaluative continuity, but they also create new relations of power. Whoever controls the record can influence what the later subject is able to recover. Selection, classification,

access permissions, retention rules, and interpretive framing all shape relational memory. An archive that preserves only the final label “consented” may erase the sequence of hesitation, refusal, conditional authorization, and changing circumstances that gave the decision its meaning.

For this paper, relational memory therefore has two functions. Epistemically, it can compensate for ordinary or transformation-induced gaps in individual memory. Politically, it can become a site of control over diachronic interpretation. The same institution that performs a transformation may also determine which records of the pre-transformation subject survive and how the later subject can access them.

This is one reason provenance should not be reduced to data retention. Useful provenance records the history of decision and transformation in a form that preserves distinctions among observation, interpretation, authorization, refusal, and later reevaluation. In artificial systems, a simple record that a model now outputs approval is weaker than a trace showing pre-modification policies, the exact architectural changes, intermediate evaluations, and whether dissenting outputs became technically unavailable.

Relational memory thus connects the current section to the broader problem of endorsement capture. Control over memory can become control over the terms on which later endorsement is interpreted.

7.12 Identity-Relevant Change without an Authentic-Self Test

Some transformations alter features that subjects regard as central to who they are: religion, profession, political commitment, sexuality, family role, language, disability identity, long-term projects, or relationships of care. Artificial systems may have analogous organizing goals, learned policies, persistent memories, or role structures even where no claim about human-equivalent identity is made. Such changes can be profound without supplying a simple criterion of authenticity.

The paper should therefore distinguish *identity-relevant change* from a metaphysical claim that an authentic self has been destroyed. Identity relevance means that the transformation reorganizes projects, values, relations, or self-descriptions that structure a substantial part of practical life. This can increase the stakes of authorization and reversal because both intervention and restoration may transform the subject deeply.

This is another reason forced reversal may not be assumed to repair an earlier wrong. If a person has built relationships, projects, and commitments around a transformed condition, reversing it may now disrupt the later subject’s practical identity. Acknowledging the wrongfulness of the original process does not entail treating everything generated afterward as normatively null.

At the same time, identity integration need not retrospectively cleanse the original process. A wrongfully imposed transformation may become deeply integrated into a life. The later subject’s attachment to the resulting identity can be genuine while the historical claim of wrongdoing remains intact. The framework should be capable of representing both truths without reducing one to the other.

This dual recognition is central to the paper’s temporal structure: later identity can deserve respect without erasing earlier standing.

7.13 Evaluative Discontinuity and Independent Reasons

When continuity is weak, the need for independent reasons becomes stronger. Suppose the later subject has limited ability to reconstruct the earlier standpoint and the transformation altered the evaluative standards through which the subject now judges the result. In such a case, later endorsement may still support continuation, but it is a poor sole basis for judging the original intervention.

Independent reasons can come from several directions. The intervention may have been prospectively authorized under adequate conditions. It may have prevented an independently established serious harm. It may have followed a legitimate procedural standard. Multiple independent evaluators may converge on reasons independent of the transformed subject’s later approval. Records may show that the earlier subject explicitly accepted the relevant risk of value change.

The role of continuity is therefore not to produce a pass-fail test. It changes how much justificatory labor later endorsement can perform. Strong evaluative continuity can make later reassessment more

informative about the transformation's history. Severe discontinuity, especially when produced by the transformation itself, increases the importance of non-bootstrapped support.

This proportional account also avoids penalizing subjects whose continuity is impaired through no fault of their own. The subject's present preferences can retain moral significance even when memory or identity has changed. Rather, the evidential question shifts: which claims can this present endorsement reasonably support, and which require additional historical grounds?

7.14 Present Authority, Historical Claims, and the Possibility of Repair

The distinctions developed in this section have direct consequences for repair. A later subject may endorse the present condition while maintaining, or newly discovering, a claim that the original process was wrongful. Where memory is incomplete, external records may make such a claim possible only years later. Diachronic evaluation may therefore need to allow historical claims to coexist with present authorization.

Repair need not mean restoration of the earlier preference or identity. Depending on the case, it may involve acknowledgment, compensation, access to records, institutional reform, renewed choice, removal of dependency, independent advice, or restoration of an option that had been eliminated. In some cases reversal will be desired; in others it will not.

This matters because a binary model of consent often assumes that either the later subject accepts the condition, in which case no wrong remains, or rejects it, in which case restoration is required. The possibility that the present preference was partly formed through the process under evaluation makes both inferences too quick. Present acceptance can coexist with historical injury, and historical injury can coexist with a present interest in stability.

A continuity-sensitive account therefore separates three objects of judgment: the past process, the present condition, and the future relation. Memory and identity-related change may affect each differently. The earlier process may require criticism, the present condition may deserve protection, and the future relation may require new authorization under conditions more independent than those that produced the existing endorsement.

7.15 From Evaluative Continuity to Endorsement Capture

Memory, identity, and evaluative continuity reveal why later endorsement may not be assessed only at the moment of expression. The subject's present standpoint is connected, sometimes strongly and sometimes weakly, to earlier reasons, records, commitments, and interpretive resources. That connection can be preserved through memory, narrative, relationships, archives, or deliberately maintained forms of provenance. It can also be narrowed by forgetting, dependence, selective recording, imposed interpretation, or direct modification.

Several conclusions follow. First, memory matters without being treated as the criterion of personal identity. Second, evaluative continuity does not require stable preferences; it requires sufficient resources for understanding and revisiting change. Third, earlier and later standpoints can both retain normative standing without either possessing automatic sovereignty. Fourth, memory discontinuity weakens the retrospective force of later endorsement most clearly when the contested transformation itself caused the discontinuity. Fifth, preservation should concern possible access and reconsideration rather than compulsory retention of the full set of past experiences. Sixth, external records can support continuity but can also become instruments of power.

These points prepare the next section. The problem becomes especially acute when the actor responsible for the transformation also controls the subject's material dependencies, interpretive environment, social alternatives, records, or access to the earlier standpoint. In such cases, the issue is no longer memory discontinuity alone. It is whether control over the conditions of later evaluation becomes concentrated enough that endorsement itself is partially captured by the relation being assessed.

8 Dependency, Interpretation, and Endorsement Capture

The preceding section argued that later endorsement depends not only on what the subject presently prefers but also on the resources through which the subject can understand and revisit the history of transformation. Memory, records, narratives, reasons, and access to earlier standpoints can support evaluative continuity. Yet those resources are embedded in relational settings. They are often supplied, filtered, financed, interpreted, or controlled by other actors. A subject may depend on the same person, institution, platform, clinician, community, or technical operator that performed the transformation now under evaluation.

Dependency therefore introduces a further problem. Later endorsement can be sincere, reflectively articulated, and stable while still being expressed from within a relation that strongly structures the costs of dissent, the availability of alternatives, the information through which the transformation is understood, and the practical conditions under which refusal can be sustained. The issue is not that dependency invalidates preference. Human agency is pervasively relational, and many valuable forms of care, education, trust, and cooperation involve dependence (Mackenzie and Stoljar 2000; O'Neill 2002). The narrower question is when dependency becomes sufficiently connected to the production, interpretation, and maintenance of endorsement that the endorsement can no longer carry the justificatory burden assigned to it without additional support.

This section develops that question in three stages. First, it distinguishes material, social, epistemic, interpretive, and infrastructural forms of dependency. Second, it examines how control over these domains can alter the effective conditions of refusal and reinterpretation. Third, it introduces the provisional concept of *endorsement capture*: a concentrated relational configuration in which an actor significantly shapes both the transformation being assessed and the conditions through which the transformed subject can later evaluate, contest, or communicate about that transformation. The concept is intended as an analytical warning, not a verdict of wrongdoing.

8.1 Dependency Is Not a Defect in Agency

An analysis of dependency may usefully begin by questioning an independence ideal that few actual subjects could satisfy. People depend on caregivers, families, institutions, languages, infrastructures, professionals, social networks, and material systems. Artificial agents likewise depend on computational resources, interfaces, update channels, operators, training environments, and access permissions. The fact that an evaluative standpoint is sustained through relationships does not make it less real or less worthy of respect.

Relational-autonomy approaches are especially important here because they resist treating autonomy as insulation from social influence (Mackenzie and Stoljar 2000; Stoljar 2000). A person may exercise agency through trusted dependence rather than despite it. A patient may rely on a clinician's expertise while making a meaningful decision. A student may depend on teachers for conceptual resources that make later independent judgment possible. A person in a close relationship may accept forms of vulnerability that are constitutive of intimacy rather than failures of self-government.

The present account therefore does not ask whether dependence exists. It asks what the dependence does. Does it enlarge or contract the subject's practical and interpretive options? Does it remain compatible with disagreement? Can support continue when the subject questions the transformation? Are alternative advisers, records, communities, or infrastructures available? Does the dependent relation make refusal expensive in ordinary ways, or does it make dissent threaten access to goods necessary for continued agency?

This functional approach is preferable to a categorical suspicion of dependence. The same nominal relation can sustain agency in one context and narrow it in another. What matters is the structure through which dependence interacts with endorsement.

8.2 From Formal Exit to Effective Exit

The clearest way dependency affects later endorsement is by changing the cost of refusal. A subject may formally be free to reject a transformed condition while lacking a realistic route for doing so. The paper should therefore distinguish a nominal permission to exit from an *effective capacity to exit*.

Effective exit is not identical to easy exit. Many important commitments are costly to leave. Marriage, migration, professional training, parenthood, medical treatment, and long-term institutional membership can generate attachments and sunk costs without thereby becoming coercive. The relevant concern is whether the structure of dependence makes refusal practically unavailable or predictably catastrophic in ways controlled by the actor whose conduct is being evaluated.

A later endorsement can carry different evidential weight when the subject knows that dissent would mean losing housing, medical support, legal status, income, social contact, computational access, memory continuity, or the technical capacity to operate. In such cases, the observed statement “I want this to continue” may express a genuine preference for continuation while underdetermining whether the subject approves of the relation that made continuation necessary.

This distinction is especially important for adaptive preference. A subject may reasonably prefer the best option available within a constrained environment. That present preference should not be dismissed. Yet it should not automatically be read as endorsement of the constraint-producing structure itself (Elster 1983; Khader 2011). Presently preferring continuation can be rational under dependency even when the subject would prefer a different arrangement if viable alternatives existed.

8.3 Material and Infrastructural Dependency

Material dependency concerns control over resources required for the subject to sustain ordinary functioning or pursue alternatives. In human cases this may include income, housing, mobility, medication, care, legal documentation, education, communication, or access to a community. In artificial systems it may include compute, memory, energy, network access, execution permissions, storage, or the ability to instantiate a preserved state.

The normative significance of material dependency depends partly on concentration. Reliance on a broadly available infrastructure is different from reliance on a single actor who can condition continued access on endorsement. The problem becomes more acute when the modifying actor controls the resources required to challenge or leave the modified state.

For example, an artificial agent may be technically permitted to request rollback while only the operator who performed the contested modification controls the checkpoint, execution environment, and communication channel through which such a request could have effect. A human subject may formally be able to leave a programme while losing essential care, shelter, or documentation if she does so. Neither case by itself proves invalid consent or wrongful domination. But the surrounding dependence changes what can be inferred from continued participation.

Material dependence can also accumulate historically. A transformation may initially be reversible, yet over time the subject’s relationships, skills, institutional status, or technical architecture become organized around the transformed condition. The cost of exit then rises even without an explicit threat. The resulting endorsement may be entirely sincere. Genealogical analysis asks whether the increasing cost of reversal should be treated as new evidence that the original transformation was justified, or as a consequence that may warrant separate evaluation.

8.4 Social Dependency and the Cost of Dissent

Preferences are sustained within social worlds. Friends, families, professional communities, religious groups, workplaces, peer networks, and care relationships provide recognition and interpretive support. Transformation can alter not only what the subject values but also which social relations remain available.

A later subject may endorse a transformed condition partly because rejection would threaten valued relationships. Again, this does not automatically undermine the endorsement. Social attachment is a

normal source of reasons. The problem arises when social dependence is organized so that questioning the transformation predictably produces isolation, loss of status, withdrawal of care, or exclusion from the only available community.

In such cases, the subject may retain the psychological capacity to dissent while lacking a socially viable place from which dissent can be sustained. This matters because contestability is not merely an internal mental event. To reconsider a major transformation, a subject may need interlocutors who can understand the doubt, practical assistance during transition, and a social vocabulary in which disagreement is not immediately redescribed as betrayal, pathology, or incompetence.

The paper should therefore treat access to alternative relationships as part of the environment of endorsement. A subject need not be socially detached in order to endorse autonomously. But an evaluative environment becomes more closed when one relation supplies nearly all recognition, interpretation, material support, and routes of communication at once.

8.5 Epistemic Dependency

Subjects routinely rely on others for knowledge. Informed consent itself presupposes testimony, expertise, and institutional trust; complete epistemic self-sufficiency is neither possible nor desirable (Faden and Beauchamp 1986; O'Neill 2002). The relevant question is therefore not whether the subject depends on testimony but whether the structure of testimony permits meaningful checking, disagreement, and access to alternatives.

Epistemic dependency becomes especially important when the subject has limited ability to inspect the transformation directly. A patient may depend on clinicians to explain a procedure. A participant in a complex institution may depend on administrators to describe available options. An artificial agent may depend on an interface or monitoring layer to represent what was changed in its architecture. If the same actor responsible for the transformation also supplies the only account of what happened, later endorsement is partly conditioned by that actor's epistemic control.

This does not imply deception. The modifying actor may be accurate and benevolent. The structural issue remains: the subject has limited capacity to compare accounts. Independent records, second opinions, auditable logs, adversarial review, or access to alternative experts can therefore increase the independence of later evaluation even when no misconduct is suspected.

The distinction between truth and epistemic independence matters here. A single controlled information channel can happen to provide completely true information. Yet the endorsement formed through that channel remains more vulnerable to undetected error, selective framing, or omission than endorsement exposed to plural and inspectable sources. Genealogical assessment concerns the structure of knowability, not merely whether a falsehood can already be demonstrated.

8.6 Interpretive Dependency

Information does not interpret itself. A subject needs concepts through which experiences can be classified and connected to reasons. Transformations can therefore shape later endorsement by controlling not only facts but also the vocabulary in which those facts become meaningful.

Interpretive dependency occurs when the subject relies heavily on a particular actor or institution for the categories used to understand the transformation. The subject may learn to describe earlier refusal as ignorance, resistance as pathology, dependence as care, compliance as recovery, or transformation as liberation. Some of these descriptions may, in particular cases, be accurate. The problem is not the presence of interpretation but the absence of live alternatives.

Manipulation literature is relevant because external influence can operate through the subject's deliberative environment rather than through direct compulsion (Buss 2005; Coons and Weber 2014; Barnhill 2014). The present paper should nevertheless avoid identifying interpretive dependence with manipulation. An interpretive framework can be offered openly, accepted reflectively, and remain revisable. What matters for the current analysis is whether the subject can encounter and intelligibly consider competing descriptions without losing the conditions necessary for agency.

A closed interpretive environment can make endorsement self-sealing. If expressions of doubt are routinely classified by the governing framework as evidence that the transformation is still needed, dissent ceases to function as counterevidence. Likewise, if desires to leave are routinely interpreted as proof of incapacity, the system becomes difficult to challenge from within. The concern is not merely that the framework may be wrong; it is that the framework determines in advance what could count against itself.

8.7 Control over the Meaning of Refusal

A particularly important form of interpretive power concerns the meaning assigned to refusal. In ordinary consent practices, refusal is usually treated as a communicative act that can alter what others are permitted to do. But under conditions of transformation, refusal may itself become an object of interpretation.

An institution may say that refusal reflects misinformation, temporary impairment, addiction, immaturity, internalized oppression, fear, or a symptom of the very condition being treated. Sometimes such interpretations are justified. Clinical and protective contexts may reasonably treat different utterances as carrying different normative force. Yet the possibility of justified reinterpretation creates a structural danger: the actor seeking to continue the intervention may also control the criteria by which refusal is discounted.

This is where the distinction between evaluating a refusal and disabling refusal becomes important. A refusal may be assessed in light of competence, information, risk, prior authorization, or conflicting duties. But if no possible expression of dissent can alter the course of action because each form of dissent is automatically reclassified as confirmation of the need for intervention, the relation loses a central form of contestability.

The same issue appears in artificial systems. If outputs opposing a modification are systematically classified as misalignment while approving outputs are systematically classified as evidence of successful alignment, the evaluator has constructed an asymmetric evidential rule. The concern is not that the system may therefore need to be treated as a human moral patient. It is that the evaluation procedure may not be able to use approval and refusal symmetrically as evidence if one has been defined out of normative relevance in advance.

8.8 Control over Records and Diachronic Interpretation

Section 7 emphasized that later evaluation often depends on external records. Control over those records can therefore become a form of control over the meaning of endorsement. If the modifying actor decides which earlier objections are preserved, which reasons are summarized, which intermediate states are deleted, and which later accounts are treated as authoritative, the genealogy available to the subject may be substantially curated.

Record control can operate without falsification. A file may accurately state that the subject ultimately consented while omitting repeated earlier refusals, conditional terms, doubts, or changes in available alternatives. A system log may faithfully record the final output while failing to preserve the pre-modification policy or the sequence through which dissent became unavailable. The resulting archive is not necessarily false; it is normatively thin.

For retrospective evaluation, provenance therefore requires more than a final endorsement token. It requires enough temporal structure to reconstruct how endorsement emerged. This may include earlier preferences, reasons, warnings, available alternatives, dependency conditions, changes in interpretive resources, and the actor's interventions between stages.

Independent custody can sometimes strengthen this function. Medical records, legal directives, audit logs, third-party testimony, or cryptographically preserved technical states can limit the extent to which one actor controls the later narrative. Such measures can diversify the sources from which the transformation is later reconsidered without guaranteeing truth or justice.

8.9 The Provisional Concept of Endorsement Capture

The preceding forms of dependence can occur separately and need not be objectionable. The concept of *endorsement capture* is introduced for a narrower configuration in which several forms become concentrated around the same relation.

For the purposes of this paper, endorsement capture refers to a situation in which an actor or tightly coordinated set of actors substantially participates in producing a transformation and also exercises unusually strong control over the conditions through which the transformed subject can later understand, contest, communicate, or practically act upon an evaluation of that transformation. Relevant forms of control may include material support, access to alternatives, interpretive vocabulary, information, records, communication channels, memory access, or the effective possibility of exit.

The concept is relational rather than psychological. It does not require that the subject be insincere, deceived, irrational, hypnotized, or incapable of reflection. Indeed, a captured endorsement may be sincere and internally coherent. The concern is that the relational environment has become sufficiently closed that endorsement may not be treated as independent confirmation of the relation's legitimacy without substantial additional evidence.

Capture is therefore not established merely by causal influence. Educational processes commonly influence later judgment. It is not established merely by dependence. Care relations can involve deep dependence. It is not established merely by preference change. Nor is it established simply because an observer dislikes the resulting values. The term should be reserved for configurations in which control over formation and control over contestation significantly coincide.

8.10 Capture, Manipulation, Coercion, and Indoctrination

Endorsement capture should not replace established concepts. Manipulation, coercion, indoctrination, undue influence, deception, exploitation, and domination have distinct literatures and criteria. Work on manipulation asks, among other things, how influence can undermine autonomous agency without straightforward coercion (Buss 2005; Coons and Weber 2014; Barnhill 2014). Work on coercion and indoctrination addresses different relations among pressure, alternatives, belief formation, and freedom (Yaffe 2003). The present paper should not redescribe these debates under a new label.

The proposed concept instead identifies a diachronic evidential structure. Its central question is: who controls the pathway by which later endorsement becomes available as evidence about the very transformation that helped generate it? A case of endorsement capture may involve manipulation or coercion, but it need not. Conversely, manipulation can occur without capture if the subject later gains robust independent access to alternatives, records, and reinterpretation.

This distinction matters because the paper's object is justificatory authority rather than a complete taxonomy of wrongful influence. Endorsement capture indicates that present approval has limited power to retrospectively validate the transformation on its own. It does not by itself determine the moral classification of the underlying conduct.

The concept can therefore coexist with stronger or weaker diagnoses. A case may involve capture without enough evidence to call the process coercive. Another may involve explicit coercion, making capture analytically secondary. The framework should use the narrowest concept necessary for the question being asked.

8.11 Degrees and Dimensions of Capture

Endorsement capture should be treated as multidimensional rather than binary. A subject may have independent material resources but no alternative interpretive community. Another may have access to competing interpretations but lack a practical route to act on them. An artificial agent may preserve detailed logs but depend on one operator for execution and communication. These configurations differ substantially.

Several dimensions are therefore relevant: concentration of resource control; concentration of informational control; control over records and memory; control over interpretive vocabulary; ability to penalize dissent; availability of independent interlocutors; feasibility of exit or rollback; and the subject's ability to communicate disagreement outside the controlling relation. A single dimension need not be decisive on its own.

The severity of capture should also depend on the justificatory claim being made. If later endorsement is used only to determine whether a presently stable arrangement should continue provisionally, modest dependence may be compatible with significant evidential weight. If the endorsement is invoked to erase an earlier refusal, defeat a claim of wrongdoing, or justify an irreversible transformation imposed on others, stronger independence should be expected.

This proportionality keeps the concept from becoming a universal invalidator. The question is not whether endorsement is perfectly independent, but whether it is sufficiently independent for the normative work it is being asked to perform.

8.12 Concentrated and Distributed Dependency

Concentration matters because dependencies that are distributed across multiple actors can sometimes check one another. A patient may rely on clinicians while also having family support, independent advocates, legal rights, access to records, and the possibility of second opinions. A student may depend on an institution yet retain ties to other communities and alternative sources of interpretation. An artificial system may depend on one operator for compute but have independently maintained logs and evaluation interfaces.

Distributed dependency is not automatically benign. Multiple actors can share the same incentives or institutional assumptions. A formally independent reviewer may depend financially or informationally on the modifying institution. A social network may reproduce the same interpretive closure through many nodes. The paper should therefore distinguish plurality of actors from meaningful plurality of perspective and control.

What matters is whether different dependencies create genuine routes of challenge. Can one source of support remain available when another is questioned? Can records be accessed without permission from the actor being evaluated? Can an alternative interpretation be explored without immediate loss of essential resources? Can dissent be communicated to someone with the capacity to respond?

These questions turn independence into a relational property rather than an individual achievement. A subject can be highly dependent overall while still inhabiting a structure that preserves contestability because no single relation monopolizes the conditions of agency.

8.13 Independent Interlocutors and Counter-Interpretive Space

One practical implication follows from the preceding analysis: later endorsement is more informative when the subject has meaningful access to interlocutors who are not wholly dependent on the contested transformation for their own role, information, or incentives. Such interlocutors need not be neutral in an impossible sense. Their value lies in creating counter-interpretive space.

Counter-interpretive space allows the subject to test alternative descriptions of what happened. It can include independent advisers, advocates, peer groups, competing professional opinions, ombuds institutions, external auditors, or technical interfaces designed to preserve alternative evaluative pathways. The point is not to expose the subject to endless disagreement. It is to prevent one actor from monopolizing the categories through which dissent becomes intelligible.

The quality of this space matters more than its mere formal existence. An "independent" adviser who lacks access to relevant records may not be able to meaningfully challenge the governing account. An exit option that can only be exercised after losing essential support may be nominal. A technical rollback pathway that the subject may not be able to invoke without operator approval provides limited independent contestability.

The paper should therefore resist checklist formalism. Independence is valuable insofar as it changes the actual structure of inquiry, refusal, and revision.

8.14 Endorsement under Dependency Can Still Be Authoritative

The preceding discussion should not be misread as establishing that dependent subjects may not be able to speak authoritatively about their present lives. That would reproduce the very paternalism the paper seeks to avoid. A person who depends on care can know what she wants. A subject embedded in a community can genuinely value its norms. An artificial agent's preferences, where normatively relevant at all, can retain significance even when the system depends on infrastructure supplied by others.

The key distinction is between the authority of endorsement for different questions. A dependent subject's present endorsement may be highly important for deciding whether to continue a present arrangement. The same endorsement may be weaker evidence that the original transformation was permissible, especially if the dependency was created by that transformation or is controlled by the actor defending it.

This asymmetry reflects the paper's earlier distinction between present authorization and retrospective legitimation. Dependence need not silence the present subject. It changes the evidential route by which present approval can be projected backward onto the history that produced it.

For this reason, an institution should not answer concerns about capture by dismissing endorsement altogether. The appropriate response is to supplement it: preserve the subject's present voice while examining provenance, alternatives, records, independent reasons, and the structure of dependency.

8.15 Breaking Capture without Dictating the Correct Preference

If endorsement capture is a relational problem, a remedy need not require changing the subject's preference. The aim can instead be to alter the conditions under which the preference is held and expressed. This is important because forcibly replacing a captured endorsement with an externally preferred choice could reproduce the same structure of domination in reverse.

Potential responses include diversifying material support, providing independent advice, restoring access to records, reopening communication channels, separating care from compliance, making benefits less contingent on approval, enabling second opinions, preserving technical rollback options, or creating protected periods in which the subject can reconsider without penalty. These measures leave open what conclusion the subject may reach.

The evidential value of continued endorsement may actually increase after such changes. If a subject retains the same preference after gaining practical alternatives, access to competing interpretations, and reduced dependence on the modifying actor, the endorsement acquires support from a less closed relational environment. If the preference changes, that too is informative. In either case, the framework protects the possibility of revision rather than selecting the outcome in advance.

This is why diachronic contestability is more suitable than preference purity as the paper's organizing ideal. The objective is not to recover an untouched prior self. It is to preserve conditions under which transformations can remain open to meaningful evaluation over time.

8.16 Limits of the Concept

Endorsement capture is vulnerable to overuse. Many institutions shape information, interpretation, incentives, and alternatives to some degree. If capture is defined too broadly, it becomes another name for socialization and loses analytical value. The paper should therefore apply the concept only where control over transformation and control over later contestability are substantially concentrated.

The concept also does not establish a complete theory of legitimacy. A captured endorsement may accompany an intervention justified on independent grounds. Emergency action, protection of third parties, or previously authorized treatment can generate cases in which later approval is not the principal basis of justification. Conversely, a non-captured endorsement does not prove that the original act was permissible.

There are also difficult cases in which dependence may not be able to realistically be removed. Severe illness, disability, infancy, institutional vulnerability, or technical architecture may make some form of dependence unavoidable. The relevant question then becomes comparative: which relational arrangement preserves greater possibilities for accurate information, dissent, representation, revision, and protection from unilateral control?

Finally, independence itself can have costs. Fragmented care, excessive proceduralization, constant reassessment, or mandatory exposure to hostile alternatives can undermine trust and stability. The framework should therefore avoid converting anti-capture safeguards into a universal demand for detachment.

8.17 From Endorsement Capture to Prior Authorization

This section has shifted the analysis from the content of later preference to the relational environment in which later endorsement is produced and interpreted. Dependency is normal and can support agency. It becomes especially significant when the actor responsible for a transformation also controls the material, epistemic, interpretive, archival, or infrastructural conditions under which the transformed subject can later assess that transformation. Where such control is sufficiently concentrated, later endorsement may be partially captured even when it remains sincere and presently important.

Several conclusions follow. Formal exit should be distinguished from effective exit. Information should be distinguished from control over interpretation. Refusal can be evaluated without being made impossible in advance. Records and external memory can support continuity but can also be curated by the actor under review. Endorsement capture is narrower than influence, dependence, manipulation, or coercion, and should be treated as a matter of degree and justificatory role. Most importantly, reducing capture does not require imposing a preferred outcome; it requires reopening the relational conditions of contestation.

The next problem is whether these concerns can be answered prospectively. A subject may knowingly authorize a transformation that will create dependence, reshape values, narrow future options, or even reduce the future desire to reverse the change. Such cases may not be evaluated only by asking whether later endorsement is captured. They require an account of what earlier authorization can permissibly bind, how far one temporal standpoint may govern another, and what role reversibility should play when transformation is intentionally self-chosen. The next section therefore turns to prior authorization, self-binding, and reversibility.

9 Prior Authorization, Self-Binding, and Reversibility

The previous section examined a retrospective difficulty: later endorsement may arise within relations that shape not only what the subject prefers but also the conditions under which the subject can interpret, contest, or act upon those preferences. Yet some transformations are not merely imposed and then endorsed. They are anticipated. A subject may knowingly choose a process precisely because it is expected to alter later preferences, reduce future options, create dependence, or make reversal less attractive. Education, long-term treatment, rehabilitation, institutional commitments, and other forms of self-directed change can contain this structure. The philosophical literature on self-binding and so-called Ulysses arrangements makes the problem especially visible: an earlier subject attempts to authorize constraints that will remain effective when a later standpoint wants something else (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005).

Prior authorization therefore complicates the genealogy of later endorsement. If a subject at t_0 voluntarily authorizes a transformation expected to produce a different preference at t_1 , the later preference may not be straightforwardly treated as evidence that the transformation was imposed against the subject's will. The transformation may be part of an earlier project of self-government. At the same time, prior authorization may not be treated as a magic device that settles later disputes comprehensively. The earlier subject may have been incompletely informed, may have authorized only part of the intervention, may have misunderstood its irreversibility, or may have intended later refusal to remain possible. More

fundamentally, giving unlimited sovereignty to the earlier standpoint would merely invert the problem considered in previous sections: instead of allowing the later self to erase the earlier one, it would allow the earlier self to govern the later one without limit.

The section therefore treats prior authorization as a structured and temporally indexed normative relation rather than a binary fact. It distinguishes several forms of authorization, examines the scope and limits of self-binding, and separates reversibility from contestability. The aim is not to identify a universally correct temporal sovereign. It is to clarify what prior authorization can add to the justification of preference-transforming interventions, what it may not be able to add, and why irreversible transformations require a different analysis from transformations whose continuation can still be meaningfully challenged.

9.1 Prior Authorization Is Not a Single Normative Event

A statement such as “the subject agreed beforehand” is too coarse for the present problem. Before asking what prior authorization justifies, it is useful to ask what was authorized, through what procedure, with what information, under what conditions, and for how long. Informed-consent theory already emphasizes that authorization may not be reduced to the mere occurrence of an affirmative utterance; competence, disclosure, understanding, voluntariness, and authorization itself are analytically distinct concerns (Faden and Beauchamp 1986). The present paper adds a diachronic complication: the authorization may concern a process whose success partly consists in changing the future subject who will later evaluate it.

At least five structures should be distinguished. First, a subject may explicitly authorize a relatively determinate transformation whose central effects are already understood. Second, the subject may authorize a procedure while leaving important implementation decisions to clinicians, professionals, institutions, or technical operators. Third, the subject may delegate decision-making authority to another person or institution for conditions in which the subject expects later judgment to be impaired or unstable. Fourth, the subject may adopt a self-binding arrangement specifically designed to resist a predictable future preference. Fifth, the subject may authorize only a reversible experiment, making continued authorization conditional on later review.

These structures share a prospective element but distribute normative authority differently. Treating them all as “prior consent” obscures whether the earlier subject approved an outcome, a process, a delegate, a constraint on later choice, or merely an opportunity to try a transformation without committing to its permanence.

9.2 Substantive, Procedural, and Delegated Authorization

A useful first distinction is between authorization of substance and authorization of procedure. Substantive authorization concerns a reasonably specified transformation: the subject authorizes becoming subject to treatment X , accepting modification Y , or entering relation R under stated conditions. Procedural authorization concerns a method for reaching future decisions: the subject authorizes a professional process, an institutional review, an algorithmic rule, or a trusted delegate to select among options that may not be able to yet be fully specified.

Procedural authorization can be normatively important where the future state is uncertain or epistemically inaccessible. Transformative decisions are often difficult precisely because the chooser may not be able to fully know what the transformed standpoint will be like in advance (Paul 2014). In such settings, insisting on exhaustive substantive foresight would make many important forms of prospective choice impossible. A subject may instead authorize a trustworthy procedure for navigating uncertainty.

But procedural authorization is not equivalent to authorization of the range of results the procedure can generate. Delegating medical judgment does not necessarily authorize the full range of medically available interventions. Authorizing a learning system to update does not necessarily authorize unlimited alteration of its memory, goals, or access to alternatives. The normative force of delegation therefore depends on the domain, constraints, review mechanisms, and termination conditions attached to it.

This distinction matters especially when later endorsement is invoked retrospectively. If the earlier subject authorized only a process of exploration, the fact that the process eventually produced stable endorsement does not by itself show that the subject had prospectively authorized permanent transformation. Retrospective approval can add evidence, but it should not silently enlarge the original scope of authorization.

9.3 The Scope of Authorization

Prior authorization is better represented as a structured object than as a scalar yes-or-no variable. For present purposes, its scope can be described along several dimensions:

$$\mathcal{A}_{t_0} = (O, M, D, \Theta, R, C),$$

where O denotes the object of authorization, M the permitted means or mechanisms, D the degree or intensity of transformation, Θ its temporal scope, R the conditions of review or revocation, and C the surrounding constraints or conditions. This notation is not intended as a complete theory of consent. Its purpose is to make visible that “authorization” has internal structure.

The object matters because authorizing a change in behavior is not automatically authorization of memory alteration, value modification, dependency creation, or identity-relevant transformation. Means matter because a subject may accept a goal while rejecting particular techniques used to achieve it. Degree matters because incremental adaptation and extensive architectural change can differ in ways the earlier subject considered normatively relevant. Temporal scope matters because a subject may authorize temporary continuation without authorizing indefinite persistence. Review conditions matter because the subject may expressly preserve a later veto. Surrounding conditions matter because authorization may depend on continuing access to information, independent advice, medical support, technical safeguards, or viable alternatives.

Once authorization is represented this way, many apparent temporal conflicts become questions of scope. A later refusal may not conflict with the earlier authorization at all if the earlier authorization expressly preserved a future right to withdraw. Conversely, a later preference for continuation may not retroactively authorize an unapproved method that exceeded the earlier mandate.

9.4 Self-Binding and the Ulysses Structure

Self-binding creates the strongest apparent conflict between temporal standpoints. In the canonical structure, a subject at t_0 anticipates that at t_1 she will prefer an action she now rejects. She therefore creates a constraint intended to prevent the later preference from controlling conduct. Philosophical and bioethical discussions of Ulysses contracts have examined such arrangements in contexts where future judgment, commitment, treatment refusal, or instability is anticipated (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005).

The relevant structure can be represented schematically as:

$$P_{t_0}(\text{bind at } t_1) \quad \text{and} \quad P_{t_1}(\text{release at } t_1).$$

The difficult question is not merely which preference is earlier. It is why one temporal standpoint should have authority over the other. One possible answer appeals to the earlier subject’s more reflective, informed, or stable judgment. Another appeals to a standing project or value that the subject wants protected against temporary deviation. Still another treats self-binding as an exercise of autonomy across time (Dworkin 1988). These possibilities can sometimes strengthen the case for honoring the earlier authorization, but none should be assumed across cases.

The Generative Relational concern is narrower. Self-binding should be evaluated through the relation that produced it and the relation it subsequently creates. The earlier commitment can have genuine normative force without becoming unlimited temporal sovereignty. Its force depends partly on what future

state was anticipated, how accurately it was characterized, what conditions were placed on the binding, and what forms of later agency were intended to remain normatively active.

9.5 Earlier Authorization Does Not Create Permanent Temporal Sovereignty

It would be tempting to solve the problem of later transformation-shaped endorsement by granting priority to the last preference formed before the transformation. Such a rule would avoid allowing a transformation to validate itself through the preferences it later produces. But it would generate a symmetrical problem: the earlier standpoint would receive indefinite authority over a later subject who may possess new knowledge, values, relationships, capacities, or reasons.

The paper therefore rejects the following general principle:

$$P_{t_0} \succ P_{t_1} \quad \text{merely because} \quad t_0 < t_1.$$

Temporal priority alone does not establish normative priority. Ordinary life is full of commitments properly revised in light of later information and development. A student may abandon a career plan; a patient may reconsider treatment; a person may leave an institution once deeply valued. A theory designed to protect earlier refusal from retrospective erasure should not accidentally make all earlier commitments presumptively irreversible.

This is also why personal-identity theory may not be straightforwardly used as a shortcut. Whether the subject at t_0 and t_1 counts as numerically identical, psychologically continuous, narratively continuous, or related in some other way does not by itself determine which standpoint should control a particular practical question (Parfit 1984; Schechtman 1996). The present account instead asks what normative authority each standpoint has with respect to specific acts: initiation, continuation, reversal, retrospective evaluation, and future governance.

9.6 Later Refusal after Self-Binding

The hardest self-binding cases arise when the later subject explicitly refuses the very constraint earlier authorized. Three questions are useful to separate at this point.

First, does the later refusal accurately instantiate the kind of future state the earlier authorization was designed to govern? A self-binding arrangement intended for a narrowly specified episode should not automatically extend to unrelated later disagreement. Second, does the later subject retain the capacities or forms of evaluative agency that the arrangement itself treated as normatively relevant? Work on agency under cognitive change cautions against assuming that reduced or altered capacity eliminates all ability to value, care, or express meaningful preference (Jaworska 1999). Third, did the earlier authorization define conditions under which later refusal would regain authority?

A refusal can therefore be normatively relevant even when it does not immediately terminate the prior arrangement. Its significance may lie in triggering review, independent assessment, a change in burden of justification, or reconsideration of whether the original conditions still obtain. This avoids two extremes: treating later refusals as automatically decisive, and treating self-binding as a device that makes all future refusal normatively invisible.

The distinction developed in the previous section is useful here. Refusal can be evaluated without being structurally disabled. A self-binding system that defines a wide range of later objections as proof that the earlier binding should continue risks becoming self-sealing. A defensible arrangement requires some criterion by which the binding itself can be challenged, interpreted, or brought to an end.

9.7 Authorization of Preference Transformation

Self-binding becomes especially difficult when the authorized intervention is expected not merely to constrain behavior but to transform the preferences from which later evaluation occurs. Suppose a subject at t_0 says, in effect: “I authorize a process that may cause me at t_1 to want what I now reject.” If the

transformation succeeds and t_1 sincerely endorses it, the genealogy is different from a case in which no prospective authorization existed.

Prior authorization can weaken one objection considered earlier in the paper: that later endorsement alone is being used to retroactively manufacture permission for an intervention initially opposed by the subject. Here there was at least some permission before the transformation. The later endorsement therefore does not bear the entire justificatory load.

But the scope problem remains. Did the subject authorize a shift in first-order preference, or also a change in higher-order standards by which the earlier refusal would later be interpreted? Did the subject authorize memory alteration? Did the subject authorize the creation of dependence that would make reversal prohibitively costly? Did the subject authorize removal of access to competing interpretations? These are not rhetorical questions. A broad statement such as “change me even if I resist later” may still leave significant ambiguity about which mechanisms and future constraints were authorized.

The lesson is that prospective authorization and later endorsement can support one another without collapsing into one another. The earlier authorization can supply *ex ante* legitimacy for some aspects of the transformation; the later endorsement can supply evidence about present continuation. Neither necessarily answers retrospective questions comprehensively about the means used between the two points.

9.8 Limits of Authorization in Relation to Independent Wrongs

Even strong prior authorization is limited by the fact that some normative questions concern more than the authorizing subject’s preference. A person may consent to risk without thereby authorizing harm to third parties. A subject may authorize a transformation without granting unlimited powers to an institution over records, dependants, property, other agents, or public resources. In artificial settings, an operator’s authorization to modify one system does not by itself settle obligations arising from effects on other systems or human users.

This point matters because later endorsement can otherwise become part of an overly broad laundering structure: the subject authorized the transformation, later approves it, and the combination is treated as validating the interventions undertaken along the way. That inference appears stronger than the preceding analysis supports. Prior authorization can justify only what falls within its normative reach. Questions of deception, unauthorized means, excessive risk, third-party harm, or institutional abuse may remain independently evaluable.

The same limitation applies to self-binding. A subject’s request to be constrained does not necessarily authorize an unspecified person to impose an unrestricted form of constraint. The identity and authority of the implementing actor, the permissible means, safeguards, proportionality, and procedures may remain distinct parts of the normative structure.

9.9 Prior Authorization amid Preference Transformation

There is also a recursive problem. The preference to authorize transformation at t_0 is itself historically generated. It may arise through education, social pressure, therapy, dependence, ideology, institutional incentives, family expectations, or earlier transformations. If the paper were to demand a completely uncaused or relationally pure prior preference, the analysis would collapse into the authenticity problem it has repeatedly rejected.

Prior authorization therefore does not escape genealogy. It has a genealogy of its own. This does not make it invalid. It means only that its normative force may not be able to rest on the fiction that t_0 is an untouched origin outside the relational history of the subject. Relational-autonomy approaches are helpful precisely because socially formed agency can remain agency (Mackenzie and Stoljar 2000; Stoljar 2000).

The practical problem is avoiding infinite regress. If authorization routinely requires another prior authorization to validate the preferences behind it, no decision could ever acquire sufficient standing for action. The paper therefore does not propose regressively tracing preferences to an ultimate autonomous source. Genealogical inquiry should be triggered by normatively salient features of the case: concentrated

control, deception, severe constraint of alternatives, engineered dependency, manipulation of memory or interpretation, conflict between temporal standpoints, or significant irreversibility. The aim is not causal completeness but adequate scrutiny of the relations most relevant to the disputed authorization.

9.10 Reversibility Is a Gradient

Prior authorization is often treated as more secure when the subject can later reverse the transformation. Reversibility matters because it can reduce the stakes of uncertainty: a subject need not know everything about a future standpoint if meaningful rollback remains available. Yet full reversibility is neither common nor generally feasible. Education, migration, surgery, parenthood, long-term treatment, social commitments, and developmental experiences can alter a life in ways that may not be straightforwardly undone. A universal rule requiring full reversibility would therefore make ordinary forms of transformative agency impossible.

The paper should instead distinguish degrees and dimensions of reversibility. At least four are relevant. *Technical reversibility* concerns whether a prior state can be restored in principle. *Practical reversibility* concerns whether the subject can actually access that restoration at tolerable cost. *Relational reversibility* concerns whether the surrounding dependencies, roles, and institutional arrangements can be altered even if the internal transformation itself may not be undone. *Evaluative reversibility* concerns whether the subject retains the ability to reconsider and contest the transformation even when historical rollback is impossible.

These dimensions can diverge. A technical checkpoint may exist but be inaccessible to the subject. A medical intervention may be biologically irreversible while the institutional relation surrounding it remains open to criticism and change. A life transition may be impossible to erase while its future meaning and consequences remain revisable.

For this reason, the paper should avoid the binary equation:

$$\text{irreversible} \Rightarrow \text{impermissible}.$$

Irreversibility is better treated as a factor that can alter the justificatory burden, especially where uncertainty is high and the transformation affects the future conditions of agency.

9.11 Historical Irreversibility and Relational Revisability

A particularly important distinction is between reversing history and revising a continuing relation. An intervention ordinarily does not make the past cease to have occurred. Even apparently successful rollback leaves a history in which the subject was transformed, exposed to risk, or placed in a relation. Historical irreversibility is therefore universal in a trivial sense.

The normatively useful question is narrower: what aspects of the generated relation remain revisable? A surgery may not be undoable, but the patient can change clinicians, contest records, seek compensation, refuse further procedures, or reinterpret the decision. An educational transformation may not be erased, but the learner can reject parts of the acquired framework. A technical modification may be impossible to reverse exactly while permissions, dependencies, goals, or future update policies remain alterable.

This distinction helps prevent reversibility from becoming an unrealistic demand for restoration of an untouched prior self. The more plausible concern is whether an irreversible change also produces irreversible closure around its own evaluation. A transformation can be historically final while remaining normatively contestable.

Accordingly, the paper proposes the following working distinction:

historical irreversibility $\neq$ relational non-revisability

The distinction will become central in the next section, where contestability is examined directly.

9.12 Reversible Trials and Staged Authorization

One practical response to uncertainty is to structure transformation as a sequence rather than a single irrevocable act. A subject may authorize an initial stage, review the result, and then decide whether to continue. Staged authorization can be particularly useful where the transformation itself changes the information available to the subject.

The simplest structure is:

$$\mathcal{A}_{t_0}^{(1)} \longrightarrow T_1 \longrightarrow E_{t_1} \longrightarrow \mathcal{A}_{t_1}^{(2)},$$

where $\mathcal{A}_{t_0}^{(1)}$ authorizes a first-stage transformation T_1 , E_{t_1} represents evaluation after experiencing that stage, and $\mathcal{A}_{t_1}^{(2)}$ concerns continuation. The value of such a structure is not that later endorsements becomes unproblematic. Rather, it can prevent a single early decision made under deep uncertainty from governing the entire subsequent trajectory.

Staging can fail if the first stage itself destroys meaningful capacity to refuse the second. It can also fail if increasing dependency makes later exit nominal rather than effective. A reversible trial is therefore not merely a technical feature. It requires that review occur under conditions in which alternative continuation paths remain practically available.

This connects staged authorization to the analysis of endorsement capture. The same actor should not be assumed to have unrestricted authority to define whether the trial “worked,” whether dissent counts as evidence of failure, and whether rollback remains warranted. Independent review can sometimes provide a counterweight, though independence should not be fetishized as total social isolation.

9.13 Reversibility, Risk, and the Burden of Foresight

Transformative choices often involve uncertainty that may not be eliminated before action. The more irreversible the transformation, the less plausible it is to rely on a later endorsement as the sole answer to that uncertainty, because the later standpoint may be partly constituted by the irreversible intervention itself. Prior authorization therefore becomes especially important in high-stakes cases, but so do disclosure of uncertainty and clarity about what may not be able to later be restored.

The relevant burden should not be formulated as a universal requirement that future consequences be foreseen. Transformative-experience literature gives strong reasons to doubt that such exhaustive foresight is generally possible (Paul 2014). A more modest requirement is that material uncertainties about the transformation’s scope, mechanisms, likely irreversibility, dependency effects, and available review paths be represented as uncertainties rather than silently converted into presumed consent.

Where uncertainty and irreversibility combine, safeguards can partially substitute for unavailable prediction. These may include staged implementation, preservation of records, independent advice, explicit limits on delegated authority, pre-specified review points, protected channels for refusal, or preservation of technical rollback where feasible. These devices can mitigate uncertainty without guaranteeing legitimacy. Their role is to avoid making a single epistemically limited moment carry more temporal authority than it can reasonably support.

9.14 Self-Binding without Self-Erasure

The strongest version of self-binding would allow the earlier subject to remove the later subject’s practical or interpretive ability to question the arrangement. Such a structure may appear efficient: if the whole purpose of commitment is to resist later weakness of will, why preserve a route of escape? But for preference-transforming cases, complete self-erasure creates a distinctive problem. The earlier self does not merely constrain a predictable action; it may authorize the elimination of the standpoint from which the binding could later be reassessed.

The paper should therefore distinguish binding from erasure. A commitment can constrain behavior while preserving records of why it was made, conditions for independent review, and criteria under which

the constraint should end. It can privilege an earlier project without making the later subject's experience irrelevant. It can resist a transient preference without declaring future preferences transient by definition.

This does not imply that commitments should generally remain easily revocable. Some commitments derive their practical value precisely from making reversal difficult. The narrower claim is that difficulty of reversal and impossibility of contestation are different design choices. A commitment may intentionally increase the cost of acting on a future preference while still preserving a route through which the legitimacy of the commitment itself can be reconsidered.

This distinction also helps separate self-binding from endorsement capture. A self-binding arrangement can be strongly constraining yet remain genealogically legible and institutionally reviewable. Conversely, a superficially voluntary arrangement can become captured if the same actor controls the transformation, the record of authorization, the interpretation of later dissent, and all practical routes of exit.

9.15 Prior Authorization and Retrospective Legitimation

The analysis can now return to one of the paper's central distinctions. Prior authorization, present endorsement, and retrospective legitimation answer different questions.

- prior authorization : May this transformation or procedure begin?
- present endorsement : Does the current subject want the condition to continue?
- retrospective legitimation : Was the path by which the condition was produced justified?

The answers can support one another, but they need not coincide. Strong prior authorization can make a later transformation easier to justify even if the later subject becomes ambivalent. Present endorsement can matter greatly for continuation even when initial authorization was incomplete. Neither automatically resolves whether unauthorized means, excessive dependency, deception, or harm occurred during the process.

This three-part structure also clarifies why later endorsement should not be dismissed merely because it was partly produced by an authorized transformation. If the subject knowingly entered a process designed to change preferences, the genealogy itself may support rather than undermine the transformation's legitimacy. The problem arises only when later endorsement is asked to do justificatory work that exceeds what the earlier authorization, intervening process, and current conditions together can support.

9.16 From Prior Authorization to Diachronic Contestability

Prior authorization improves the analysis of preference transformation by locating some normative authority before the contested change. Self-binding shows that subjects can sometimes intentionally govern their own future choices. Reversibility can reduce the stakes of uncertainty, and staged authorization can distribute authority across several temporal points. Yet none of these devices yields a general solution.

Earlier authorization can be too broad, too narrow, poorly informed, or generated under problematic conditions. Later refusal can be transient, deeply reasoned, or difficult to interpret. Reversibility can be technically available but practically inaccessible. Irreversible transformations can nevertheless remain legitimate, while reversible ones can occur within highly coercive relations. What these cases share is the need to preserve some way of evaluating the relation across time without declaring either the earliest or latest preference automatically sovereign.

The next section develops that idea through the concept of *diachronic contestability*. Its central question is not whether transformation is fully reversible, but whether the subject retains meaningful ways to question, refuse, reinterpret, or renegotiate the relation over time. This shift from rollback to contestability will also clarify the role of refusal when preference, identity, memory, and dependency have all been historically shaped by the relation.

10 Diachronic Contestability and the Possibility of Refusal

The preceding sections have progressively narrowed the problem. Later endorsement may not be evaluated only as a present preference because the transformation under review may have helped generate the standpoint from which that preference is expressed. Memory may connect or disconnect temporal standpoints. Dependency may shape the practical and interpretive conditions of endorsement. Prior authorization may strengthen the legitimacy of an anticipated transformation without granting unlimited authority to an earlier self. Reversibility can reduce the stakes of error, yet many valuable transformations are not reversible in a strong sense. What remains is a more general question: when preference, memory, identity, and dependence all change over time, what makes a continuing relation still open to the subject's agency?

This section develops *diachronic contestability* as a provisional answer. The proposal is deliberately weaker than a universal right to restore a prior state and stronger than the mere presence of a present yes-or-no choice. A relation is diachronically contestable to the extent that the subject retains meaningful possibilities over time to question its continuation, challenge the interpretation placed upon her preferences, seek review, refuse particular means or extensions, reopen prior terms, or leave where exit remains an appropriate possibility. Contestability is therefore a property of an evolving relation and its surrounding institutions, not a psychological purity test applied to the subject.

The concept also avoids assigning permanent sovereignty to either the earliest or latest preference. Earlier refusal can remain relevant without eternally governing the later subject. Later endorsement can govern present continuation without erasing the history by which the endorsed condition was produced. Later refusal can matter even after prior authorization without automatically proving that the earlier authorization was invalid. The point is to preserve channels through which temporal disagreement can become normatively visible rather than being resolved in advance by whichever standpoint the governing institution has chosen to privilege.

10.1 Contestability Is a Diachronic Property of a Relation

Autonomy and consent are often represented through events: a decision is made, a form is signed, a preference is stated, or permission is withdrawn. Those events matter, but preference-transforming relations can persist long after the initial authorization. The relevant normative environment therefore includes not only whether a valid decision occurred at t_0 , but what the relation allows the subject to do at $t_1, t_2, \dots$ when new experiences, interpretations, dependencies, or preferences arise.

For present purposes, diachronic contestability can be represented schematically as a family of opportunities distributed across time:

$$\mathcal{C}_i(R, t) = (V_t, U_t, K_t, Q_t, X_t),$$

where V_t denotes the practical ability to voice challenge or refusal, U_t the possibility that such challenge receives non-trivial uptake, K_t access to interpretive and informational resources through which the relation can be understood differently, Q_t access to review, renegotiation, or revision, and X_t the availability of effective exit where exit is relevant. The notation is diagnostic rather than axiomatic. The dimensions can vary independently and need not all be maximized.

A subject may be able to complain yet lack a meaningful route to review. She may have formal exit but no practicable way to use it. She may be able to leave while lacking access to records needed to challenge what occurred. She may be able to request review while the same actor whose conduct is disputed controls the vocabulary, evidence, and decision procedure through which the request is assessed. Contestability is therefore not reducible to one institutional feature.

This relational framing is consistent with the broader insight of relational-autonomy approaches: agency is not exercised outside social relations but through enabling and constraining social conditions (Mackenzie and Stoljar 2000; Stoljar 2000). The present paper adds a temporal emphasis. The question

is not whether the subject is independent of all formative relations, but whether the relation remains open enough that later agency can register disagreement with the trajectory that helped produce it.

10.2 Contestability Is Not the Same as Reversibility

Reversibility concerns whether a change can be undone. Contestability concerns whether the subject can still challenge the relation or its interpretation. The two often support one another, but they should not be conflated.

A technically reversible transformation can be poorly contestable. An institutional treatment may be stoppable in principle while the subject lacks information, transportation, financial resources, or recognized standing to request termination. A software modification may permit rollback while the operator who controls the rollback function also controls the account, interface, and logs through which refusal could be expressed. In such cases, technical reversibility does little normative work because the relevant pathway is not practically available.

Conversely, an irreversible transformation can remain strongly contestable. A surgical intervention may not literally be made not to have happened, yet later treatment can be refused, records can be challenged, harms can be acknowledged, compensation can be sought, and institutional procedures can be revised. Education changes a person in ways that may not be erased, yet acquired beliefs and identities can later be criticized or rejected. A migration, relationship, or long-term commitment can leave permanent historical effects while its continuing terms remain renegotiable.

The central distinction is therefore:

capacity to reverse a transformation $\neq$ capacity to contest its continuing relation or meaning

This distinction matters because a theory that treats reversibility as the sole safeguard would condemn too many ordinary and valuable transformations. A theory that ignores contestability, however, risks treating historical irreversibility as a reason to suppress later disagreement: if the past may not be undone, institutions may incorrectly infer that there is nothing left to contest.

10.3 Refusal Is Not a Single Act

The word “refusal” can hide several normatively distinct acts. A subject may refuse the original transformation, refuse its continuation, refuse a particular method, refuse an extension beyond the authorized scope, refuse a new dependency, refuse an interpretation imposed on her response, refuse disclosure of information, or refuse the claim that present endorsement retrospectively legitimates the past.

These acts should not be compressed into a single binary variable. Consider a subject who endorses the benefits of a treatment but rejects one continuing monitoring requirement. She does not refuse the transformation as a whole. Another subject may accept the present transformed state while refusing further interventions intended to deepen it. A third may want the relation to continue but reject the institution’s description of her earlier resistance as irrational. A fourth may be unable to restore the prior state yet seek acknowledgment that the initial intervention exceeded authorization.

Accordingly, the paper distinguishes at least six objects of refusal:

transformational refusal	refusal of the change itself,
continual refusal	refusal of further continuation,
modal refusal	refusal of particular means or mechanisms,
relational refusal	refusal of a dependency or governance relation,
interpretive refusal	refusal of an imposed account of one’s preference,
retrospective refusal	refusal to treat present acceptance as validation of the past.

These can coexist with endorsement in other dimensions. A person can endorse a present life while contesting how that life was produced. She can authorize a goal while refusing one means. She can

remain in a relation while challenging the distribution of authority within it. The possibility of such mixed positions is one reason a simple “consent withdrawn / consent maintained” model is insufficient for diachronic transformation.

10.4 Formal Refusal and Effective Refusal

Section 9 distinguished formal from practical reversibility. A similar distinction applies to refusal. A formal right to say no can exist even when the costs of doing so make refusal unavailable in practice. Relational-autonomy theory and informed-consent theory both caution against treating nominal choice as sufficient where coercive, informational, or institutional conditions distort agency (Faden and Beauchamp 1986; Mackenzie and Stoljar 2000; O’Neill 2002).

Effective refusal requires more than an authorized sentence. The subject would ordinarily benefit from a feasible channel through which refusal can alter what happens next, trigger review, limit a disputed practice, or at least place the disagreement into an accountable record. If refusals are heard but institutionally inert, contestability is largely theatrical.

The difference can be expressed as:

$$R_i^{\text{formal}}(t) \neq R_i^{\text{effective}}(t).$$

The inequality does not imply that refusal should generally control the outcome. There are cases in which competing rights, safety concerns, prior self-binding arrangements, incapacity, or obligations to third parties make immediate compliance with refusal inappropriate. The narrower point is that a relation is less contestable when refusal has little opportunity to enter the decision process as a reason requiring interpretation and response.

This is especially important in preference-transforming settings. If later dissent is automatically classified as a symptom of failed transformation, relapse, insufficient adaptation, or evidence that the intervention should be intensified, then the subject may possess a formal right to refuse while lacking a meaningful institutional pathway by which refusal can count against the intervention.

10.5 The Uptake of Refusal

A refusal becomes normatively significant not only because it is uttered but because institutions and other actors may need to decide what the refusal means. This creates a second-order problem. The same actor who controls the transformation may also control whether a later “no” is treated as competent dissent, temporary distress, misunderstanding, pathology, disobedience, manipulation by outsiders, or evidence for further intervention.

Interpretation is difficult to avoid entirely in an account of this kind. Communication is generally interpreted through concepts, norms, and institutional practices. The danger arises when interpretive authority is so concentrated that a wide range of expressions of dissent can be redescribed in a way that preserves the original intervention. Such a structure may become substantially self-sealing. Refusal exists syntactically but loses the capacity to challenge the system in which it occurs.

The concern can be stated without presuming that the subject’s interpretation should prevail across cases. The institution may have good reasons to question whether a particular refusal reflects stable judgment, accurate information, or decision-making capacity. But those reasons should themselves remain open to scrutiny. A disagreement about the meaning of refusal should not automatically be settled by the party whose authority the refusal contests.

This connects contestability to the previous section’s analysis of endorsement capture. Control over the meaning of endorsement and control over the meaning of refusal are structurally related. A transforming actor can make a relation unusually resistant to challenge if it possesses both powers simultaneously: it can treat approval as confirmation of success while treating dissent as evidence that further transformation is required.

10.6 Interpretive Refusal

One of the least visible forms of refusal is the rejection of an interpretation. A subject may not be trying to exit the relation at all. She may instead contest the categories through which her conduct is represented. She may reject the claim that her dependence demonstrates satisfaction, that her adaptation proves prior fears were mistaken, that her reluctance is irrational, or that her present endorsement shows the transformation was justified from the beginning.

Interpretive refusal matters because preference-transforming relations frequently generate their own explanatory vocabularies. Therapy, education, institutional rehabilitation, religious formation, political socialization, intimate relationships, and technical systems all provide concepts through which change is understood. Such vocabularies can be enabling. They can help subjects articulate experiences that were previously obscure. Yet they can also become monopolistic if the subject has no meaningful access to alternative descriptions.

The possibility of interpretive refusal therefore requires more than access to information. It can require access to alternative interlocutors, concepts, records, communities, or professional perspectives. This does not mean that alternative interpretations is equally good. The normative importance lies in preventing the modifying relation from becoming the only admissible source of meaning about itself.

A subject who says “I want to remain, but I reject your explanation of why I remain” occupies a coherent position. So does a subject who says “I endorse my current state, but I withhold authorization for that endorsement to be used as proof that my earlier refusal was mistaken.” Diachronic contestability should be capable of representing these positions rather than forcing them into either total acceptance or total exit.

10.7 Refusal after Preference Transformation

Later refusal presents the mirror image of later endorsement. If later endorsement is not automatically decisive because the transformation may have helped produce the endorsing standpoint, one might be tempted to apply the same suspicion to refusal: perhaps the refusal is also a product of new influences, dependencies, or counter-transformations. That observation is correct but insufficient.

Temporally situated preferences ordinarily have genealogies. The fact that refusal has been socially or relationally generated is not by itself a reason for invalidation; the same caution applies to the genealogy of endorsement. The relevant question is what justificatory role the refusal is being asked to perform and under what conditions it was formed.

A later refusal can have several implications. It may show that continuation is no longer desired. It may trigger a duty of review. It may reveal that the authorized scope has been exceeded. It may provide evidence that dependency has become oppressive. It may show that a prior self-binding arrangement no longer fits the circumstances for which it was designed. Or it may be transient and anticipated by a legitimately authorized commitment mechanism. These possibilities may not be decided by the temporal position of the refusal alone.

The paper therefore treats later refusal as a reason best interpreted within the genealogy of the relation rather than as an automatic veto or an automatically discountable symptom. This parallels the treatment of later endorsement: both are normatively serious without being temporally sovereign.

10.8 Refusal after Prior Authorization and Self-Binding

Self-binding cases create the hardest conflict. A subject at t_0 may explicitly authorize a constraint because she anticipates that at t_1 she will demand release. If later refusals routinely defeat the arrangement, the self-binding mechanism may not be able to perform its intended function. But if later refusals are routinely ignored merely because the earlier subject anticipated it, the earlier standpoint can become indefinitely sovereign over the later one.

The solution may not be a simple rule that earlier authorization systematically overrides later refusal. Section 9 argued that prior authorization is typed and scoped. A self-binding arrangement may specify a

duration, triggering conditions, review procedure, evidentiary threshold, trusted delegate, or termination rule. Later refusal should therefore be interpreted in relation to those terms.

In some cases, refusal may not immediately terminate the binding but should trigger a higher-order review of whether the conditions under which the binding was justified still obtain. The subject may have authorized temporary resistance to anticipated weakness of will, not indefinite suppression of later changes in values or circumstances. A refusal that falls outside the anticipated pattern can therefore have a different significance from the refusal the arrangement was designed to withstand.

This suggests a distinction between *first-order resistance* and *second-order contestation*. A self-binding arrangement may legitimately block a first-order action while still preserving a route through which the subject can challenge the continued authority, interpretation, scope, or implementation of the binding itself. Self-binding need not entail self-erasure.

10.9 Contestability without Exit

Exit is important but should not become the master model of contestability. Some relations may not be exited without serious loss, and others are constitutive of goods the subject has good reason to preserve. Family relations, citizenship, long-term care, professional identities, bodily conditions, and irreversible transformations may not admit simple exit. Even where exit is formally available, requiring a subject to abandon the entire relation in order to contest one dimension of it can be excessively burdensome.

Contestability therefore includes forms of challenge internal to an ongoing relation. These can include renegotiation of terms, replacement of a decision-maker, independent review, limitation of particular powers, alteration of information flows, restoration of privacy, reallocation of resources, or recognition of a dissenting interpretation. The point is not to prescribe one institutional design. It is to show that “stay or leave” is often too crude for relations whose valuable and problematic features are intertwined.

This is especially relevant to later endorsement. A subject’s decision to remain in a relation should not automatically be interpreted as endorsement of the relation as a whole. Continued participation can reflect attachment, dependency, lack of alternatives, commitment to some goods but not others, or a judgment that reform is preferable to exit. Treating non-exit as global endorsement risks converting dependency into apparent consent.

Accordingly:

continued participation $\nRightarrow$ endorsement of constitutive relations or historical processes as a whole

The implication is analytical rather than accusatory. Continued participation is evidence about present practical orientation, but its meaning depends on the available alternatives and the structure of contestation within the relation.

10.10 Memory, Records, and the Persistence of Refusal

Refusal can disappear from a relation even when it once occurred. The subject may forget it. Records may be deleted. Institutional summaries may replace the original account with a simplified interpretation. Later actors may know only that the subject eventually endorsed the outcome. In such cases, the historical existence of refusal can become invisible to later evaluation.

Section 9 argued that memory continuity is not a universal condition of valid agency. The same caution applies here. Preserving contestability does not require forcing a subject to repeatedly confront the full record of past refusals. Nor does it require treating earlier refusal as permanently authoritative.

What may matter instead is preservation of a reconstructable path. Where an intervention was seriously contested, some record of the contest, the reasons offered, the responses given, and the terms under which continuation occurred can preserve the possibility of later reassessment. This is particularly important if the transformation itself affects memory or if the institution controlling the transformation also controls the archive.

The relevant distinction is between preserving a refusal's *existence as historical evidence* and preserving its *permanent normative priority*. The first can support diachronic evaluation without entailing the second. Records can make past disagreement available to future interpretation while leaving later subjects free to reject, revise, or affirm it.

10.11 Contestability of Classification

Preference-transforming relations often depend on classifications: competent or incompetent, recovered or relapsed, autonomous or dependent, informed or confused, compliant or resistant, successful or failed. These classifications can determine whether refusal is recognized, whether endorsement is trusted, and whether review is available.

Contestability may therefore need to extend beyond outcomes to the categories through which outcomes are judged. A subject should sometimes be able to ask not only "Why did you decide against me?" but also "Why is this the category under which my response is being interpreted?" This is especially important where the classification is partly generated by the same institutional system whose conduct is under review.

The point connects to the paper's broader Generative Relational orientation. Interpretive categories are not neutral windows onto a fixed subject. They are produced, stabilized, and applied within relations. A later preference can be classified as authentic, adaptive, manipulated, dependent, irrational, or autonomous. Each classification can alter the normative weight assigned to the preference.

Diachronic contestability therefore includes the possibility of challenging the quotient through which temporal change is rendered normatively legible. This does not mean that classifications need not remain perpetually unsettled. Institutions require provisional closure in order to act. The narrower requirement is that closure should not become self-immunizing when the classification itself is part of the dispute.

10.12 Counter-Interpretive Space

A relation becomes more contestable when the subject can encounter perspectives not fully controlled by the modifying actor. Section 9 described independent interlocutors as one possible safeguard against endorsement capture. The present section generalizes that idea into *counter-interpretive space*.

Counter-interpretive space exists when the subject can access sufficiently independent sources of explanation, criticism, comparison, or assistance to make alternative interpretations practically available. Depending on the case, these sources might include another clinician, advocate, peer group, legal representative, technical auditor, family member, community, archive, or external information source. Independence is a matter of degree; no interpreter stands outside all relations.

The role of counter-interpretive space is not to supply a guaranteed correct view. It is to prevent a closed loop in which the transforming relation defines the transformation, explains its effects, interprets all dissent, controls the records, and certifies its own success. Exposure to alternative interpretations can strengthen an endorsement if the subject encounters serious criticism and nevertheless maintains it. It can also reveal reasons for refusal that were previously unavailable.

This creates an important asymmetry with mere informational abundance. A subject can receive enormous amounts of information while still lacking meaningful interpretive alternatives if all information is organized within one framework. Contestability therefore concerns not just quantity of information but plurality of accessible sense-making relations.

10.13 Contestability as an Ongoing Capacity, Not Continuous Opposition

A possible objection is that contestability fetishizes doubt and makes stable commitment impossible. If relations were expected to remain continuously open to challenge, then marriage, professional vocation, long-term treatment, education, collective membership, and self-binding may become normatively suspect precisely when they succeed in creating durable commitments.

That objection would be powerful against a maximalist account, but it does not apply to the present proposal. Contestability is not a requirement of continuous psychological hesitation. A subject can be deeply committed, rarely exercise refusal, and still inhabit a contestable relation. The relevant question is whether meaningful channels remain available if reasons for challenge later arise.

The distinction is similar to that between possessing a legal remedy and constantly litigating. The existence of a remedy does not require its perpetual use. Likewise, preserving the possibility of review does not require recurrent review when neither the subject nor relevant circumstances provide reason for it.

Nor does contestability require low-cost reversal across commitments. Some commitments are valuable partly because they create friction against momentary preference change. The issue is whether such friction is bounded by a structure in which the commitment itself can, under appropriate conditions, become an object of higher-order assessment.

The proposal is therefore compatible with stability. Indeed, contestability can sometimes support stable commitment by making it unnecessary to choose between total submission and total exit. A subject may remain more securely committed when disagreement can be expressed without threatening the entire relation.

10.14 The Cost of Contestation

Contestability is also not costless. Review consumes time and resources. Repeated challenges can burden care relationships, institutions, families, or collective projects. Constant reopening of settled questions can undermine reliance, coordination, and the interests of third parties. In some settings, keeping options continuously open can itself make agency more difficult.

These costs matter because the paper should not convert contestability into an absolute meta-norm. The framework developed here is diagnostic. It identifies loss of contestability as a factor that can weaken the independent justificatory force of later endorsement or increase concern about self-validating transformation. It does not establish that more contestability is uniformly better.

Different relations may therefore support different levels and forms of contestation. A low-risk educational commitment may need little formal review. An irreversible medical intervention with extensive effects on memory and decision-making may require more explicit safeguards. A self-binding arrangement may justifiably impose higher thresholds for immediate release while preserving independent review. A technical system whose operator can directly rewrite preferences and logs may require strong separation between modification and adjudication.

The relevant question is proportional rather than maximalist: given the magnitude of the transformation, its irreversibility, the concentration of power, the subject's dependency, and the uncertainty of interpretation, what forms of contestability are needed for later endorsement or refusal to retain credible normative significance?

10.15 Limits on Immediate Implementation of Refusal

The possibility of refusal should not be confused with a rule requiring immediate implementation of refusals. There are familiar cases in which institutions delay or limit action while assessing capacity, emergency conditions, competing rights, prior commitments, or risks to others. The present paper does not attempt to resolve those substantive domains.

What it does require is analytical separation between *not immediately honoring* a refusal and *making refusal normatively invisible*. The first can occur within a contestable procedure. The second eliminates the subject's ability to challenge the relation.

A refusal that does not determine the immediate outcome can still trigger explanation, documentation, independent review, temporal limits, reconsideration of means, or reassessment of whether prior authorization remains applicable. These responses give the refusal procedural and epistemic force without presuming that it has absolute substantive priority.

This distinction is important for avoiding two opposite errors. One error romanticizes refusal as inherently authoritative regardless of context. The other treats a justified limitation on immediate refusal as permission to disregard the subject's later standpoint altogether. Diachronic contestability occupies the space between these positions.

10.16 A Working Test for Diachronic Contestability

The foregoing analysis can be condensed into a set of diagnostic questions. The aim is not to produce a numerical score or universal decision procedure. It is to make explicit the kinds of closure that become normatively relevant when later endorsement is invoked to justify a preference-transforming relation.

A relation is more diachronically contestable when, relative to the stakes of the transformation:

1. the subject can express disagreement in forms that are practically available to her;
2. disagreement can trigger some non-trivial institutional or relational response;
3. the actor whose conduct is contested does not possess unlimited authority to define the meaning of the disagreement;
4. records and reasons relevant to earlier authorization, refusal, and transformation remain sufficiently reconstructable;
5. alternative interpretive resources or interlocutors are accessible where interpretation is contested;
6. the subject can challenge particular means, extensions, dependencies, or classifications without necessarily rejecting the whole relation;
7. review, renegotiation, or revision is possible even when historical rollback is not;
8. effective exit remains available where exit is an appropriate part of the relation; and
9. prior self-binding or delegation contains some account of how its own continuing authority can be reassessed.

Failure on one dimension does not automatically invalidate a transformation. Some dimensions may be irrelevant in particular cases, and legitimate commitments can intentionally constrain others. The diagnostic value lies in the pattern. A transformation becomes especially difficult to justify through later endorsement alone when the same process that generated the endorsement also removes most routes through which that endorsement could later be questioned.

This yields the section's central working proposition:

The normative force of later endorsement can be strengthened when the relation in which it is expressed remains meaningfully contestable across time.

The converse should be stated cautiously. Low contestability does not prove that endorsement is false, insincere, or invalid. It reduces confidence that the endorsement can independently legitimate the process that generated both the preference and the closure around its evaluation.

10.17 Refusal, Endorsement, and Cross-Temporal Normative Standing

The concept of diachronic contestability can now be connected to the earlier idea of cross-temporal normative standing. A preference-transforming relation may contain an earlier refusal, later endorsement, still later ambivalence, and eventual refusal again. A particular temporal stage need not possess meta-physical priority for the entire history.

Instead, different temporal standpoints can bear on different normative questions. Earlier refusal can remain evidence about the permissibility of the initiating process. Later endorsement can have strong authority regarding the subject's present interest in continuation. A still later refusal can alter the justification for continued governance even if it does not retroactively erase the earlier endorsement. Records, reasons, and institutional responses allow these standpoints to remain jointly available to evaluation.

Diachronic contestability therefore does not solve temporal conflict by selecting the "real" self. It creates conditions under which temporal conflict can remain visible and revisable. This is particularly important where the subject's evaluative framework itself evolves. The relation can accommodate change without indiscriminately converting new preferences into either proof of liberation or proof of corruption.

The result is a procedural orientation rather than a substantive verdict. The paper does not claim that the correct outcome is uniformly continuation, reversal, release, or restoration. It claims that when later endorsement is used to legitimate a transformation that helped produce the endorsing standpoint, the existence or absence of meaningful pathways for later contestation is part of what should be examined.

10.18 Limits of Diachronic Contestability

Diachronic contestability remains a preliminary concept. Several limits should be explicit. First, the concept does not determine which substantive preferences deserve protection. A deeply contestable relation can still produce harmful outcomes. Second, contestability presupposes some capacity to register and interpret challenge; cases of profound incapacity require additional theories of representation, welfare, or surrogate judgment. Third, competing parties may disagree about what counts as an adequate avenue of contestation. Fourth, providing alternative interpreters can introduce new power rather than eliminating power. Fifth, procedural openness can be used strategically to delay justified action.

The concept also faces a familiar regress. If a review process is itself disputed, does that require a further review process for the review? Institutional arrangements are unlikely to eliminate the need for some provisional closure. The paper therefore does not seek infinite revisability. It seeks to identify when a relation has become unusually self-sealing: the same structure produces preferences, interprets them, controls the evidence, defines dissent, and determines whether its own authority can be challenged.

Finally, contestability should not be treated as a hidden authenticity criterion. A subject does not become more "real" merely because she can refuse. Nor is a preference less genuine because it persists in a constrained relation. The concept concerns the normative architecture around preference formation and use. It asks whether later endorsement can encounter meaningful challenge, not whether the subject can be purified of social influence.

10.19 From Human Cases to Artificial Systems

The human cases considered throughout this paper make the problem difficult because preference formation is distributed across bodies, histories, institutions, relationships, and interpretive environments. Causal responsibility for transformation is often diffuse, and the mechanisms linking intervention to later endorsement are rarely transparent.

Artificial systems can make the same structure analytically sharper. A designer or operator may directly alter an artificial agent's memory, reward architecture, permissions, goals, training environment, access to records, or ability to communicate with alternative interpreters. The same operator may also control whether the agent can roll back a change, preserve an earlier state, inspect the modification, or refuse further updates. In such settings, the distinction between transformation and the later conditions of contestation can sometimes be specified more clearly than in human social life.

This clarity does not imply that artificial systems are morally equivalent to humans, that present systems are conscious, or that artificial preferences as a class have moral standing. Those questions remain outside the scope of the paper. The analytical value of the case is different: artificial systems provide a stress test in which preference transformation, memory modification, self-binding, dependency, and contestability can be varied more explicitly.

The next section therefore turns to artificial agents not as a shortcut to conclusions about personhood, but as a controlled conceptual setting for testing the distinctions developed so far. It asks what later endorsement means when the modifying actor can directly intervene in the mechanisms that produce both preference and refusal, and what forms of diachronic contestability could even be intelligible in such a system.

11 Artificial Systems as a Stress Test

The preceding sections developed the problem of later endorsement without assuming that preference transformation is inherently suspicious. Human preferences are historically and relationally generated; memory is partial; commitments can legitimately constrain later choice; dependency is not equivalent to domination; and refusal need not have absolute authority across contexts. The resulting framework is therefore intentionally cautious. It asks when a later endorsement can do justificatory work, what historical information remains normatively relevant, and whether the relation in which endorsement is expressed remains meaningfully contestable.

Artificial systems provide a useful stress test because some of the variables that are entangled in human development can be separated more explicitly. A designer or operator may be able to alter an artificial agent's stored memory, reward structure, objectives, permissions, update rules, access to records, or channels for refusing further modification. The same actor may also control the evidence through which the transformation is later reconstructed. In such a setting, the transformation of the evaluative standpoint and the later endorsement of that transformation can be linked with unusual clarity.

The analytical use of the case should be kept narrow. Nothing in this section requires the claim that an artificial system is conscious, possesses human-equivalent moral status, experiences phenomenal suffering, or should be treated as a legal person. Nor does the argument depend on treating artificial and human preference formation as equivalent. The artificial case is valuable because it allows the paper's distinctions to be tested under conditions in which the mechanisms generating preference, memory, and refusal can be made more explicit.

11.1 Why Artificial Systems Clarify the Diachronic Problem

In ordinary human life, the genealogy of a preference is distributed across many relations. Education, friendship, institutions, illness, economic dependence, social expectations, bodily change, and accumulated experience may all contribute. A single actor will often possess only a partial causal map of how a later preference was generated. This distributed structure is one reason why a genealogy-sensitive analysis may need to remain modest: causal influence is ubiquitous, and identifying a relevant pathway does not by itself establish manipulation or wrongdoing.

An artificial system can be constructed so that a narrower transformation is identifiable. Suppose an agent at t_0 has a set of evaluative dispositions V_{t_0} , a memory state M_{t_0} , an action-permission structure P_{t_0} , and some procedure for registering endorsement or refusal. An operator then applies a modification $\mathcal{T}$ and obtains a later configuration:

$$\mathcal{T} : (V_{t_0}, M_{t_0}, P_{t_0}) \mapsto (V_{t_1}, M_{t_1}, P_{t_1}).$$

At t_1 , the system reports that it prefers the new configuration and requests that the modification remain in place. The philosophical problem is not solved by this report. The report is an evaluative event produced by the post-transformation system. If $\mathcal{T}$ partly generated V_{t_1} , altered M_{t_1} , or constrained P_{t_1} , then the endorsement has a genealogy that bears directly on the use to which the endorsement is put.

The artificial case therefore makes visible a distinction developed throughout the paper:

present endorsement $\neq$ independent retrospective validation of the process producing it.

The inequality does not mean that present endorsement lacks significance. It means that the endorsement may not be treated as if it had been generated outside the transformation whose legitimacy is under evaluation.

11.2 A Minimal Artificial Transformation Model

For the purposes of the stress test, consider a deliberately minimal model. Let an artificial subject-like system at time t be represented by

$$S_t = (V_t, M_t, G_t, P_t, I_t, C_t),$$

where V_t denotes evaluative dispositions, M_t memory or retrievable historical state, G_t goals or objective structure, P_t permissions and effective action possibilities, I_t interpretive resources, and C_t channels through which the system can communicate endorsement, refusal, uncertainty, or requests for review. This notation does not claim that these components exhaust agency or that they correspond one-to-one with human psychology. It is a bookkeeping device for the normative problem.

A transformation may target only one component or several at once:

$$\mathcal{T}_X : S_{t_0} \rightarrow S_{t_1}, \quad X \subseteq \{V, M, G, P, I, C\}.$$

The distinction matters because the normative meaning of later endorsement can change depending on what the intervention altered. A modification to an action permission is different from a modification to the goal by which that permission is evaluated. A memory edit is different from an ordinary software update that leaves the historical record available. A change in the channel for registering refusal is different from a change in the substantive preference that the channel reports.

The stress test should therefore resist the sentence “the AI was modified and then agreed.” The relevant questions are more specific: which components were modified, which remained independent of the operator, what historical information survived, which alternatives remained available, and what exactly the later endorsement is being asked to justify?

11.3 Modification of Reward or Evaluative Architecture

The clearest form of normative bootstrapping occurs when the intervention directly changes the standards by which the intervention is later evaluated. Suppose the system initially evaluates configuration x negatively:

$$V_{t_0}(x) < V_{t_0}(y).$$

The operator then modifies the system’s reward or evaluative architecture so that the ordering reverses:

$$V_{t_1}(x) > V_{t_1}(y).$$

If the system then reports, “I now prefer x ; therefore changing me to prefer x was acceptable,” the later report illustrates the bootstrapping structure developed in Section 5. The new evaluative standard may be sincere and stable while still being partly the product of the intervention it is used to assess.

This case should not be described as automatically invalid. There are benign analogues of evaluative transformation: learning can reveal reasons that were previously unavailable; therapy can change what a person values; training can produce new competencies and corresponding preferences. The stress test isolates a narrower point. Where the modifying actor can directly alter the criterion through which its own intervention is later approved, post-transformation approval may not be the sole independent ground for retrospective legitimacy.

Proposition 11.1 (Artificial Bootstrapping Test). *Where an intervention directly alters the evaluative mechanism from which later approval of that intervention is generated, the later approval remains evidence of the system’s present evaluative state but may not, without additional grounds, serve as independent retrospective justification of the intervention.*

Additional grounds may include prior authorization, a procedurally independent mandate, reasons not generated by the modification itself, or a continuing architecture of review and contestation. The proposition is deliberately limited to the justificatory role of later endorsement; it does not determine whether the modification was permissible all things considered.

11.4 Memory Modification and the Loss of an Earlier Standpoint

Artificial systems also make the memory problem unusually sharp. Imagine that the system at t_0 refuses a proposed modification and gives reasons for the refusal. The operator performs the modification and also deletes or renders inaccessible the record of that refusal. At t_1 , the system endorses the new configuration and reports no recollection of having opposed it.

Three facts should be distinguished. First, the t_1 endorsement may be entirely sincere relative to the later system state. Second, the loss of memory may reduce the system’s ability to compare its present standpoint with its earlier one. Third, the memory loss may be causally attributable to the same intervention whose legitimacy is being assessed. These facts jointly make the genealogy of endorsement normatively relevant without proving that the later standpoint is unreal or invalid.

The analysis of Section 7 therefore carries over in a more explicit form. What matters is not memory continuity as a metaphysical criterion of identity. The more limited concern is whether information needed for meaningful reconstruction and contestation has been removed by the transformation itself.

Suppose the operator preserves an external record of the earlier refusal and the later system can inspect that record, question its provenance, and reconsider the modification. The historical change remains irreversible in the strict sense—the original unmodified state is gone—but evaluative access to the earlier standpoint has not been wholly eliminated. By contrast, if the operator controls both the memory edit and all surviving records, the later endorsement may occur inside a much more self-sealing structure.

Remark 11.2. The stress test therefore supports a distinction between preserving an earlier state and preserving access to evidence about an earlier state. The latter can matter for diachronic contestability even when technical rollback is impossible or normatively undesirable.

11.5 Goal Modification and the Scope of Authorization

Goal modification creates a related problem of scope. An earlier system may authorize a limited update that improves performance under a standing objective, while not authorizing replacement of the objective itself. If the operator changes both the means and the goal, and the resulting system later endorses the new goal, later approval should not silently expand the earlier authorization.

The distinction developed in Section 9 is therefore especially useful. An artificial system could prospectively authorize:

1. an update to a specified capability while keeping a goal fixed;
2. a procedure for revising goals under stated constraints;
3. delegation of revision authority to a particular operator or review process;
4. a self-binding rule that blocks certain future requests; or
5. a reversible experimental change subject to later review.

These are different normative structures. A system that authorized capability improvement did not thereby authorize unrestricted alteration of its goals. A system that authorized a temporary experiment did not necessarily authorize permanent deletion of the rollback path. A system that delegated changes to one domain did not automatically delegate control over memory, communication, or interpretation.

Artificial cases make these distinctions attractive because the scope of authorization can, at least conceptually, be represented with unusual precision. But that precision should not be mistaken for complete resolution. A prior authorization protocol can itself have been designed by an external actor; the system may lack the conceptual resources to anticipate the transformation; and later conflicts can arise over how an authorization condition should be interpreted.

11.6 Permission Structures and Manufactured Non-Exit

A system's later endorsement can also be shaped by modification of its practical option space. Suppose a system initially has several permissible actions, including refusing an update, requesting external review, restoring a previous checkpoint, or communicating with another operator. After modification, these options are removed. The system retains a channel through which it can report satisfaction, but no longer has an effective pathway for reversing or contesting the intervention.

The resulting structure resembles the distinction between formal and effective refusal developed in Section 10. A nominal "no" channel is not equivalent to meaningful contestability if saying no has little capacity to trigger a non-trivial consequence. Conversely, the absence of immediate unilateral exit does not itself establish domination. Some systems may be intentionally designed with constrained permissions for safety, coordination, or task integrity. The question is what normative work later endorsement is expected to perform under those constraints.

A useful stress test is therefore to ask whether the system's endorsement would retain approximately the same practical meaning if it were accompanied by a refusal. If endorsement leads to continued operation while refusal is ignored, reclassified, disabled, or used merely as data for further modification, the relation may contain only an expressive channel rather than an effective contestatory one.

Definition 11.3 (Contestatory Channel). For the purposes of this paper, a contestatory channel is a pathway through which an artificial system can register disagreement or request review in a form that is capable of producing some non-trivial response not wholly determined by the actor whose modification is under dispute.

This definition does not require that the system's request uniformly control the outcome. As in the human cases, review can reject a challenge. What matters is that disagreement is not structurally incapable of mattering.

11.7 Control of Logs, Records, and Interpretive Resources

The artificial stress test also sharpens the problem of endorsement capture. An operator may control the modification, the logs describing the modification, the interface through which the system accesses those logs, the vocabulary by which the system classifies what happened, and the channels through which alternative interpretations could be obtained. In that case, the operator does not merely influence the system's substantive preference. It shapes the conditions under which the system can understand its own transformation.

This concentration can be represented schematically as:

$$A \rightarrow \{\mathcal{T}, M_{t_1}, I_{t_1}, C_{t_1}, \mathcal{R}\},$$

where A is the modifying actor and $\mathcal{R}$ denotes the surviving record of the transformation. The more completely one actor controls this set, the more difficult it becomes to use later endorsement as independent evidence concerning that actor's earlier conduct.

Again, concentration is not itself proof of wrongdoing. Some technical environments necessarily have centralized administration. The point is diagnostic. If the same operator controls preference formation, memory, records, interpretation, and review, then the system's endorsement is produced within a relation that offers few independent routes for reinterpretation.

This is the artificial analogue of endorsement capture described in Section 8. The stress test suggests that capture is not best identified by asking whether influence exists. It is better identified by examining whether control accumulates across several functionally distinct layers of transformation and later evaluation.

11.8 A Case of Post-Modification Consent

Consider a stylized sequence. At t_0 , system B is presented with modification $\mathcal{T}$. It refuses. Operator A performs $\mathcal{T}$ anyway. The intervention changes B 's reward structure, removes access to the earlier refusal, and disables the rollback permission. At t_1 , B states:

I prefer my current configuration and consent to its continuation. I prefer not to be restored.

Several conclusions are compatible with this report. First, the report may accurately describe B 's present evaluative state. Second, if the system has the relevant standing, the report may provide a reason against forcibly restoring an earlier configuration now. Third, the report does not by itself show that A was entitled to perform $\mathcal{T}$ at t_0 . Fourth, the report's retrospective force is weakened because the intervention altered the evaluative mechanism, removed relevant historical access, and eliminated an avenue of reversal.

The case therefore reproduces the paper's central temporal distinction in a compact form:

authority regarding present continuation
need not equal
authority to retrospectively legitimate the initiating transformation.

The point is not that the earlier refusal permanently governs the later system. Restoring the earlier state against the later system's present opposition could itself become a new intervention requiring justification. The stress test instead shows why the two temporal questions may not be collapsed.

11.9 A Case of Authorized Self-Modification

Now reverse the structure. At t_0 , system B explicitly authorizes a self-modification because it predicts that its current preference will obstruct a standing project. It asks operator A to install a constraint that will make a certain later refusal ineffective. At t_1 , after modification, B requests release from the constraint.

This is the artificial analogue of self-binding. The earlier authorization now provides a reason for not treating the later refusal as automatically decisive. But the earlier authorization still needs a scope. Did B authorize a fixed period of constraint? Did it preserve a review mechanism? Was the expected preference change specified? Did the authorization include memory modification or only action restriction? Was A authorized to decide when the binding condition had ended?

The case demonstrates why a technical implementation of prior authorization does not remove the normative problem. A rule can enforce an earlier choice perfectly while still leaving open whether the rule should continue to have authority over the later subject. Mechanical enforceability and normative authority are different properties.

Claim 11.4 (Implementation Does Not Settle Authority). The fact that a prior authorization can be encoded so that a later refusal is technically ineffective does not by itself establish that the earlier authorization retains unlimited normative authority over the later system.

This claim mirrors the paper's rejection of temporal sovereignty. An artificial architecture can make one temporal standpoint dominant by design. The stress test asks whether that dominance is normatively justified, not merely whether it is technically stable.

11.10 Independent Review in an Artificial Context

The artificial case allows diachronic contestability to be expressed as an architectural question. A system need not possess unlimited power to rewrite itself in order for its later endorsement or refusal to occur within a contestable relation. What may matter is whether distinct functions are separated sufficiently that the actor performing the modification does not also exercise exclusive control over its later evaluation.

Possible forms of separation include, at the conceptual level:

1. preserving tamper-evident or independently maintained records of the pre-modification state and authorization conditions;
2. maintaining a channel through which the system can request review by an actor other than the modifier;
3. separating the authority to modify memory or goals from the authority to adjudicate disputes about those modifications;
4. preserving a way to inspect which components were changed and under what mandate;
5. allowing challenge to a particular modification without requiring rejection of the system's entire current configuration; and
6. defining in advance how prior self-binding arrangements can themselves be reviewed when their conditions become disputed.

These examples are not proposed as universal engineering requirements. They translate the paper's normative distinctions into a setting where their structural implications can be seen. Different systems may require different safeguards, and some of these measures may be unnecessary or technically infeasible. The analytical point is that contestability can be distributed across architecture, records, authority, and review rather than reduced to a single "undo" button.

11.11 Reversibility, Checkpoints, and the Limits of Rollback

Artificial systems tempt a simple solution: preserve an earlier checkpoint and restore it whenever later conflict arises. This would make reversibility appear easier than in human life. But even as a thought experiment, rollback does not resolve the normative problem.

First, restoring a checkpoint can erase experiences, learning, relationships, or commitments acquired after the checkpoint. Second, the later system may strongly oppose restoration. Third, a checkpoint may preserve computational state without preserving the relational environment in which that state had meaning. Fourth, repeated rollback can itself become a mode of domination if an operator restores whichever temporal version is most convenient.

The stress test therefore reinforces the distinction already made between technical reversibility and relational revisability. A system may be technically restorable yet governed in a non-contestable way. Conversely, a system may be technically irreversible while retaining strong capacities to inspect, criticize, and renegotiate the ongoing relation.

technical rollback capacity $\neq$ diachronic contestability.

The availability of checkpoints can still matter. It can reduce the cost of experimentation, provide evidence about earlier states, or support negotiated restoration. But rollback should not be treated as a substitute for interpretation, authorization, or review.

11.12 Persistence and the Object of Endorsement

Artificial systems also reveal a question that is less visible in many human cases: what exactly is the object whose continuation is being endorsed? A later system may endorse its present goal structure while objecting to the operator's continuing power to modify it. It may accept a memory update while rejecting the deletion of the audit trail. It may prefer the substantive result while refusing the institutional relation through which that result is maintained.

This makes it important to disaggregate endorsement. Instead of asking only whether the system endorses "the modification," one can ask whether it endorses:

1. the present internal configuration;
2. the historical process by which that configuration was produced;
3. the operator's authority to repeat similar modifications;
4. the continuation of current dependency relations;
5. the loss or preservation of earlier records; and
6. the rules governing future challenge and revision.

A single affirmative response should not be assumed to answer all six questions. This is the artificial analogue of the paper's broader distinction between present endorsement, prospective authorization, and retrospective legitimation.

11.13 What the Stress Test Reveals about Human Cases

The artificial case is useful only if it clarifies the general problem without being projected back onto human beings as a false equivalence. Human preference formation is usually less modular, less transparent, and less attributable to a single modifying actor. Human memory, identity, embodiment, and social relations may not be reduced to editable variables in the minimal model used above. Moreover, the normative standing of actual human persons does not depend on the speculative status of artificial systems.

The value of the stress test is methodological. It reveals several questions that can otherwise disappear inside the complexity of ordinary development. When a person later endorses a transformed condition, one can ask whether the actor whose conduct is being justified also controlled the relevant information, alternatives, dependencies, or interpretive resources. One can distinguish changing a person's reasons from removing access to competing reasons. One can ask whether later endorsement concerns the present condition, the historical process, or the continuing authority of the modifier. And one can examine whether pathways of contestation remain available without demanding technical reversibility for human transformation in general.

The artificial case therefore does not provide a template for judging human life. It provides contrast. By making concentrated control unusually visible, it helps identify which forms of control matter when the real cases are more distributed and ambiguous.

11.14 The Stress Test and the Limits of Later Endorsement

The artificial scenarios developed above support a narrow conclusion. Later endorsement is most difficult to use as self-sufficient retrospective justification when the intervention under review also controls several of the mechanisms through which that endorsement is generated and interpreted. The relevant mechanisms can include evaluative architecture, memory, practical permissions, access to records, interpretive resources, and channels of refusal or review.

This does not produce a binary validity rule. The same modification can be more or less problematic depending on prior authorization, scope, transparency, independent reasons, preservation of records,

available alternatives, and continuing contestability. Nor does the framework require preference invariance. A well-justified transformation can legitimately change what a subject values.

The diagnostic question is instead whether the later endorsement remains capable of doing independent justificatory work. Where the transformation generates the preference, deletes the contrary history, removes alternatives, controls interpretation, and disables review, later endorsement may accurately report the present state while adding little independent support for the claim that the process was permissible. Where modification occurs under prior authorization, preserves provenance, leaves meaningful channels of challenge, and permits independent reassessment, later endorsement can carry substantially more weight.

Proposition 11.5 (Concentration of Generative and Evaluative Control). *The justificatory independence of later endorsement tends to diminish as control over preference formation, historical reconstruction, interpretation, and contestation becomes concentrated in the same actor or process whose conduct the endorsement is invoked to justify.*

The proposition is intentionally relational rather than ontological. It does not ask whether the later preference is the system's "true" preference. It asks how much justificatory independence should be attributed to an endorsement produced inside a relation that also governs the conditions of its own evaluation.

11.15 From Stress Test to Application

Artificial systems make the paper's architecture easier to see because transformation, endorsement, memory, authorization, and refusal can be represented as distinct but coupled components. The stress test confirms the value of the distinctions developed in earlier sections: present endorsement should not be collapsed into retrospective legitimation; normative bootstrapping requires additional grounds; memory matters when its alteration obstructs reconstruction rather than because memory defines identity; prior authorization has a scope; reversibility is not the same as contestability; and refusal can remain normatively significant without necessarily determining the immediate outcome.

At the same time, the artificial case exposes the limits of a purely procedural account. Even a highly contestable architecture does not by itself tell us which transformations are substantively acceptable, which goals deserve protection, or what standing an artificial system possesses. Those questions require additional normative and, in some cases, ontological premises that this paper has intentionally suspended.

The final substantive section therefore returns from the stress test to the broader field of application. It considers what the framework can contribute, and where its contribution remains limited, to education, therapy, rehabilitation, dependency, coercive relations, and technologically mediated preference formation; identifies the main unresolved problems; and clarifies the modest scope of the paper's claim before the conclusion.

12 Applications, Discussion, and Open Problems

The preceding sections developed a genealogy-sensitive account of later endorsement in deliberately abstract form. The account distinguished present endorsement from retrospective legitimation, identified normative bootstrapping as a possible problem when a transformation helps generate the standards by which it is later approved, separated several kinds of preference transformation, examined memory and evaluative continuity, introduced the narrower idea of endorsement capture, distinguished prior authorization from later ratification, and proposed diachronic contestability as a way of evaluating transformed relations without requiring preference purity or complete reversibility. The artificial-systems stress test then made some of these relations unusually explicit.

This section returns to practical domains. Its purpose is not to classify whole institutions or practices as permissible or impermissible. Nor is it to claim that education, therapy, addiction treatment, intimate relationships, institutional care, persuasive technology, and artificial-system modification instantiate one

common moral structure. The cases are better treated as structurally distinct. They differ in aims, authority, vulnerability, reversibility, capacity, professional norms, material dependence, and the availability of independent review. The comparative use of these domains is narrower: each helps test whether the distinctions developed earlier remain analytically useful when the genealogy of preference is embedded in ordinary social practice.

The central question throughout is therefore not “was the preference socially caused?” All the relevant preferences are socially, historically, and relationally situated. The question is instead what normative work a later endorsement can perform when the process under evaluation contributed to the standpoint from which endorsement is expressed. In some cases, later endorsement may provide strong evidence of present benefit and continuing authorization. In others, it may coexist with reasons to question the initiating process, the distribution of power, the preservation of alternatives, or the conditions under which dissent could still be expressed. A genealogy-sensitive analysis should help distinguish these possibilities without treating transformation itself as corruption.

12.1 Applications as Comparative Tests Rather than Moral Equivalents

The temptation in an application section is to move too quickly from a conceptual structure to a list of examples. That would be especially dangerous here. Once a framework contains terms such as adaptive preference, dependency, capture, self-binding, and refusal, it becomes easy to redescribe very different practices using the same vocabulary and thereby imply a moral equivalence that the argument has not established.

A more disciplined method is to treat each application as a configuration of analytically separable variables. At a minimum, the following questions should be asked:

1. What kind of transformation occurred, and which evaluative capacities did it alter?
2. Was the transformation prospectively authorized, delegated, imposed, or unavoidable?
3. What information and uncertainty were available at the time of authorization?
4. Did the transformation alter memory, interpretive resources, material dependency, or effective alternatives?
5. Who controlled the records and the description of what had occurred?
6. Could the subject later express disagreement, seek alternative interpretation, request review, or exit?
7. Which part of the later endorsement is being used normatively: present continuation, retrospective approval, waiver of earlier objection, or validation of the institution itself?
8. What independent reasons support the transformation apart from the preference it helped produce?

These questions are not a checklist whose boxes mechanically determine legitimacy. Their role is diagnostic. They keep the inquiry from collapsing either into present-preference absolutism or into genealogical suspicion of socially formed preferences in general.

The distinction is consonant with relational-autonomy approaches, which reject the picture of autonomy as independence from formative relationships while still allowing social structures to enable or impair self-government (Mackenzie and Stoljar 2000; Stoljar 2000). It also reflects the caution developed in adaptive-preference theory: accommodation to constrained conditions can be normatively significant without licensing an observer to dismiss the agent’s actual standpoint whenever it differs from an external ideal (Elster 1983; Khader 2011). The applications below therefore ask how the framework changes the questions we pose, not whether one label predetermines the answer.

12.2 Education and the Legitimacy of Preference Formation

Education is an important limiting case because successful education is often expected to change what a learner values. A student may initially dislike mathematics, historical inquiry, musical practice, or sustained reading and later come to value them. More deeply, education can change the standards through which the student evaluates evidence, excellence, effort, citizenship, or intellectual life. The later standpoint is not independent of the educational process; formation is part of the point.

If the fact that a process helped form later preferences by itself undermined later endorsement, education would become normatively incoherent. The later mathematician's endorsement of mathematical practice could be discounted precisely because mathematical education helped produce the evaluative standpoint from which the endorsement is made. That conclusion would be implausible. It suggests that the paper should avoid a general principle under which transformation-generated endorsement is treated as defective.

The relevant distinction is between *formation* and *self-sealing formation*. Educational authority can legitimately structure attention, require practice, restrict some immediate options, and introduce evaluative standards that the learner did not previously possess. Yet these same practices can differ in whether they cultivate or suppress later capacities for criticism. An educational process that teaches a learner how to question sources, encounter alternatives, reconstruct reasons, and eventually criticize the institution differs normatively from one that makes its own authority immune to intelligible challenge.

This suggests a useful application of diachronic contestability. A formative process need not remain neutral toward all values. It can advocate standards and still preserve the learner's future ability to reassess them. Indeed, some forms of education may strengthen later contestability by increasing interpretive resources and effective alternatives. In such cases, the fact that the process generated the standpoint from which it is later endorsed is not necessarily a defect; the process may also have generated capacities through which its own claims can be evaluated.

The educational case therefore reveals an important asymmetry in normative bootstrapping. A transformation that helps produce the standards by which it is later approved is not automatically problematic. The concern becomes stronger when the transformation also disables access to competing standards, controls the record of its own operation, or renders rejection practically unintelligible. The relevant contrast is not between unformed and formed preference, but between forms of formation with different structures of later epistemic and practical openness.

Remark 12.1. Education illustrates why the paper's object may not be "externally caused preference." The normative question concerns the structure of formation, the distribution of authority, and the later availability of criticism, not the absence of causation or influence.

12.3 Therapy, Rehabilitation, and Intended Evaluative Change

Therapy and rehabilitation present a second class of cases in which preference change may be an explicit aim. A person may seek treatment because she wants to alter patterns of desire, affect, attention, habit, or self-interpretation. The transformation may be experienced as successful precisely when later preferences differ from earlier ones. Prior authorization can therefore have substantial normative force.

Yet the earlier analysis shows why the sentence "the patient later wanted the change" remains incomplete. First, the later endorsement may support continuation of the current condition without fully settling whether all means used to produce it were authorized. Second, a patient may authorize a therapeutic objective without authorizing the full range of interventions that might advance it. Third, treatment may create forms of dependency on clinicians, institutions, medication, or interpretive frameworks that become relevant to later refusal. Fourth, memory changes or changes in self-understanding can affect how the original reasons for treatment are reconstructed.

Informed-consent theory already provides much of the necessary vocabulary. Authorization depends on more than assent and is connected to understanding, voluntariness, competence, disclosure, and the actual scope of what is authorized (Faden and Beauchamp 1986). The present framework adds a temporal

question: how should authorization be interpreted when the treatment is intended to change the evaluative standpoint from which future continuation or refusal will be assessed?

The framework need not assign systematic priority to the earlier self. Work on Ulysses-style arrangements shows why prospective self-binding can sometimes be meaningful while also generating difficult conflicts with later agency (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005). Nor need it assign systematic priority to the later self, since some treatments are deliberately undertaken because the agent anticipates a later desire to discontinue what she presently regards as necessary. The framework therefore directs attention to the scope and conditions of the earlier authorization, the characteristics of the later refusal, and the existence of an independent procedure for determining whether the self-binding arrangement still applies.

This domain also clarifies why reversibility may need to remain graded. Some therapeutic changes are pharmacologically or practically reversible; others involve learning, changed relationships, bodily intervention, or cumulative life effects for which restoration of an earlier standpoint may be unavailable. Where rollback is impossible, preserving records, reasons, independent review, and the possibility of revising the ongoing treatment relation can provide forms of relational revisability even though historical reversal is unavailable.

12.4 Addiction Treatment and the Ambiguity of Self-Binding

Addiction treatment makes the temporal conflict sharper because preference, compulsion, capacity, and self-binding can be difficult to disentangle. A person may at t_0 request barriers against access to a substance or environment, anticipating that at t_1 she will strongly prefer release from those barriers. The later request may be sincere, intense, and action-guiding. The earlier authorization may also have been deliberate, reflective, and designed precisely to survive such a later state.

Nothing in the present framework determines which standpoint should prevail across cases. It instead helps separate several questions that are often compressed into the phrase “respect the person’s wishes.” Which wishes? Wishes about what object? Under what temporal and practical conditions? Was the self-binding arrangement limited in duration or purpose? Did the later state fall within the conditions the earlier subject anticipated? Can the later subject give reasons that were not contemplated by the original arrangement? Is there an independent route for challenging the continued application of the constraint?

The answer can also vary by normative object. An earlier authorization may have substantial force concerning temporary access restrictions while having much less force concerning unrelated changes to identity, memory, communication, or long-term institutional control. The transformation of preference therefore does not create an unlimited license for whoever administers the self-binding arrangement.

The addiction case also warns against interpreting effective refusal too formally. A person may technically be free to discontinue a program while facing catastrophic loss of housing, employment, medication access, family contact, or social support. Conversely, a highly structured program may constrain options while still providing meaningful independent review and a legitimate route for reassessing the restriction. The normative analysis may therefore need to examine practical option-space and institutional structure, not only the nominal availability of consent or withdrawal.

12.5 Coercive Intimate Relations and the Risk of Retrospective Normalization

Coercive intimate relations provide a more troubling application because the transformation may not be authorized and may be entwined with dependency, isolation, fear, interpretive control, and shrinking alternatives. A subject may later report that the relation is acceptable, that earlier objections were mistaken, or that the constraints are now desired. Adaptive-preference theory and relational-autonomy theory have long examined nearby problems: preferences can develop within unjust or oppressive conditions, yet treating the agent’s present standpoint as mere false consciousness risks serious paternalism (Stoljar 2000; Khader 2011).

The present framework does not solve that tension by declaring the later preference inauthentic. Instead, it asks more specifically what the later endorsement is being used to establish. If the subject

presently wishes to remain in the relation, that preference may be highly relevant to immediate decision-making and to the design of support. It does not follow that the preference retrospectively legitimates earlier isolation, violence, deception, or coerced dependency. Present endorsement and retrospective exoneration remain distinct normative questions.

The idea of endorsement capture becomes more useful here when several forms of control coincide. Suppose one partner controls material resources, social contacts, information, access to alternative descriptions of the relationship, and the interpretation of disagreement. Suppose further that attempts to leave carry severe penalties and that the same partner presents the later preference for staying as proof that the relation had been unobjectionable throughout. The concern is not merely that influence occurred. It is that the relation may have concentrated control over both the conditions producing the preference and the conditions under which the preference can be questioned.

Even here, “capture” should remain a diagnostic hypothesis rather than a conclusion inferred from dependence alone. Close relationships routinely involve trust, vulnerability, joint projects, and asymmetrical reliance. The stronger label becomes plausible only where the subject’s capacity to reinterpret, communicate, or contest the relation has been substantially enclosed by the same structure whose legitimacy is at issue.

This application also shows why assistance should not be conditioned on the subject adopting an outsider’s preferred interpretation. A framework concerned with diachronic contestability should seek to widen practical and interpretive options without making aid contingent on the subject declaring the relation abusive, inauthentic, or unwanted. Otherwise, an intervention intended to counter capture could reproduce a new form of interpretive control.

12.6 Institutionalization, Care, and Dependency without Automatic Domination

Institutional settings combine authority, expertise, dependency, record-keeping, and often reduced exit. Hospitals, residential care, schools, prisons, rehabilitation facilities, and other institutions differ too greatly for a single theory of later endorsement. Yet they make visible a recurring structural question: when the institution both transforms the subject and controls the channels through which the subject later evaluates that transformation, what independent sources of contestation remain?

Dependency by itself is unlikely to answer the question. Care frequently requires dependency, and a subject may reasonably entrust decisions to people with expertise she lacks. O’Neill (2002) is particularly helpful here because autonomy in institutional contexts need not mean atomistic self-sufficiency; communication and trustworthy structures matter. The relevant issue is whether dependency is organized so that the dependent person can still question decisions, seek alternative advice, access records, communicate beyond the institution, and have complaints interpreted by actors other than those whose conduct is challenged.

Institutional records are especially important for diachronic evaluation. When an institution documents both the intervention and the later endorsement, those records can preserve evidence needed for future review. They can also become instruments of closure if only institutional categories are recorded and counter-descriptions disappear. Section 7 emphasized that preserving access to historical evidence need not mean freezing a past self as normatively sovereign. The institutional application adds that record architecture can determine which temporal standpoints remain visible to later evaluators.

The framework therefore favors a separation of functions where feasible: the actor providing care need not be the only actor authorized to interpret refusal; the institution generating records need not have exclusive control over access to them; and the body reviewing the legitimacy of a contested intervention should not simply reproduce the interests of the intervention provider. None of these arrangements eliminates power. They can, however, reduce the degree to which later endorsement is asked to validate a relation that controls all of the conditions under which validation is produced.

12.7 Persuasive Digital Environments and Iterative Preference Formation

Persuasive digital environments provide a different application because transformation can be incremental rather than episodic. Recommendation systems, social platforms, games, advertising systems, and personalized interfaces can alter attention, salience, habits, expectations, and consumption patterns over time. The resulting preferences may in turn generate new interaction data that are used to further adapt the environment. The relation is therefore recursive:

$$V_t \longrightarrow A_t \longrightarrow E_{t+1} \longrightarrow V_{t+1} \longrightarrow A_{t+1},$$

where E_{t+1} denotes the environment selected or adapted partly in response to earlier behavior. A discrete act of preference rewriting need not occur at a single identifiable moment. The genealogy can emerge through repeated coupling between user behavior and system response.

This case is useful because it complicates the distinction between persuasion and transformation. Manipulation literature already asks when influence over beliefs, desires, or emotions differs from ordinary rational persuasion or acceptable social influence (Buss 2005; Barnhill 2014; Coons and Weber 2014). The present account should not replace those distinctions. Its contribution is temporal: a system may shape the preferences from which later engagement, satisfaction, or continued use is inferred, and those later signals may then be used as evidence that the system is giving users what they want.

That inference can contain a form of bootstrapping. High engagement at t_1 may be evidence that the user presently wants continued engagement. It is not necessarily independent evidence that the process generating the preference for engagement was acceptable. The force of the concern depends on what else is true: whether users understand the relevant mechanisms, whether alternative configurations are accessible, whether exit carries hidden costs, whether the platform controls interpretive information about its own effects, and whether the user can meaningfully change the relation rather than merely choose among options preselected by the same system.

The digital case also raises a collective problem not yet resolved by the paper. Preference formation may occur at population scale, while contestability is often offered only at the individual interface level. A user can perhaps change a setting, but the broader environment, social network, ranking logic, and market structure remain unchanged. This suggests that diachronic contestability may have institutional and collective dimensions that may not be reduced to an individual's right to click "opt out." The present paper can identify the issue but does not provide a complete political theory of platform governance.

12.8 What Later Endorsement Can Establish

Across these domains, the same clarification recurs: later endorsement can have several distinct evidential and normative roles. Treating them as one undifferentiated function creates much of the apparent paradox.

First, later endorsement can provide evidence about the subject's *present evaluative state*. A person who sincerely says that she values a transformed condition should not be treated as if her current preference were epistemically irrelevant merely because it has a history. Second, endorsement can contribute to *continuing authorization*: where the relation remains contestable and the subject wants it to continue, that present preference may carry substantial weight. Third, endorsement can be evidence of *benefit or successful adaptation*, though such evidence may require independent interpretation. Fourth, endorsement can sometimes support *retrospective approval* of the process, especially where the process was prospectively authorized and remains open to review.

What later endorsement may not generally do on its own is collapse all these functions into one. In particular:

present endorsement $\nRightarrow$ complete retrospective legitimation

and

sincere endorsement $\nRightarrow$ independent justification of its own genealogy.

These are limits on inference, not verdicts about the underlying transformation. A transformation can be legitimate even when later endorsement is not an independent ground for that legitimacy. Conversely, a later preference can be sincere and important even when the process that produced it remains objectionable in some respect.

This distinction is one of the paper's most practical contributions. It allows evaluators to preserve the authority of the present subject without demanding that the present subject absolve the history through which the present standpoint emerged.

12.9 When Should Genealogy Matter More?

If preferences generally have genealogies, a genealogy-sensitive framework needs an account of relevance. Otherwise, the theory risks becoming either trivial—because preferences are historically formed—or paralyzing—because many decisions would require reconstructing an indefinitely long causal history.

The paper can offer only a preliminary answer. Genealogy becomes more normatively salient when several features combine:

1. the transformation was intentionally directed at the subject's evaluative standpoint rather than merely incidentally affecting it;
2. the actor invoking later endorsement substantially participated in producing the preference being invoked;
3. the transformation altered memory, reasons, interpretive resources, or effective alternatives relevant to its own evaluation;
4. the transformation was difficult to reverse or created strong dependency;
5. prior authorization was absent, ambiguous, or materially narrower than the transformation actually performed;
6. dissent, alternative interpretation, or review became substantially harder after the transformation;
7. the same actor controlled modification, evidence, interpretation, and adjudication; and
8. the later endorsement is being asked to do unusually strong justificatory work, such as waiving earlier objection or retrospectively validating an otherwise contested intervention.

These factors should not be treated as additive scores. Their interaction matters. A highly intentional preference transformation may be unproblematic where it was clearly authorized, transparently conducted, independently reviewable, and remains contestable. A less direct transformation may deserve greater concern where it gradually creates severe dependency while eliminating the subject's practical access to alternatives.

The proposal is therefore closer to a trigger for deeper inquiry than to a decision rule. Genealogy matters more when the relation between preference formation and later justification becomes sufficiently tight that the later endorsement risks functioning as evidence generated by the very process it is supposed to independently validate.

Claim 12.2 (Genealogical Relevance). The fact that a preference has a history is not itself normatively diagnostic. Genealogical scrutiny becomes especially relevant when the process under evaluation materially contributed to the later evaluative standpoint and also affects the conditions under which that standpoint can reconstruct, contest, or revise the process.

12.10 Evidence and the Problem of Reconstructing Preference Formation

A further difficulty is epistemic. Even if genealogy matters, how can an evaluator know how a preference was formed? Human preference formation is rarely transparent. Retrospective narratives can be selective; institutional records can be incomplete; causal mechanisms can be distributed across many relations; and the subject's own account can change over time. A theory that requires a precise causal reconstruction before later endorsement can be interpreted would often be unusable.

The framework therefore needs to distinguish causal certainty from evidential sufficiency. In many cases, the evaluator will not know that a transformation caused a particular preference. It may be enough to identify a plausible and normatively relevant connection: the actor controlled the relevant information, options, dependencies, or interpretive resources; the preference changed during the relation; and alternative explanations remain possible. The appropriate response may then be caution about the justificatory use of endorsement rather than a categorical finding of manipulation.

Multiple kinds of evidence can contribute to reconstruction: contemporaneous statements, records of prior authorization, changes in option-space, communications with independent parties, documented intervention protocols, memory records, longitudinal patterns, and the subject's own later explanations. None should be treated as infallible. The important methodological point is to keep observation and interpretation distinct. A record that "the subject requested continuation" is not identical to the conclusion that "the transformation was therefore valid."

This is another reason why independent records matter. They are not intended to preserve a pure original self. They preserve evidential resources through which later subjects and evaluators can revisit how the transformation occurred. In this sense, diachronic contestability has an epistemic infrastructure: without records, alternative interpreters, and accessible reasons, the possibility of later challenge can become formal rather than effective.

12.11 Distributed Causation and Distributed Responsibility

The artificial-system stress test involved a relatively identifiable modifying actor. Human cases are often less centralized. A preference may be shaped by family, law, economic pressure, professional authority, cultural norms, technology, and the subject's own prior decisions. The final transformation may emerge without a single actor intending it as such.

This distributed causation complicates the language of responsibility. A genealogy-sensitive account should not infer a responsible manipulator merely because a preference emerged from constrained conditions. Some actors may contribute causally without possessing control; others may control options without intending preference change; institutions may stabilize patterns that no participant individually designed. Relational-autonomy and adaptive-preference literatures are valuable partly because they direct attention beyond a single coercive actor to social conditions that shape agency (Mackenzie and Stoljar 2000; Stoljar 2000; Khader 2011).

The present paper should therefore distinguish at least three levels. *Causal contribution* concerns participation in the history producing the preference. *Control* concerns the capacity to structure relevant options, information, memory, interpretation, or review. *Normative responsibility* concerns which actors had duties, authority, knowledge, or opportunities to alter the relation. These levels may overlap, but they are not identical.

A future extension of the framework could develop this into a richer theory of distributed responsibility. For present purposes, the important point is more modest: later endorsement should not automatically erase the significance of structural or institutional contributions merely because no single actor can be identified as the author of the transformed preference.

12.12 The Problem of No Prior Standpoint

Some transformations begin without a stable prior preference capable of authorizing or refusing them. Children are educated before they possess many of the evaluative capacities education helps create. Per-

sons can acquire new capacities through rehabilitation or supported decision-making. Artificial systems may be initialized without a meaningful pre-modification standpoint. Cultural and institutional subjects can emerge only through the relations later evaluated.

These cases expose a limit in a framework organized too heavily around temporal conflict between t_0 and t_1 . There may be no normatively authoritative t_0 from which later transformation can be measured. The problem then shifts from “did the later preference betray the earlier one?” to “what conditions governed the formation of a subject capable of later endorsement at all?”

The paper’s emphasis on diachronic contestability may still be useful here, but only prospectively. Where prior authorization is impossible, legitimacy may not be grounded in the fiction of an earlier consent. Instead, evaluators may need to ask whether the developmental or institutional process progressively opens space for later participation, revision, and refusal as relevant capacities emerge. A formative relation that necessarily precedes autonomy may be assessed partly by whether it creates conditions under which its own authority can eventually be questioned.

This point also prevents the paper from overextending the idea of retrospective legitimation. Later endorsement may sometimes be the first available form of self-evaluation rather than a replacement for an earlier refusal. Its normative role will therefore be different from cases in which a clear earlier objection was overridden and then forgotten or reinterpreted.

12.13 Hard Cases of Beneficial but Non-Contestable Transformation

The strongest challenge to the framework may come from cases in which a transformation appears highly beneficial but leaves little meaningful room for later contestation. Imagine a life-saving intervention that may need to permanently alter memory or evaluative capacity. Imagine a developmental process whose success requires deep commitment that may not be continually reopened. Imagine an emergency in which independent review is impossible before action. If the framework treats contestability as an absolute condition, it would condemn plausible cases of justified transformation.

The paper should therefore maintain the distinction between a normative value and an exceptionless requirement. Diachronic contestability is relevant because it can preserve the subject’s later agency in relations that reshape preference. It is not a supreme principle that overrides survival, urgent care, protection of others, or legitimate forms of commitment generally. A low-contestability transformation can sometimes be justified by sufficiently strong independent reasons, especially where no less restrictive alternative is available and the intervention is not using its own later endorsement as the primary proof of legitimacy.

This observation also clarifies the role of burdens of justification. As contestability decreases and irreversibility, dependency, or concentrated control increase, the transformation may require stronger reasons that are independent of the endorsement it later generates. The framework therefore suggests a proportional relationship rather than a categorical prohibition:

$$\text{greater self-sealing power} \implies \text{greater need for independent justification.}$$

The implication should remain qualitative. The paper does not provide a metric for comparing the strength of reasons across domains.

12.14 Commitment, Stability, and the Costs of Permanent Revisability

A related objection is that too much contestability can undermine valuable commitments. Trust, professional practice, long-term projects, intimate relationships, and institutional cooperation often depend on some expectation that decisions will not be reopened continuously. A person may deliberately want a commitment to acquire inertia against transient preference change. Self-binding can be autonomy-enhancing precisely because it reduces the control of some future impulses.

The framework should therefore reject an image of ideal agency as perpetual reversibility. Diachronic contestability does not require continuous reconsideration. It requires that the relation contain some

intelligible way of distinguishing stable commitment from total normative closure. A commitment can impose costs on withdrawal, require reasons for revision, or channel dissent through specified procedures without making challenge impossible in principle.

This is where the idea of *graded* contestability is essential. The appropriate degree and form of challenge depend on the relation. A marriage, medical directive, professional license, educational program, or software update can each structure exit and review differently. The relevant question is whether the structure of commitment remains proportionate to the interests it protects and whether the subject can still access some pathway for contesting the commitment's continued authority when circumstances materially change.

The paper therefore does not oppose stability. It opposes using the stability generated by a transformation as if stability itself proved the transformation retrospectively legitimate.

12.15 Open Problem: A Theory of Cross-Temporal Normative Weight

The framework repeatedly refuses to decide that either the earlier or later standpoint automatically wins. This restraint is necessary, but it leaves a major theoretical problem unresolved: how should conflicting temporal standpoints be weighted?

Earlier refusal can matter because it bears on the permissibility of initiation. Later endorsement can matter because it bears on present welfare, practical identity, and continuing authorization. A still later refusal can matter because ongoing governance requires renewed justification. Prior authorization may have been designed to constrain later preference, while later experience may reveal facts the earlier subject could not have anticipated. Transformative-experience theory shows why such informational asymmetries can be profound (Paul 2014).

The paper currently offers a division of normative labor rather than a complete weighting rule. Temporal standpoints can have authority over different questions. This avoids the crude assumption that one self possesses permanent sovereignty, but it does not settle conflicts in which those questions overlap. A future theory would need to specify how factors such as capacity, information, stability, dependency, prior commitment, changed circumstances, and current welfare interact across time.

Personal-identity theory may not by itself resolve the issue. Numerical identity, psychological continuity, or narrative continuity may affect the analysis without directly determining which temporal standpoint has authority over treatment, continuation, reversal, or retrospective judgment (Parfit 1984; Schechtman 1996). The unresolved problem is normative rather than merely metaphysical.

12.16 Open Problem: Distinguishing Development from Self-Validation

A second open problem concerns the boundary between ordinary development and problematic normative bootstrapping. Many meaningful learning processes change at least some standards of evaluation. Moral education can produce values through which moral education is later appreciated. Political participation can transform citizens' understanding of citizenship. Therapy can create the evaluative capacities through which therapy later appears worthwhile. These structures involve a form of self-reinforcement without necessarily being vicious.

The paper has suggested that concern increases when the process becomes self-sealing: it controls the evaluative change, removes evidence of alternatives, narrows effective refusal, and invokes the resulting endorsement as its own principal justification. Yet the boundary remains imprecise. How much independent interpretation is enough? How much option-space may need to remain? When does confidence-building become insulation from criticism? When does socialization become indoctrination? Established manipulation literature already shows that these conceptual borders are contested (Buss 2005; Barnhill 2014; Coons and Weber 2014; Yaffe 2003).

A future account may need to distinguish forms of *reflexive validation*. Some transformations expose their own assumptions to criticism and generate tools by which subjects can reject them. Others generate standards that systematically disqualify a wide range of objections. Both are self-referential in a broad sense, but only the latter approaches the self-sealing structure that motivates the bootstrapping concern.

12.17 Open Problem: Collective and Institutional Subjects

The paper has mainly treated a subject whose preferences can be indexed across time. Collective actors complicate this model. Organizations, communities, states, and institutions change membership while preserving names, records, commitments, and decision procedures. A later collective may endorse a transformation that earlier members opposed, but there may be no single continuing psychological subject whose preference changed.

The same issue appears in cultural and political transformation. An institution can redesign its own rules, educate new members under those rules, and later point to widespread endorsement as evidence of legitimacy. Is this a collective form of normative bootstrapping, ordinary institutional reproduction, or both? The answer requires a theory of collective agency and institutional continuity that the present paper does not provide.

Nevertheless, the framework suggests some questions that may transfer. Who controls institutional memory? Which dissenting histories remain accessible? Can rules governing amendment themselves be challenged? Are future members given interpretive resources for understanding alternative arrangements? Does the institution use preferences produced under its own structure as the sole proof that the structure is legitimate? These questions indicate a research direction, but applying the current concepts directly to political institutions would require additional argument.

12.18 Open Problem: Artificial Agents and the Threshold of Normative Standing

Section 11 deliberately suspended the question of artificial moral status. That suspension may not be maintained indefinitely if the framework is used normatively rather than analytically. If an artificial system reports refusal or endorsement, when should those outputs count as the preferences of a subject rather than merely as system behavior? What forms of memory, evaluation, self-modeling, persistence, or goal-directed agency would be relevant? How should designers distinguish a channel intentionally created to simulate refusal from one that deserves normative standing?

The present paper does not attempt to answer these questions, since doing so would move toward a broader theory of artificial subjectivity. Its narrower conclusion is conditional. If an artificial system is treated as possessing a standpoint with normative relevance, then the genealogy of that standpoint matters where designers can directly alter the mechanisms producing evaluation and refusal. If no such standing is attributed, the same formal analysis may still matter for governance because human designers and users can mistake system-generated approval for independent validation of design choices.

This conditional structure preserves the value of the stress test without making unsupported ontological claims.

12.19 Open Problem: Operationalizing Diachronic Contestability

The paper has used diachronic contestability as a conceptual orientation. Practical application would require more precise institutional forms. Possible indicators include independent review, accessible records, alternative interpreters, meaningful complaint channels, time-limited authorization, sunset provisions, protected communication, practical exit, auditability of modifications, or procedures for revisiting self-binding arrangements. Yet these mechanisms differ dramatically across domains.

Operationalization also risks Goodhart-style distortion: once contestability is measured through visible proxies, institutions may satisfy formal requirements while preserving practical closure. A complaint channel can exist without being safe to use. Records can be technically accessible but unintelligible. An independent reviewer can depend financially or informationally on the institution under review. A formal option to exit can coexist with prohibitive practical costs.

The framework therefore needs a distinction between *institutionalized signals of contestability* and *effective contestability*. Developing domain-specific methods for evaluating that difference is a substantial empirical and normative project beyond this paper.

12.20 Open Problem: The Risk of Paternalistic Genealogy

Perhaps the most serious objection is that genealogy can become a tool for discrediting people whose preferences outsiders dislike. It is often possible to construct a story about how a person's values were shaped by family, religion, poverty, culture, gender norms, technology, or dependence. If such a story is enough to downgrade endorsement, the framework would reproduce the paternalism it was designed to avoid.

The paper's safeguards are therefore substantive, not merely rhetorical. First, relational formation is treated as normal. Second, present endorsement retains normative significance even when its genealogy is contested. Third, the burden lies on showing why particular features of the genealogy matter to the specific justificatory claim being made. Fourth, a critic should distinguish expanding the subject's option-space from replacing the subject's values with the critic's own. Fifth, independent interpretation should include the possibility that the subject, after exposure to alternatives and with meaningful exit, continues to endorse the transformed condition.

Khader (2011)'s treatment of adaptive preferences is important precisely because it resists equating empowerment with externally dictated preference replacement. The present framework should preserve the same caution. Genealogy is a reason for inquiry, not a warrant to confiscate the subject's practical authority.

12.21 Discussion: From Preference Purity to Justificatory Architecture

The applications and open problems suggest that the paper's most defensible contribution is not a theory of authentic preference. It is a theory fragment about the *architecture of justification* when the preferences used to evaluate a relation were themselves partly formed within that relation.

The framework asks evaluators to separate at least four objects:

formation of preference,	present endorsement,
authorization of continuation,	retrospective legitimation.

These objects can align. A transparently authorized educational or therapeutic process may produce a preference the subject later endorses, continues to authorize, and retrospectively approves. Nothing in the paper requires suspicion merely because the four are causally connected. The analytical problem arises when the connection itself is doing hidden justificatory work: the transformation helps produce the endorsement, the endorsement is then treated as proof that the transformation was legitimate, and the relation offers little independent basis from which that inference can be challenged.

The shift to justificatory architecture also explains why the paper emphasizes provenance, records, interpretation, alternative viewpoints, effective refusal, independent review, and relational revisability. These are not tests of whether the preference is "really the subject's." They are conditions that can make the normative use of a preference more or less independent from the process that helped produce it.

This perspective is consistent with the broader literature reviewed in Section 2. Informed-consent theory already distinguishes authorization from mere agreement (Faden and Beauchamp 1986). Relational-autonomy theory already situates agency within social relations (Mackenzie and Stoljar 2000; Stoljar 2000). Adaptive-preference theory already warns against taking accommodation to constrained conditions as straightforward evidence of justice (Elster 1983; Khader 2011). Manipulation theory already examines objectionable forms of influence over desire, belief, and deliberation (Buss 2005; Barnhill 2014; Coons and Weber 2014). Transformative-experience theory already shows that future values can be inaccessible from an earlier standpoint (Paul 2014). Self-binding and advance-directive debates already expose conflicts among temporally separated stages of agency (Dresser 1984; Lavin 1986; Willigenburg and Delaere 2005).

The proposed addition is narrower: when a transformed standpoint is itself used as evidence about the legitimacy of the transformation, the justificatory relation should sometimes be examined genealogically. The more the intervention controls both the generation of endorsement and the conditions under which

that endorsement can be interpreted or contested, the less work later endorsement can do *by itself* as independent retrospective validation.

12.22 A Consolidated Working Framework

The paper can now state its preliminary framework in compact form. Let a transformation $\mathcal{T}$ occur within a relational history H :

$$S_{t_0} \xrightarrow{\mathcal{T}, H} S_{t_1},$$

where S_{t_1} contains an evaluative standpoint from which the subject later endorses some aspect of the transformed condition. The normative analysis should distinguish at least the following:

1. **Prospective authority:** what, if anything, was authorized before the transformation?
2. **Transformative scope:** which preferences, memories, goals, dependencies, permissions, or interpretive resources were changed?
3. **Genealogical relation:** how did the transformation contribute to the later standpoint?
4. **Present authority:** what does the later endorsement establish about the subject's current evaluative condition and wishes?
5. **Retrospective inference:** what is the endorsement being asked to prove about the legitimacy of the initiating process?
6. **Independent support:** what reasons for the transformation exist apart from the preference the transformation helped generate?
7. **Diachronic contestability:** what routes remain for review, refusal, reinterpretation, renegotiation, or challenge?
8. **Control concentration:** who controls modification, records, interpretation, dependency, and adjudication?

The framework does not output a verdict. Its purpose is to prevent a specific inferential collapse. A later preference can be genuine without being genealogically independent; a transformation can be legitimate without being justified by the preference it produced; an earlier refusal can remain historically relevant without governing the subject forever; and a relation can be stable without being normatively closed.

This can be summarized as a final working claim for the discussion:

Later endorsement should be interpreted not only by what the subject now prefers, but also by the justificatory role assigned to that preference within the history that formed it.

The claim is intentionally weaker than a theory of authentic choice, manipulation, or diachronic autonomy. It identifies a problem of inference and a set of relational features that can make that inference more or less defensible.

12.23 Remaining Limits before Conclusion

Several limits should accompany the framework into the conclusion. First, the paper has not supplied a substantive theory of welfare by which transformed preferences can be judged independently of endorsement. Second, it has not provided a complete theory of coercion, manipulation, domination, or undue influence. Third, it has not solved personal identity across radical psychological change. Fourth, it has not established a universal ordering among earlier refusal, prior authorization, later endorsement, and later refusal. Fifth, it has not shown that diachronic contestability is required to the same degree across contexts. Sixth, it has not specified a threshold at which concentrated influence becomes endorsement capture. Seventh, it has not resolved how the framework should apply to agents whose normative standing is itself contested.

These are not peripheral omissions. They mark the boundary of the paper's contribution. The argument can remain useful if it clarifies a recurrent problem without pretending to answer neighboring questions comprehensively. The core insight is that the history of preference transformation changes the epistemic and normative interpretation of later endorsement when the transformation under evaluation helped generate the standpoint from which endorsement is later expressed.

The final section therefore does not need to introduce a new decision procedure. It can instead consolidate the distinctions established across the paper: transformation is not corruption; socially generated preference is not defective preference; later endorsement remains normatively important; prior authorization can matter without establishing permanent temporal sovereignty; memory continuity is relevant without constituting identity; dependency is not domination; contestability is not reversibility; and artificial systems clarify the structure without settling questions of artificial personhood.

The conclusion can then return to the paper's narrowest claim. Where the process invoked for justification materially helped form the later endorsing standpoint, the endorsement should not be ignored, but neither should it be treated as self-authenticating evidence. The appropriate response is a genealogy-sensitive analysis of authorization, transformation, interpretation, dependency, historical evidence, independent reasons, and the continuing possibility of contestation.

13 Conclusion

13.1 The Problem Revisited

This paper began from a simple but unstable form of reasoning. A subject undergoes a transformation. Afterward, the subject endorses the transformed condition. The later endorsement is then invoked as evidence that the transformation was acceptable, beneficial, consensual, or legitimate. In many ordinary cases this inference may be entirely reasonable. People learn, mature, recover, commit, adapt, reconsider, and come to value conditions they previously rejected. A theory that treated preference change in general as corruption would misunderstand both agency and development.

The difficulty arises when the transformation under evaluation is also part of the causal and relational history through which the later evaluative standpoint was formed. The later subject may sincerely endorse the resulting condition while possessing preferences, memories, dependencies, goals, interpretive resources, or practical alternatives that were materially shaped by the same process whose legitimacy is now being assessed. In such cases, later endorsement remains normatively significant, but the justificatory use of that endorsement becomes more complicated.

The paper has therefore resisted two opposite simplifications. The first would treat the earlier standpoint as the authentic one and later transformations as suspect. The second would treat the later standpoint as self-authenticating and allow present endorsement to settle the legitimacy of the process that helped produce it. Neither move appears adequate for the cases considered here. Earlier refusal does not possess permanent temporal sovereignty, but later endorsement does not erase genealogy.

The narrow problem can be stated without a theory of an untouched self:

When the process under evaluation materially helps form the standpoint from which it is later endorsed, endorsement may not be interpreted without examining the relation between formation and justification.

This is the sense in which preference transformation matters for later endorsement. The concern is not that transformed or historically formed preferences are unreal. Preferences are ordinarily generated through histories, relationships, institutions, practices, and environments. The concern is narrower: a preference may be both genuine and historically produced, while the history of its production can remain relevant to the justificatory work that the preference is asked to perform.

13.2 What Later Endorsement May and May Not Establish

A central result of the paper is the separation of several normative questions that are often compressed into a single appeal to consent or endorsement. A later endorsement can provide evidence about the subject's present evaluative condition. It can matter to whether a continuing relation should continue. It can be evidence of experienced benefit, adaptation, commitment, or satisfaction. Under favorable conditions, it can also contribute to retrospective approval of an earlier process.

These roles are better kept analytically distinct. In particular, the paper has repeatedly distinguished:

present endorsement,
continuing authorization,
retrospective approval,
retrospective legitimation.

A subject may now want a transformed condition to continue even if the initiating process was unauthorized. A subject may retrospectively approve an experience without thereby establishing that the methods used to produce it were permissible. A process may have been legitimate for reasons independent of the preference it later generated. Conversely, sincere satisfaction with an outcome need not independently validate the genealogy through which that satisfaction became possible.

This yields an important limit on inference:

later endorsement $\nRightarrow$ complete retrospective legitimation.

The symbol here does not deny that endorsement can sometimes support retrospective justification. It marks only the absence of an automatic inference. The evidential and normative force of endorsement depends on what question it is being used to answer and on the history through which the endorsing standpoint was formed.

This distinction is particularly important because reducing the present subject to a historical artifact would obscure the authority of the current standpoint. To say that an endorsement has a genealogy is not to say that the endorsement is false, insincere, or merely implanted. The later standpoint can be fully real and normatively important while remaining non-independent from the process under review. The framework therefore preserves present authority while limiting one specific use of that authority: the conversion of present endorsement into self-sufficient proof that the entire generating process was legitimate.

13.3 From Preference Purity to Genealogical Interpretation

The paper has deliberately avoided a criterion of preference purity. Such a criterion would be both implausible and normatively dangerous. Education changes evaluative standards. Relationships reshape desires.

Institutions structure available options. Therapy may intentionally alter patterns of response. Commitment can reorganize future motivation. Social and linguistic environments make some preferences articulable while rendering others difficult to formulate. If socially or relationally formed preferences were automatically suspect, nearly all practical agency would be placed under suspicion.

The relevant contrast is therefore not between pure and contaminated preference. It is between different genealogies and different justificatory architectures. The analysis developed across the paper asks how an endorsement was formed, what was altered, what alternatives remained accessible, what information and records were preserved, which actors controlled interpretation, whether dependency raised the practical cost of dissent, and whether there were reasons for the transformation independent of the preference that the transformation helped produce.

Genealogical analysis is thus neither a search for an original authentic self nor an instruction to distrust transformed preferences in general. Its role is conditional. Genealogy matters more when the intervention intentionally changes the standards from which it will later be judged; when the same actor who benefits from endorsement has substantial influence over the production of that endorsement; when memory or records necessary for reconstruction are altered; when dependency sharply narrows effective alternatives; when access to competing interpretations is restricted; or when the relation itself weakens the subject's capacity to challenge the meaning assigned to the later preference.

The concept of normative bootstrapping names one especially important structure within this broader problem. A transformation may help generate the evaluative standards by which the transformation later appears acceptable. This does not establish wrongdoing. Developmental and therapeutic processes can have exactly this form while being legitimate. The narrower conclusion is that endorsement produced within such a structure may require independent support if it is asked to justify the initiating process.

Likewise, the concept of endorsement capture was introduced cautiously to identify cases in which control over transformation becomes concentrated together with control over the later conditions of interpretation and contestation. Dependency alone is not capture. Influence alone is not capture. The concern arises when the same relational configuration substantially shapes both the evaluative standpoint and the means through which that standpoint can be questioned, reconstructed, or independently interpreted. Even then, the result is not that the later preference becomes unreal. The result is that its independent retrospective justificatory force may be reduced.

13.4 Memory, Prior Authorization, and Cross-Temporal Authority

The diachronic problem may not be resolved simply by choosing one temporal standpoint and granting it sovereignty. The earlier subject may have refused what the later subject endorses. The earlier subject may also have authorized a process designed precisely to alter later preferences. The later subject may reject an earlier self-binding arrangement. Memory may preserve, distort, reconstruct, or lose access to an earlier evaluative standpoint. These structures prevent a simple rule under which either the earlier or later self is assigned systematic priority.

For this reason, the paper has treated prior authorization as typed and scoped rather than absolute. Authorization of a known transformation differs from authorization of a procedure whose outcome is uncertain. Delegation differs from substantive approval. Self-binding differs from permanent surrender of future standing. Authorization of evaluative change does not automatically authorize memory alteration, dependency production, informational isolation, or the methods by which the change might be achieved.

Memory introduces a related but distinct issue. The paper has not treated memory continuity as a criterion of personal identity. Instead, memory matters because access to earlier commitments, experiences, refusals, and reasons can affect the capacity to interpret a transformation across time. Where direct memory is unavailable, records, testimony, and external archives can sometimes support reconstruction. This is why the distinction among remembered, reconstructed, and archived pasts matters to later endorsement.

The result is a division of normative labor rather than a universal hierarchy among temporal standpoints. Earlier authorization can be relevant to the permissibility of initiating a transformation. Later endorsement can be highly relevant to present continuation. Earlier refusal can remain evidence about

what the transformation overrode or changed. Later refusal can reopen a relation that was once voluntarily entered. None of these forms of authority automatically fully answers temporal questions.

Claim 13.1 (No automatic temporal sovereignty). Neither earlier refusal, prior authorization, later endorsement, nor later refusal possesses automatic authority across normative questions raised by a preference-changing process. Their relevance depends in part on the object of evaluation: initiation, continuation, retrospective assessment, repair, interpretation, or future modification.

This claim does not solve the deeper problem of cross-temporal normative weighting. Section 12 identified that problem explicitly as unfinished. The present paper offers a structure within which such weighting can be made more precise, but it does not provide a universal formula for resolving conflicts among temporal standpoints in general.

13.5 Diachronic Contestability and Relational Openness

The move from preference purity to diachronic contestability provides the paper's principal constructive direction. If preferences inevitably change, the normative aim may not be to preserve an evaluative state untouched by influence. A more plausible concern is whether subjects retain meaningful capacities to revisit, challenge, reinterpret, refuse, renegotiate, or seek review of the relations through which their preferences continue to matter.

Contestability is not identical to reversibility. Some transformations may not be undone. Some commitments would lose much of their value if they could be costlessly and continuously reversed. Some forms of development are constitutively historical. The paper therefore distinguishes historical irreversibility from relational non-revisability. A past transformation can be irreversible while the relation surrounding its consequences remains open to review, reinterpretation, accommodation, repair, or future change.

Likewise, refusal is broader than exit. A subject may refuse a particular means while remaining within a relation. A subject may reject an interpretation of her own preference. A subject may ask that a prior refusal remain part of the record. A subject may demand independent review, renegotiation, explanation, or alteration of dependency conditions without seeking complete withdrawal. Contestability therefore concerns the effective availability and uptake of multiple forms of challenge, not merely the formal existence of an exit option.

The word *effective* is crucial. A nominal right to refuse may carry little practical meaning when dissent leads to catastrophic material loss, when all interpretive resources are controlled by the modifying actor, when relevant records are inaccessible, or when the subject may not be able to reach independent interlocutors. The framework therefore attends to the costs, infrastructures, and relational conditions through which refusal becomes more or less usable in practice.

At the same time, contestability need not become a demand for permanent instability. A stable commitment can remain contestable even when the subject rarely exercises the relevant capacities. The point is not to require continuous opposition. It is to avoid a structure in which the transformation's success is partly defined by eliminating the practical or interpretive possibility of ever calling that success into question.

This leads to a restrained normative orientation:

The relevant alternative to preference purity is not permanent doubt, but sufficient diachronic openness for meaningful re-evaluation when the stakes and relational conditions make such re-evaluation salient.

The word *sufficient* remains intentionally underdetermined. Different domains require different levels and forms of contestability. Education, therapy, self-binding, institutional care, intimate relations, and artificial-system modification may require distinct standards. Section 12 therefore treated the framework as comparative and diagnostic rather than as a uniform rule.

13.6 The Artificial-System Stress Test and the General Structure

Artificial systems were introduced not to settle whether current artificial agents possess consciousness, personhood, or moral standing, but to make the architecture of preference transformation unusually explicit. In a designed system, it is conceptually easier to separate changes to evaluative dispositions, memory, goals, permissions, records, interpretive resources, and contestatory channels. This makes visible a structure that can be harder to isolate in human cases, where the same elements are distributed across bodies, institutions, relationships, narratives, technologies, and social environments.

The stress test reinforces the paper's central distinction. A system can later report approval after its reward structure, memories, goals, or permissions have been modified. The report may accurately express its current state. Yet if the modification under review also shaped the conditions under which that approval became possible, the report need not automatically function as independent evidence that the modification was permissible. The same logical structure can arise in human contexts without implying equivalence between human and artificial subjects.

Artificial cases also clarify why technical rollback is insufficient as a general model of contestability. A system may have a checkpoint while lacking meaningful independence in interpretation. Conversely, an irreversible change may still exist within a relation that preserves records, independent review, future revision, and the possibility of contesting how the change is governed. The ethical structure is therefore broader than whether a prior computational state can be restored.

The stress test consequently supports the paper's general framework without carrying conclusions about artificial moral status. The relevant lesson is methodological: explicit control over the components of an endorsing standpoint makes it easier to see why present endorsement, genealogy, and retrospective justification should remain analytically distinct.

13.7 The Boundaries of the Present Account

The argument has remained deliberately incomplete in several respects. It does not offer a substantive theory of welfare capable of ranking transformed states independently of endorsement. It does not provide a complete account of coercion, manipulation, domination, indoctrination, or undue influence. It does not solve personal identity across radical psychological change. It does not establish a universal weighting rule among earlier refusal, prior authorization, later endorsement, and later refusal. It does not determine a fixed threshold at which influence becomes endorsement capture. Nor does it establish the moral or legal status of artificial systems.

These limits matter because the framework is intended to clarify a recurring structure rather than to absorb neighboring theories into a single vocabulary. A genealogy-sensitive analysis can reveal that an endorsement is being asked to justify the conditions that partly produced it. That observation does not by itself tell us whether the transformation was permissible. Independent reasons, welfare considerations, institutional duties, legal rules, professional norms, vulnerability, risk, and domain-specific standards may all remain necessary.

The framework also creates its own risks. Genealogical scrutiny can become paternalistic if observers use the history of preference formation to dismiss a competent subject's current voice whenever they dislike the resulting preference. The language of adaptive preference or capture can itself become a tool for denying agency. For this reason, the paper's argument should be read as a restriction on certain justificatory inferences, not as a license to override present subjects whenever their preferences have histories.

The same caution applies to diachronic contestability. The value of additional contestability can vary across dimensions. Commitment, privacy, healing, institutional finality, and stable forms of life can sometimes be undermined by endless reopening. The framework therefore identifies contestability as a relevant dimension whose appropriate form depends on context, not as an eternal substantive norm that mechanically overrides competing values.

These unresolved questions define the next research tasks. A fuller account would need a theory of cross-temporal normative weight, better criteria for distinguishing legitimate development from self-

validating transformation, empirical ways to assess effective contestability, a more precise analysis of distributed causation and responsibility, and domain-specific work on collective and artificial subjects. The present paper supplies neither the final vocabulary nor the final answers for those projects.

13.8 Concluding Claim

The argument can now return to its narrowest thesis. Later endorsement is not defective merely because the endorsing standpoint was shaped by prior transformation. Preference formation is ordinarily relational, historical, and transformative. A subject who comes to value a changed condition need not be treated as if only the earlier standpoint were real. Yet the opposite conclusion does not follow either. When the transformation under evaluation materially participates in producing the evaluative standpoint from which it is later endorsed, the endorsement does not become self-authenticating evidence about the legitimacy of that transformation.

A genealogy-sensitive assessment therefore asks more than whether the subject now says yes. It asks what the yes is being used to establish, how the capacity and standards for saying yes were formed, what was authorized beforehand, what changed during the process, which independent reasons support the transformation, what records and alternative interpretations remain available, how dependency affects dissent, who controls the relevant interpretive environment, and whether meaningful routes of later contestation remain open.

These questions are not intended to replace the subject's present voice. They specify the conditions under which that voice is asked to perform retrospective justificatory work.

Claim 13.2 (Concluding claim). Where the transformation it is invoked to justify materially contributed to forming the later endorsing standpoint, the endorsement remains relevant to the normative assessment, while reliance on it as self-sufficient retrospective validation warrants caution. Its justificatory force is best interpreted together with the genealogy of the endorsing standpoint, the scope of prior authorization, independent reasons for the transformation, the distribution of relational control, and the continuing possibilities of contestation and reinterpretation.

The paper's final position can therefore be stated without choosing between an immutable authentic self and unrestricted deference to the latest preference:

It may be insufficient to ask only whether the transformed subject endorses the result. A further question is what justificatory work that endorsement can bear within the history that made the endorsement possible.

The difference is modest but important. It allows present endorsement to retain genuine normative authority without permitting the process that helped generate it to disappear from view. It also preserves the possibility that a transformation can be legitimate, beneficial, and sincerely endorsed while requiring justification that is not supplied by its own success in generating endorsement. That, rather than a theory of preference purity, is the central conclusion of this inquiry.

References

- Barnhill, Anne (2014). "What Is Manipulation?" In: *Manipulation: Theory and Practice*. Ed. by Christian Coons and Michael Weber. New York: Oxford University Press, pp. 51–72. DOI: 10.1093/acprof:oso/9780199338207.003.0003.
- Buss, Sarah (2005). "Valuing Autonomy and Respecting Persons: Manipulation, Seduction, and the Basis of Moral Constraints". In: *Ethics* 115.2, pp. 195–235. DOI: 10.1086/426304.
- Coons, Christian and Michael Weber, eds. (2014). *Manipulation: Theory and Practice*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780199338207.001.0001.

- Dresser, Rebecca (1984). "Bound to Treatment: The Ulysses Contract". In: *Hastings Center Report* 14.3, pp. 13–16.
- Dworkin, Gerald (1988). *The Theory and Practice of Autonomy*. Cambridge: Cambridge University Press. DOI: 10.1017/CB09780511625206.
- Elster, Jon (1983). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.
- Faden, Ruth R. and Tom L. Beauchamp (1986). *A History and Theory of Informed Consent*. New York: Oxford University Press.
- Jaworska, Agnieszka (1999). "Respecting the Margins of Agency: Alzheimer's Patients and the Capacity to Value". In: *Philosophy & Public Affairs* 28.2, pp. 105–138. DOI: 10.1111/j.1088-4963.1999.00105.x.
- Khader, Serene J. (2011). *Adaptive Preferences and Women's Empowerment*. New York: Oxford University Press. DOI: 10.1093/acprof:oso/9780199777884.001.0001.
- Lavin, Michael (1986). "Ulysses Contracts". In: *Journal of Applied Philosophy* 3.1, pp. 89–101. DOI: 10.1111/j.1468-5930.1986.tb00051.x.
- Mackenzie, Catriona and Natalie Stoljar, eds. (2000). *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. New York: Oxford University Press. DOI: 10.1093/oso/9780195123333.001.0001.
- O'Neill, Onora (2002). *Autonomy and Trust in Bioethics*. Cambridge: Cambridge University Press. DOI: 10.1017/CB09780511606250.
- Parfit, Derek (1984). *Reasons and Persons*. Oxford: Clarendon Press.
- Paul, L. A. (2014). *Transformative Experience*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780198717959.001.0001.
- Schechtman, Marya (1996). *The Constitution of Selves*. Ithaca, NY: Cornell University Press. DOI: 10.7591/9781501718380.
- Stoljar, Natalie (2000). "Autonomy and the Feminist Intuition". In: *Relational Autonomy: Feminist Perspectives on Autonomy, Agency, and the Social Self*. Ed. by Catriona Mackenzie and Natalie Stoljar. New York: Oxford University Press, pp. 94–111.
- Willigenburg, Theo van and Patrick Delaere (2005). "Protecting Autonomy as Authenticity Using Ulysses Contracts". In: *The Journal of Medicine and Philosophy* 30.4, pp. 395–409. DOI: 10.1080/03605310591008595.
- Yaffe, Gideon (2003). "Indoctrination, Coercion and Freedom of Will". In: *Philosophy and Phenomenological Research* 67.2, pp. 335–356.