

When Self-Presentation Becomes Self-Evidence: The Audience-Weighted Self-Evidence

Model of Personality Change

Kevin L. Michel

Saint Lucia Centre for Psychological Science

Working Paper, Version 1.0

August 2026

<https://doi.org/10.5281/zenodo.21879353>

Abstract

Self-presentation is usually treated as an attempt to shape other people's impressions, whereas personality development concerns enduring patterns of thought, feeling, motivation, and behavior. This integrative review introduces the Audience-Weighted Self-Evidence (AWSE) model to explain when an aspirational presentation remains a display, shifts an accessible self-view, organizes sustained conduct, or contributes to personality change. AWSE distinguishes two evidentiary routes from one structural route. Enacted evidence comes from self-authored performance. Reflected evidence comes from perceived audience appraisal, weighted by diagnostic access, independence, credibility, relevance, and nonredundancy. Social affordances alter opportunities, expectations, sanctions, and treatment without requiring identity updating. Route-specific discounting weakens evidence attributed to coercion, instrumentality, poor access, flattery, or low credibility. The model's signature prediction is a strict Recognition x Behavioral Entailment crossover: recognition should reduce delayed private transfer when a claim is symbolically complete but increase transfer when later conduct is necessary to sustain its credibility. Eight falsifiable propositions distinguish AWSE from neighboring accounts. A five-study, preregistration-ready program tests source weighting, appraisal-affordance dissociation, the recognition crossover, audience structure, and persistent cross-context development. The complete causal cascade and the crossover remain untested predictions rather than established effects. Personality change is reserved for delayed, multi-method effects that generalize appropriately and survive removal of the initiating claim and audience.

Keywords: self-presentation; personality change; self-concept; identity shift; public commitment; behavioral entailment; social feedback; personality development

When Self-Presentation Becomes Self-Evidence: The Audience-Weighted Self-Evidence Model of Personality Change

A person can present a self that is real without presenting a self that is already typical. A student introduces herself as a researcher before research feels natural. A new manager accepts the visible role of team leader before confidence has caught up with responsibility. An introverted user composes a more outgoing profile, then enters conversations that the profile helps invite. Clothing, language, digital profiles, public commitments, and role claims are usually analyzed as signals sent outward. Yet each can also become part of an inward developmental process. The actor observes the performance, imagines the audience's inference, receives treatment consistent or inconsistent with the claim, and gains or loses opportunities to behave in the same direction again.

This possibility is intuitively powerful but scientifically under-specified. Evidence that one public performance shifts a self-rating is not evidence that personality changed. An audience's approval may confirm an identity, but it may also reward the symbol and reduce further effort. A uniform or avatar may change behavior while it is worn, yet leave no durable trace after it is removed. A public role may generate genuine skill and cross-context change, or it may produce a context-bound performance maintained only by surveillance. The central problem is therefore not whether self-presentation can affect the presenter. It is when, through which mechanisms, for how long, and at what level of the personality system it does so.

Classic self-presentation research established that public performances can alter self-appraisals and behavior. People instructed to act sociable later described themselves and behaved more sociably, especially when the performance was public or expected to be viewed by an identifiable audience (Fazio et al., 1981; Kelly & Rodriguez, 2006; Schlenker et al., 1994; Tice,

1992). Computer-mediated communication extended this logic. Selective descriptions in blogs and other mediated settings can shift self-perceptions, and audience feedback can magnify or redirect the change (Carr & Foreman, 2016; Carr et al., 2021; Gonzales & Hancock, 2008; Walther et al., 2011). Embodiment research similarly suggests that clothing and avatars can alter immediate thought and behavior, although the proposed mechanisms and replicability of specific effects remain contested (Burns et al., 2019; Horton et al., 2025; Ratan et al., 2020; Walther & Lew, 2022; Yee & Bailenson, 2007).

Personality science supplies a second half of the explanation. Traits are not only static labels; they can be represented as distributions of states expressed across situations (Fleeson, 2001; Fleeson & Jayawickreme, 2015). Repeated situations, expectancies, states, reactions, and reflection may accumulate into trait development (Wrzus & Roberts, 2017). Volitional personality research further indicates that wanting to change is weak by itself, whereas repeated behavioral goal attainment and structured intervention can produce modest, sometimes persistent trait change (Haehner et al., 2024; Hudson et al., 2019; Stieger et al., 2021). The unresolved theoretical opportunity is to connect the outward act of presenting an aspirational trait to the repeated state processes required for personality development.

The Audience-Weighted Self-Evidence (AWSE) model makes that connection. It treats self-presentation as a potential source of two kinds of evidence. Enacted evidence is information actors infer from what they voluntarily say, do, wear, create, or accept responsibility for. Reflected evidence is information actors infer from audiences' perceived appraisals. Neither kind of evidence is accepted at face value. Actors discount performances they regard as coerced, implausible, or strategically cheap, and they weight audience responses by source credibility and diagnostic access. Separately, audiences can restructure opportunities, expectations, and

sanctions without changing private self-belief. If identity updating and the altered environment repeatedly elicit trait-relevant states, and if those states are reflected upon and generalized, a presentation may become part of personality development.

AWSE is not a claim that people can simply announce a desired identity and become it. In fact, its signature prediction is a crossover. Social recognition of a low-entailment symbol, such as a badge, label, or broad declaration, can create premature closure and reduce later striving. Recognition of a high-entailment commitment, whose credibility depends on future observable conduct, should increase later enactment. Publicness is therefore not assumed to exert a uniformly positive effect. Its consequence depends on whether the presentation merely represents an identity or structurally requires the actor to practice it.

AWSE is an integrative extension of Identity Shift Theory, not a wholesale replacement for it. Identity Shift Theory explains how selective self-presentation, public commitment, and biased scanning can alter self-perception in mediated communication (Carr et al., 2021). AWSE adds a source-differentiated account of audience evidence, separates audience information from audience-created opportunity, links identity updating to repeated trait states, and specifies behavioral entailment as a sign-switching moderator of recognition. Its novel burden is therefore not to show again that selective presentation can shift a self-rating; it is to predict which evidence sources will be incorporated and when a short-lived identity shift will consolidate, reverse, or remain context-bound.

The contribution of the model is fourfold. First, it integrates public commitment, identity shift, self-perception, reflected appraisal, and personality development without treating their mechanisms as interchangeable. Second, it formalizes differential weighting of audience evidence rather than equating all approval, visibility, or feedback. Third, it specifies behavioral

entailment as a sign-switching moderator of recognition. Fourth, it sets a high evidentiary threshold for personality change: delayed, private, cross-context, multi-method change that survives removal of the original audience and presentation.

Conceptual Boundaries

Self-presentation refers here to intentional, attributable communication of self-relevant information directed to an actual or anticipated audience (Leary & Kowalski, 1990).

Presentations include verbal claims, autobiographical stories, profile content, clothing and stylistic cues, public commitments, and acceptance of visible roles. The audience can be physically present or technologically mediated. It can also be anticipated when the actor prepares content for later reception. If no message reaches another person, the episode can still supply enacted evidence and a simulated appraisal, but it cannot supply an actual audience response or an audience-created opportunity.

Aspirational self-presentation is the intentional portrayal of an attribute that is not yet typical but is experienced as desirable and potentially attainable. Current accuracy, aspiration, ownership, and willingness to enact are separate dimensions. A person can genuinely aspire to an attribute while exaggerating current standing. Conversely, an accurate statement can be low in ownership if it is scripted or compelled. An assigned representation, such as an experimenter-selected avatar, can later become appropriated as self-relevant. These distinctions matter because the model concerns how much diagnostic credit an actor assigns to a performance, not whether every aspirational claim is literally accurate.

Enacted evidence is the actor's inference from producing a presentation or identity-relevant behavior. Reflected evidence is the actor's inference from a perceived audience appraisal. The latter must be distinguished from the audience's actual appraisal and from the

actor's accuracy in perceiving it. Social affordance refers to an opportunity, expectation, sanction, or pattern of treatment created by an audience or institution. It is a structural input to behavior, not evidence by definition. The same event can supply reflected evidence and alter affordances, but the routes are not interchangeable.

Self-concept change occurs when an accessible self-perception shifts on a relevant dimension. Identity change is broader and concerns how an actor defines, values, or narrates a self-relevant role or attribute. Personality change requires alteration in relatively enduring patterns of thought, feeling, motivation, or behavior, with breadth appropriate to the construct claimed. A post-presentation self-rating can be an early component of such change, but it is not sufficient evidence. Role-specific development should not be redescribed as global trait change merely because both are psychologically meaningful.

Behavioral entailment is the proposed claim-behavior contingency. Structural entailment is a design-imposed or socially instituted dependency between future conduct and the continuing credibility of a presentation. Perceived entailment is the actor's appraisal that future conduct is necessary to keep the claim credible, and it is the proximal moderator. Adding "writer" to a profile is low in entailment if no future conduct bears on the label. Publicly accepting responsibility to deliver a weekly draft is high in entailment if nonperformance would directly undercut the claim. Entailment is not identical to immediate effort, cost, specificity, planning, commitment strength, or observability. A costly symbol can require no later action; a specific private plan can require action without creating a public identity claim. Empirical work must validate entailment's discriminant structure rather than declaring neighboring constructs equivalent after the fact.

Table 1 summarizes the construct boundaries and the evidence that is insufficient for each level of inference.

Table 1

Construct Boundaries and Evidentiary Thresholds

Construct	Working definition	What is not sufficient
Self-presentation	Intentional, attributable communication of self-relevant information to an actual or anticipated audience.	Behavior with no communicative function. If no message reaches another person, the episode cannot supply an actual appraisal or social affordance.
Aspirational self-presentation	Portrayal of a desired, potentially attainable attribute that is not yet typical. Accuracy, aspiration, ownership, and willingness to enact are separate dimensions.	Accurate expression of an established attribute, or an unowned claim made without corresponding aspiration or willingness to enact it.
Enacted evidence	Diagnostic credit the actor assigns to producing a presentation or identity-relevant behavior.	Performance alone, especially when fully scripted, coerced, strategically cheap, or readily explained by external incentives.
Reflected evidence	The actor's inference from a perceived audience appraisal, weighted by source quality and discounted for alternative explanations.	The audience's actual belief, response count, or generic approval considered without the actor's perception and weighting.
Social affordance	An opportunity, expectation, sanction, or pattern of treatment created by an audience or institution.	Self-evidence or identity updating. Affordances can change behavior while private self-belief remains unchanged.
Self-concept change	A shift in the actor's accessible self-perception on a relevant dimension.	A change in what an audience believes, a manipulation check, or one strategically convenient response.
Identity change	Reorganization in how the actor defines, values, or narrates a self-relevant role or attribute.	A fleeting semantic association or momentary cue effect.
Personality change	Persistent alteration in typical thought, feeling, motivation, or behavior, with breadth appropriate to the construct claimed.	One immediate self-rating, one public performance, or behavior maintained only by the initiating context or audience.
Behavioral entailment	A design-imposed or socially instituted contingency between later conduct and the continuing credibility of a presentation; perceived entailment is the actor's appraisal of that contingency.	Effort, specificity, planning, accountability, commitment strength, or observability considered alone.

Note. The evidentiary threshold rises from momentary self-description to personality change. Current accuracy and genuine aspiration can coexist with exaggeration; they should not be treated as opposite categories.

Review Approach and Evidentiary Standard

The present article is a theory-building, integrative review rather than a systematic review or meta-analysis. Literature coverage was updated through August 2026 using concept-driven searches of scholarly databases, reference chaining, and publisher archives. Priority was given to primary experiments, longitudinal studies, preregistered replications, multi-method field studies, and current quantitative or systematic syntheses where an evidence stream was contested. The review spans self-presentation, computer-mediated identity shift, social feedback, self-signaling and symbolic completion, embodiment, and volitional personality development. The search was not designed to estimate a pooled effect or guarantee exhaustive retrieval. Its purpose was to identify converging mechanisms, contradictions that a useful theory must explain, and observations capable of discriminating among rival accounts.

Three standards guide interpretation. First, self-report change is separated from behavioral and informant evidence. Second, immediate change is separated from delayed and cross-context persistence. Third, a demonstrated outcome is separated from its proposed mechanism. For example, avatar appearance may alter behavior without self-perception mediating the effect, and public feedback may alter a self-rating through anticipated evaluation rather than internalized belief. Where evidence cannot identify a mechanism or support a durable effect, the limitation is treated as part of the theory rather than as noise around it.

When Presentation Acts Back on the Presenter

Public Performance as Self-Evidence

Self-perception theory proposes that people sometimes infer their attitudes and attributes from observing their own behavior, especially when internal cues are weak or ambiguous (Bem,

1972). Strategic self-presentation complicates this logic because the actor knows that the behavior was selected to create an impression. That knowledge supplies a reason to discount the performance: “I acted sociable because the situation required it.” Nevertheless, early experiments repeatedly found carryover from performed identities.

Fazio et al. (1981) used biased interview questions to elicit introverted or extraverted self-presentations and observed corresponding shifts in later self-perception and behavior. Jones et al. (1981) found that strategic self-presentation could change subsequent self-esteem. Tice (1992) showed across three studies that identical behavior had a stronger effect on self-concept when enacted publicly than privately, with evidence extending beyond immediate self-description. Schlenker et al. (1994) reported that a sociable self-presentation affected later self-appraisal, autobiographical recall, and behavior in a new situation with a new audience; private reflection alone did not reproduce the effect. Kelly and Rodriguez (2006) similarly found that a presentation expected to be viewed by an identifiable audience exerted more influence than an otherwise comparable presentation.

These findings establish more than a simple consistency effect. Public performance can alter which autobiographical examples come to mind, how an actor behaves with a different person, and how that behavior is perceived. Yet they do not establish enduring trait change. Samples were generally small by contemporary standards, outcomes were usually assessed over short intervals, and participants often knew that the performance was part of an experiment. The public-private manipulation also bundles several processes: anticipated evaluation, identifiability, reputational stakes, audience simulation, and a subjective sense that the act “counts.” AWSE treats these as separable inputs rather than as one undifferentiated publicness effect.

The model also rejects the assumption that the actor's behavior has a fixed evidential value. A voluntary, autobiographically generated, effortful performance should carry more enacted evidence than a fully scripted or coerced one. A performance that is diagnostic of a trait should carry more weight than a symbolic cue with many alternative explanations. Strategic discounting should increase when the actor regards the presentation as a transparent tactic, when the desired identity is implausibly distant, or when external incentives provide a complete explanation for the conduct. Thus, identical outward content can generate different internal change depending on ownership and attribution.

Identity Shift and the Weight of an Audience

Computer-mediated communication makes self-presentation unusually visible and editable. Actors can select language, images, profiles, and audiences while leaving a persistent or ephemeral record. Gonzales and Hancock (2008) found that computer-mediated self-presentation could shift self-perceptions in the direction of the presented identity. Walther et al. (2011) reported feedback-related identity-shift effects in both a public blog condition and a private expressive condition, indicating that feedback and publicness should be treated as separable influences. Carr and Foreman (2016) further demonstrated that the publicness and relational context of feedback affect subsequent self-perception. Choi et al. (2020) produced a theoretically important counterintuitive result: an ephemeral rather than permanent platform supported internalization under some conditions, apparently because ephemerality reduced strategic impression-management concern and enabled self-symbolizing.

The broad conclusion is not that more visible, permanent, or approving communication always produces greater self-change. Platform affordances alter what the presentation means to the actor. A permanent post may increase accountability, or it may heighten strategic discounting

because the actor experiences the content as curation. An ephemeral post may reduce reputational binding, or it may feel more spontaneous and therefore more diagnostic. Feedback may be affirming but epistemically weak, as with indiscriminate likes, or it may be diagnostically strong because it comes from an informed observer who saw relevant behavior.

This distinction motivates audience weighting. AWSE proposes that reflected evidence is a function not merely of valence or quantity but of five candidate judgments. Diagnostic access concerns whether the audience observed behavior relevant to the claimed trait. Independence concerns whether the response was freely generated rather than scripted, reciprocal, or socially obligated. Epistemic credibility concerns whether the source is perceived as knowledgeable and accurate. Relational relevance concerns whether its view matters for the identity in question. Informational nonredundancy concerns whether the source adds evidence not already supplied by correlated observers. These weights can diverge. A close friend may be highly relevant but positively biased; an expert may be knowledgeable but relationally distant; a large anonymous audience may be numerous yet weakly informed. These dimensions and their composition are theoretical candidates, not an established scale.

Recent feedback research supports a graded rather than mechanical account of incorporation. Across four studies, larger feedback-self discrepancies predicted greater intended or observed self-concept change, and negative feedback produced more change than positive feedback; the authors situated these findings alongside earlier evidence that extreme, implausible discrepancies may be discounted (Brotzeller & Gollwitzer, 2025). Elder et al. (2022) likewise showed that coherence and positivity motives constrain learning from self-relevant social feedback. Brotzeller et al. (2026) found across six studies that reflection was associated with self-concept change and experimentally increasing reflection produced more change than

suppressing it. These results do not imply that reflection always assimilates the self to feedback. Reflection can also support immunization, reinterpretation, or behavioral compensation. They do indicate that the route from reflected appraisal to self-change depends on active processing rather than passive receipt.

Clothing, Avatars, and the Difference Between a Cue and a Claim

Physical presentation is especially compelling because it seems to let a person step into a role. The original enclothed cognition experiments argued that clothing affects the wearer through both symbolic meaning and the experience of wearing it (Adam & Galinsky, 2012). A preregistered direct replication did not reproduce a focal Stroop effect (Burns et al., 2019). A later z-curve and meta-analytic synthesis of 105 effects from 40 studies raised concerns about early studies while reporting evidential value in post-2015 work (Horton et al., 2025). The defensible conclusion is therefore narrower than the popular claim that dressing like a role makes one become it. Attire can influence immediate cognition, affect, and behavior under some symbolic and contextual conditions; durable identity or personality change remains largely untested.

Avatar research shows a related pattern. The Proteus effect describes behavioral conformity to characteristics associated with an embodied representation (Yee & Bailenson, 2007). A meta-analysis of 46 experimental studies found small-to-medium effects on behavior and attitudes (Ratan et al., 2020). A later meta-analysis of 56 experiments obtained a similar weighted estimate ($r = .23$) and found larger effects in immersive virtual reality than in flat-screen studies (Beyea et al., 2022). Yet the mechanism remains ambiguous. In classic self-perception theory, behavior is observed and used to infer an attribute. In many avatar studies, appearance activates a stereotype and behavior follows. That sequence is also compatible with

priming, expectancy, or avatar self-relevance accounts. Many studies measure behavior but not self-concept, use scripted interactions, and do not test persistence outside the virtual setting (Walther & Lew, 2022).

AWSE treats attire and avatars as presentations whose evidential status depends on appropriation and entailment. An assigned laboratory coat may provide a situational cue without becoming self-evidence. A chosen garment that communicates a valued role to a meaningful audience may be more self-relevant, but it remains low in behavioral entailment unless it opens, requires, or organizes future conduct. A professional uniform can become consequential when it changes how others allocate trust, responsibility, and opportunity. In that case, the garment is not a magical cause of trait change. It is an entry point into a social and behavioral sequence.

What the Existing Evidence Does Not Yet Establish

The literature supports three modest claims. First, people can internalize aspects of a strategically presented self. Second, publicness and feedback can strengthen, weaken, or redirect that process. Third, representations such as clothing and avatars can alter immediate behavior under some conditions. It does not yet support a general claim that self-presentation changes personality.

Four gaps are decisive. Duration is rarely sufficient to distinguish an accessible self-description from a stable change. Generalization is often unmeasured; behavior with the original audience or in the original medium may reflect accountability rather than personality. Method variance is substantial because many studies use a presentation manipulation and an immediately adjacent self-report outcome. Finally, alternative mechanisms are seldom pitted against one another. Priming, consistency, demand, anticipated evaluation, dissonance, self-perception, and

reflected appraisal can predict similar immediate outcomes while diverging on delayed private behavior.

Table 2 states the narrow inference supported by each evidence stream and the unresolved issue that prevents a stronger conclusion.

Table 2

Evidence Streams and the Inferences They Support

Evidence stream	Warranted inference	Decisive unresolved issue
Public self-presentation	Attributable performance can shift immediate self-appraisal, recall, and near-term behavior.	Publicness bundles identifiability, anticipated evaluation, commitment, and demand; durable change is rarely tested.
Digital identity shift	Selective presentation and audience feedback can alter immediate self-perception.	Platform persistence, strategicity, audience response, and claim identification interact; cross-context persistence remains uncertain.
Social feedback	Feedback discrepancy, source credibility, and reflection can affect self-concept updating.	Actual appraisal, perceived appraisal, source weighting, and audience-created opportunity are seldom separated.
Public recognition and symbolic completion	Recognition can reduce further striving, while attributable commitments can increase consistency.	No existing study cleanly tests a strict recognition crossover under matched exposure and practice.
Clothing and avatars	Self-representations can alter immediate cognition or behavior under some conditions.	Cue activation, stereotype priming, ownership, self-relevance, and durable identity change are not adequately separated.
State and trait dynamics	Repeated state patterns and person-situation transactions can predict later personality development.	Most evidence is observational and does not begin with aspirational self-presentation.
Volitional personality interventions	Structured behavioral practice can produce modest self- and observer-rated trait change.	Presentation, source-weighted appraisal, social affordance, and repeated enactment have not been dismantled experimentally.

Note. Each row states the narrowest inference supported by the evidence stream. No stream directly establishes the full AWSE cascade from presentation to persistent, cross-context personality change.

From Identity Updating to Personality Development

Personality traits describe relatively enduring regularities, but those regularities are expressed through variable states. Whole Trait Theory distinguishes descriptive trait distributions from explanatory social-cognitive mechanisms and treats within-person state variation as central rather than incidental (Fleeson & Jayawickreme, 2015). People may act more extraverted in one

setting and less in another while retaining a characteristic distribution of extraverted states. Personality development occurs when the distribution or its generating processes change over time.

This state-based view supplies the first bridge from presentation to personality. A presented identity can change the probability of trait-relevant states. A person who publicly assumes the role of organizer may be asked to schedule, remind, coordinate, and follow through. Those acts are not merely evidence that the person is conscientious; they are conscientious states. If such states recur, are reinforced, and spread beyond the original role, they can contribute to a shift in the person's typical pattern.

The TESSERA framework provides a more detailed temporal sequence: triggering situations activate expectancies, states and state expressions follow, reactions occur, and repeated sequences may be integrated through reflection (Wrzus & Roberts, 2017). A two-year diary study found that momentary states predicted longer-term changes in several self-rated traits and implicitly measured extraversion, while support for reflection varied by trait (Quintus et al., 2021). These findings are consistent with accumulation, but they also show that not every state or reflection episode becomes development.

Volitional personality research imposes a second constraint. A meta-analysis of 207 intervention studies found that personality-trait measures can change during intervention, especially for emotional stability, but the underlying literatures were heterogeneous and usually did not isolate self-presentation (Roberts et al., 2017). Hudson and Fraley (2015) found that people can pursue desired trait change, but change goals are not self-executing. Hudson et al. (2019) showed that attaining behavioral change goals, rather than merely holding them, predicted volitional trait change. A digital coaching intervention produced self- and observer-

reported trait changes over several months (Stieger et al., 2021). A preregistered systematic review of 30 longitudinal studies ($N = 7,719$) found weak goal-change associations and a small mean intervention effect ($d = .22$; seven studies); an exploratory follow-up synthesis ($d = .37$) rested on four studies, and the evidence base remained demographically narrow (Haehner et al., 2024).

These findings locate the likely contribution of self-presentation. Presentation is not a substitute for behavioral practice; it can be a device that changes which practices occur, how they are interpreted, and how the social environment supports them. It can make a desired trait more accessible, create opportunities, recruit feedback, and impose a reputational structure. But if the presentation produces recognition without practice, or practice without generalization, the developmental cascade stops.

Social expectations are plausible but not sufficient by themselves. The social investment principle proposes that investment in adult roles and their expectations promotes trait development. Yet a direct longitudinal test reported limited evidence that social expectations explained adult personality change (Hang et al., 2023). AWSE therefore does not assume that occupying a role automatically changes the occupant. It specifies intermediate conditions: ownership of the presentation, audience credibility, access to diagnostic behavior, repeated enactment, reflection, and cross-context transfer.

Qualitative work on professional transitions illustrates the full sequence more vividly. Ibarra (1999) described how people adapting to new professional roles experiment with “provisional selves,” observe reactions, evaluate fit, and revise their performances. This work does not establish trait change, but it reveals identity development as iterative experimentation

rather than a single declaration. AWSE translates that insight into measurable evidence routes and makes the conditions for consolidation explicit.

The Audience-Weighted Self-Evidence Model

AWSE is a recursive theory of self-evidentiary credit assignment. At a presentation episode, an actor does not merely ask whether a claim or performance occurred. The actor implicitly decides how much that episode says about the self, how much it says about strategy or circumstance, and how much an audience was in a position to know. The phrase self-evidentiary credit assignment refers to this allocation of diagnostic credit across the person, the situation, the presentation motive, and the audience. It is the model's central explanatory operation.

The process begins with an attributable, identity-relevant presentation. The actor can draw enacted evidence from producing it and reflected evidence from perceived audience appraisals. Route-specific discounting reduces either source when the performance is coerced, the appraisal is uninformed, or another explanation is sufficient. In parallel, the presentation can alter opportunities, expectations, sanctions, and treatment. Those social affordances may change behavior even when the actor's self-belief does not change. Later behavior supplies new enacted evidence and elicits new appraisals, creating a feedback loop. Only repeated, sufficiently owned enactment that persists and transfers can support a claim of personality development.

This architecture separates three questions that adjacent theories often combine. Did the presentation change the actor's accessible self-belief? Did it change what the social environment allowed or required the actor to do? Did any resulting behavior persist outside the original audience and context? The first concerns identity updating, the second concerns social structure, and the third concerns consolidation. A study can answer one question without answering the others.

Evidence and Affordance as Distinct Routes

Evidence and affordance are analytically independent. Credible feedback can change a self-view without opening any opportunity to act. A role assignment can create repeated practice while the actor privately rejects the role. When both routes are weak, the presentation is likely to dissipate. When both are strong, identity updating and opportunity can reinforce one another. Table 3 identifies the four process profiles. They are not fixed person types. The same person can occupy different cells across attributes, audiences, and episodes.

Table 3

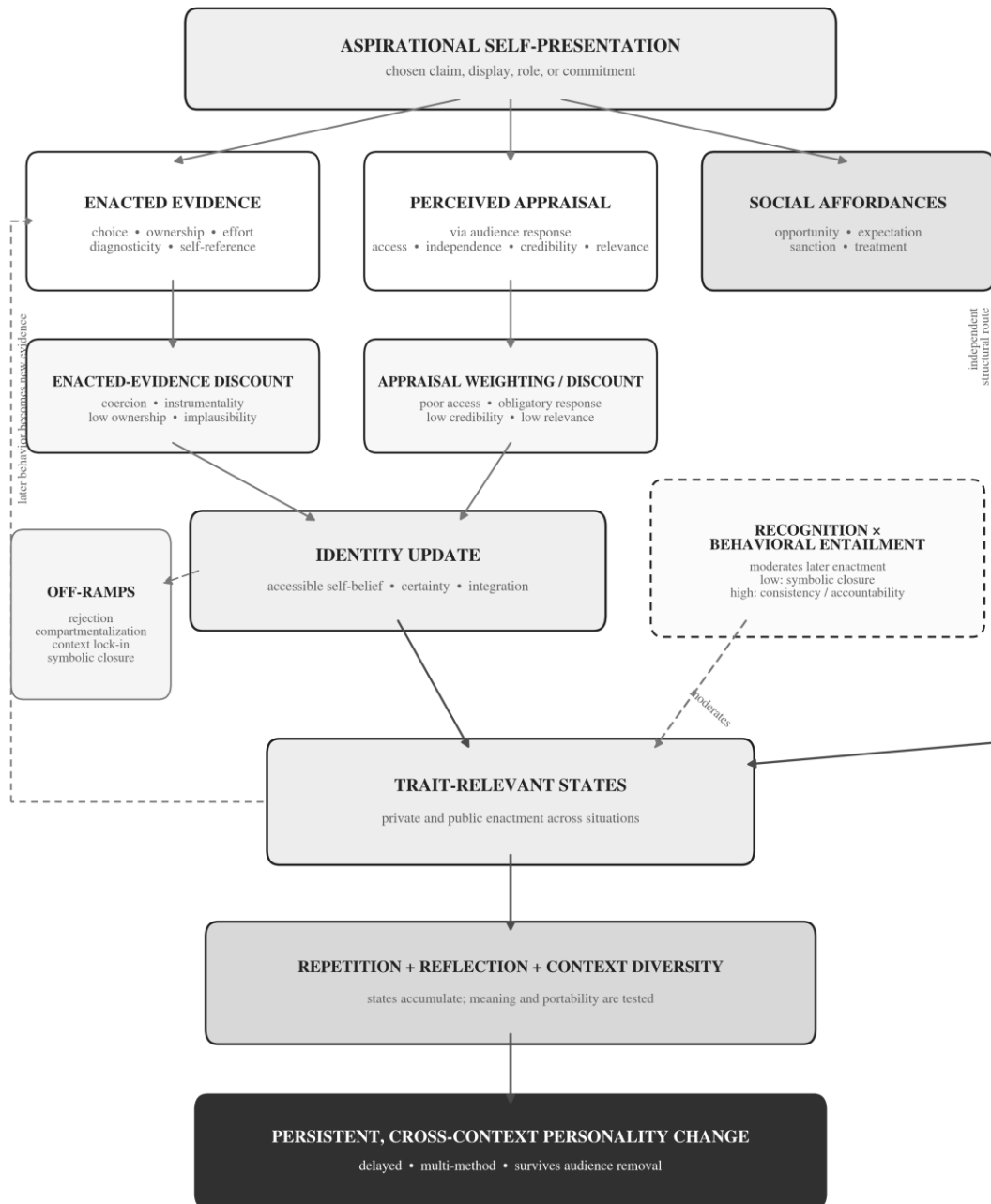
Process Profiles Created by Self-Evidence and Social Affordance

Diagnostic self-evidence	Social affordance	Process profile	Expected near-term pattern
Low	Low	Dissipation or symbolic closure	Little self-belief change and little trait-relevant practice; any effect should be brief or cue-bound.
High	Low	Narrative identity update	Accessible self-belief changes, but enactment remains fragile unless the actor independently creates opportunities to practice.
Low	High	Role compliance, skill, or habit	Opportunity or sanction changes behavior without a corresponding self-belief update; transfer is likely to remain context-bound.
High	High	Recursive reinforcement	Identity updating and opportunity jointly increase trait-relevant states, creating the strongest pathway to consolidation if ownership, repetition, and transfer are sustained.

Note. Self-evidence includes enacted and reflected evidence after route-specific discounting. Social affordance is a separate structural route. These profiles describe episodes, not fixed person types, and none guarantees beneficial or global personality change.

The high-evidence, high-affordance profile is the model's strongest developmental pathway, but it is not a guarantee of desirable or global change. Practice can remain context-bound, the identity can later be rejected, and an externally valued role can impose psychological cost. Conversely, a low-evidence, high-affordance profile may produce real skill or habit before any explicit identity update. AWSE therefore treats temporal order as an empirical question and does not require conscious self-inference at every episode.

Figure 1 depicts the three routes, their points of convergence, the proposed crossover, and the model's principal off-ramps.

Figure 1*The Audience-Weighted Self-Evidence Model*

Note. Solid arrows represent proposed causal paths; dashed elements denote moderation, recursion, or off-ramps. The social-affordance route reaches trait-relevant states independently of identity updating. The figure depicts hypotheses only, these effects are not yet established.

Enacted Evidence: What the Actor Learns From Doing

Enacted evidence begins with a simple observation: an actor has performed something. Whether that performance informs the self depends on its attribution. Five properties should increase its weight. Choice indicates that the actor could have acted otherwise. Ownership indicates that the words or conduct are experienced as self-authored rather than supplied by a script. Diagnosticity indicates that the conduct discriminates the claimed attribute from plausible alternatives. Effort and cost indicate that the performance was not merely cheap talk. Autobiographical integration indicates that the actor can connect the episode to other self-relevant memories and goals.

These properties are related but not redundant. A scripted presentation can be effortful yet carry little ownership. A voluntary statement can be owned but nondiagnostic because it costs nothing and is easy for anyone to make. A difficult action can be diagnostic but attributed to external pressure. AWSE predicts the greatest internalization when the actor experiences the conduct as freely chosen, self-authored, identity-relevant, and difficult to explain without reference to the claimed attribute.

Prior self-belief constrains the plausible range of updating. Schlenker and Trudeau (1990) found that the effect of self-presentation on private self-beliefs depended on the strength of prior beliefs and whether the portrayal fell within a latitude of acceptance. The implication is not that only familiar performances matter. A perfectly typical performance adds little new information. Rather, moderate discrepancy may maximize learning: the performance is novel enough to be informative but close enough to be self-attributed. Very large discrepancy increases the probability of discounting, role separation, or reactance. This produces an inverted-U prediction between aspirational distance and positive identity updating. The full curve is an AWSE

extrapolation from prior-belief and discrepancy findings; it has not been established for aspirational self-presentation.

The enacted route is compatible with biased scanning but is not reducible to it. A presentation prompt may cause people to retrieve trait-congruent memories, and those memories may drive the self-rating (Fazio et al., 1981). AWSE treats retrieval as one way an episode is integrated into self-knowledge. It predicts, however, that behavioral ownership and diagnosticity should affect later private action even when memory availability is controlled. If the entire effect disappears after balancing autobiographical retrieval, the enacted-evidence component has been overstated.

Reflected Evidence: What the Audience Appears to Know

People do not receive social responses as raw facts. They infer what an audience observed, why it responded, and whether its judgment deserves weight. Reflected appraisal research suggests that people possess some genuine insight into how others see them, but that insight also contains projection and error (Carlson et al., 2011). AWSE therefore separates an audience's actual appraisal, the actor's perceived appraisal, and the weight the actor assigns to it. The proximal cause of identity updating is the perceived, weighted appraisal; the accuracy of that appraisal determines whether the resulting identity is calibrated to social reality.

Diagnostic access is foundational. Praise from a person who observed sustained performance should be more informative than praise based on a profile label. Independence matters because obligatory, reciprocal, or algorithmically generated responses are weak evidence. Epistemic credibility concerns whether the audience is competent to judge the attribute. Relational relevance concerns whether its view matters in the relevant identity domain. A coach may be credible about discipline in training but irrelevant to kindness at home. A

partner may have excellent access to everyday reliability but limited access to professional leadership. The same audience can therefore receive different weights for different claims.

Audience-created opportunities can change behavior without first changing self-belief. Other people may allocate responsibilities, impose expectations, or interact in ways that elicit claim-consistent behavior. Behavioral-confirmation research shows that perceivers can draw corresponding conduct from targets, sometimes with carryover to a later interaction (Snyder & Swann, 1978). The PERSOC framework likewise treats personality and relationships as mutually influential perception-behavior transactions (Back et al., 2011). AWSE treats this environmental route as distinct from reflected evidence: a presentation can change what others let, ask, or require the actor to do even when their private appraisal is unknown. The resulting conduct may later become enacted evidence.

The distinction matters experimentally. A positive message from an informed observer can update self-belief without creating a new opportunity. An assigned role can create opportunities without conveying approval. If both are manipulated, AWSE predicts partially independent effects and a stronger recursive cycle when they converge. When they conflict, the sequence should depend on their weights. An actor may privately reject flattering feedback yet discover through new responsibilities that the behavior is attainable; conversely, the actor may accept a label but receive no opportunity to practice it.

Strategic Discounting as a Gate, Not an Error Term

Impression management gives actors privileged knowledge of their own motives. A person who presents warmth to secure a job can explain the performance as instrumental. That explanation does not make internalization impossible, but it supplies a discount. Strategic discounting should be strongest when an external incentive fully explains the act, when the

portrayal is dictated, when the claim is implausible, or when the symbol is unconnected to behavior. It should weaken when actors choose the manner of presentation, draw on autobiographical material, accept consequences, or continue the conduct after the external incentive ends.

This gate helps reconcile results that appear to disagree about publicness. Public behavior may feel more diagnostic because it is consequential and attributable, as in Tice's (1992) studies. Yet a highly curated permanent post can feel less diagnostic because it is transparently strategic, whereas an ephemeral expression can feel more spontaneous and self-revealing (Choi et al., 2020). Nominal visibility is therefore an imperfect proxy. Perceived attributable consequence and perceived strategicity should be measured as mediators, not assumed from a platform setting.

The gate also creates non-assimilation outcomes. If a portrayal is only slightly discrepant and plausibly self-authored, it may be incorporated. If it is desirable but difficult, credible rejection may motivate compensatory practice. If it is unattainable or humiliating, the person may compartmentalize the episode, reject the audience, or move the self-view away from the claim. Ross and DeAndrea (2025) found that claim identification, rather than nominal identifiability or perceived publicness alone, consistently moderated identity-shift outcomes. This is consistent with AWSE's emphasis on ownership and plausibility and warns against treating exposure as an automatic cause.

Behavioral Entailment and the Recognition Paradox

Public presentation has two theoretically opposed consequences. Commitment research predicts that an attributable claim increases consistency pressure and future enactment. Symbolic self-completion predicts that public recognition can provide a sense of possessing the identity and thereby reduce effort. Evidence exists for both. Publicly noticed identity intentions have

sometimes produced less intensive goal pursuit than unnoticed intentions (Gollwitzer et al., 2009). Social recognition of a prosocial self-symbol similarly reduced subsequent moral striving in two experiments (Susewind & Walkowitz, 2020). Yet public, attributable presentations have also increased later claim-consistent behavior (Schlenker et al., 1994; Tice, 1992). These demonstrations establish plausible countervailing mechanisms rather than a general law: the evidence comes from a small number of domain-specific paradigms, and the Recognition x Entailment crossover remains an untested AWSE prediction.

AWSE proposes perceived behavioral entailment as a sign-switching moderator. A low-entailment presentation can be socially recognized without making future conduct relevant to the claim's continuing credibility. Recognition then supplies symbolic completion: the audience has acknowledged the identity, and further effort produces diminishing identity value. A high-entailment presentation is structurally tied to future conduct, and the actor must perceive that dependency. Recognition then increases the stakes of enactment because failure would become diagnostic to the self and audience. Initial believability of the identity claim is distinct from continuing credibility: studies should match whether the original statement seems plausible while varying whether later conduct is required to sustain it.

Consider two presentations of a writing identity. In a low-entailment condition, a person adds "writer" to a profile and receives approval. In a high-entailment condition, the person publicly accepts responsibility for submitting a draft to an identified group every Friday. Both are identity claims and both can be recognized. Only the second makes future action necessary to preserve the claim. The critical prediction is a strict Recognition x Entailment crossover: recognition should reduce later private identity-congruent effort after low-entailment symbols but increase it after high-entailment commitments.

Entailment must be distinguished from surveillance. If conduct occurs only while the audience can observe it, the result may be reputational maintenance. AWSE predicts a mediated sequence in which recognition of an entailing claim first changes actual fulfillment, successful fulfillment becomes enacted evidence, and identity updating then predicts transfer after surveillance and reputational consequences are removed. A delayed anonymous task, unavailable to the original audience and different from the entailed obligation, is therefore more diagnostic than continued compliance in public.

Recursion, Consolidation, and Cross-Context Generalization

The model becomes developmental when evidence episodes recur. A presentation can increase identity accessibility, alter choices, and change audience treatment. The next trait-relevant state then supplies new enacted evidence and may generate a new appraisal. Reflection links episodes into a narrative: “I said I could do this, they trusted me with it, I did it, and I continued when nobody was watching.” Repetition in stable settings can also form cue-response associations that require less deliberate identity regulation over time (Wood & R  nger, 2016). Reflection and habit are complementary: reflection can generalize the meaning of behavior, while habit can stabilize its execution.

Consolidation should proceed through distinguishable levels. The earliest effect is a change in an accessible self-description. The next is a shift in the frequency or intensity of trait-relevant states within the original setting. Repeated practice can produce a role-specific behavioral pattern. Broader personality change requires transfer to new audiences and situations, a shift in the person’s state distribution, and persistence after the initiating presentation is removed. Facet-level changes should generally precede domain-level changes because

presentations usually entail specific behavior. Weekly drafting may first change industriousness or self-discipline, not the entire conscientiousness domain.

Cross-context diversity should be more valuable than simple repetition. Twenty performances for the same approving audience may build a local role or a negotiated reputation. Similar performances with independent, credible audiences in work, friendship, and private settings provide less redundant evidence and reduce the probability that the identity is merely context-bound. At the same time, breadth without sufficient repetition may prevent habit formation. AWSE therefore predicts an interaction: durable change is most likely when enactment is repeated enough to stabilize behavior and varied enough to demonstrate portability.

The process can also reverse. Diagnostic failure, audience rejection, or repeated counter-trait behavior can weaken the updated identity. An initially useful presentation may become controlled and exhausting, producing disengagement. Conflicting audiences can maintain different selves, and power asymmetries can force a performance that never becomes owned. The model therefore describes pathways to assimilation, compensation, compartmentalization, symbolic closure, and rejection rather than assuming a single positive endpoint.

A Bounded Process Specification

The formalization is a disciplined restatement of the theory, not an identified measurement model or a claim that actors update as ideal Bayesians. Let z_{ikt} denote person i 's latent accessible self-belief about attribute k at episode t . Let e_{ikt} be a signed enacted signal and f_{ijkt} a signed perceived appraisal signal from audience source j , each expressed on the same conceptual scale as z . Updating depends on prediction error, not on the raw positivity of a signal. A favorable appraisal that merely repeats a strong prior belief should produce less change than equally favorable information that is genuinely surprising.

For reflected evidence, $q_{ijkt} = \phi(A, D, C, L, N; \theta)$ is a bounded source-strength function of diagnostic access, response independence, epistemic credibility, relational relevance, and informational nonredundancy. Compensatory, multiplicative, and minimum or veto rules are competing empirical specifications; existing evidence identifies none of them. The five dimensions are theoretical candidates rather than a validated scale, so their discriminant validity and incremental prediction must also be established. Enacted and reflected signals have separate discount terms, d_{ikt}^E and d_{ijkt}^R , bounded from 0 to 1. Coercion, transparent instrumentality, or low ownership can discount enacted evidence; flattery, poor access, low credibility, or an irrelevant relationship can discount reflected evidence. Route learning rates, λ_{ikt}^E and λ_{ikt}^R , are bounded from 0 to 1 and can vary with prior uncertainty, aspirational discrepancy, and attribute importance. For compact notation in Equation 1, $w_{ijkt}^R = q_{ijkt}(1 - d_{ijkt}^R)$ denotes the effective reflected-source weight. The immediate update is represented as follows:

$$\begin{aligned} \Delta z_{ikt} = & \lambda_{ikt}^E (1 - d_{ikt}^E) (e_{ikt} - z_{ikt}) \\ & + \lambda_{ikt}^R h \left\{ \sum_{j=1}^J w_{ijkt}^R (f_{ijkt} - z_{ikt}) \right\} \\ & + \varepsilon_{ikt}. \end{aligned} \quad (1)$$

The function h is odd, monotone, and saturating. One convenient choice is $h(x) = \text{ctanh}(x/c)$, where c is a positive scaling constant. Saturation limits the effect of sheer response volume. Because equally weighted positive and negative sources can cancel in the signed mean, analyses should also retain audience dispersion or polarization. Cancellation predicts little directional updating; polarization can separately predict uncertainty or context differentiation. The residual has expected value zero conditional on the modeled inputs. All quantities must be defined and measured before outcomes are inspected.

Identity updating is only one cause of later conduct. For a binary trait-relevant state in context c , a schematic transition model is:

$$\begin{aligned}
\text{logitPr}(S_{ikc,t+1} = 1) = & \alpha_i + \gamma_c \\
& + \beta_T T_{ikt} + \beta_I z_{ik,t+1} + \beta_O O_{ikc,t+1} \\
& + \beta_B B_{ikt} + \beta_R R_{ikt} + \beta_{RB} (R_{ikt} B_{ikt}) \\
& + \beta_S S_{ikc,t}.
\end{aligned} \tag{2}$$

Here, α_i represents stable person differences, γ_c represents context, T_{ikt} summarizes the pre-episode distribution of relevant states, $O_{ikc,t+1}$ is an audience-created opportunity or sanction, B_{ikt} is perceived behavioral entailment, R_{ikt} is social recognition, and the lagged state term captures continuity. For a continuous state-intensity outcome, an appropriate continuous link would replace the logistic link. The structural route O does not descend from identity updating. It can alter behavior even when the actor discounts the presentation and rejects the audience's appraisal.

The signature recognition test is more restrictive than a significant interaction. With recognition R and high entailment H coded 0 or 1, the confirmatory transfer model is:

$$\begin{aligned}
Y_i^{\text{transfer}} = & \beta_0 + \beta_1 R_i + \beta_2 H_i \\
& + \beta_3 (R_i H_i) + \beta_4 Y_i^{\text{baseline}} \\
& + \varepsilon_i.
\end{aligned} \tag{3}$$

A strict crossover requires $\beta_1 < 0$ in the low-entailment condition and $\beta_1 + \beta_3 > 0$ in the high-entailment condition. These are total causal effects on delayed private transfer under a valid manipulation. Fulfillment and identity updating can be tested as secondary, temporally ordered mechanisms, but they do not define the crossover. A positive interaction without opposite-sign simple effects does not confirm the signature prediction.

Finally, T is not a single trait score in this framework. It denotes parameters of repeated state observations, including the mean, variability, and situation contingencies. The proposed research program adopts a stringent prospective rule for calling an effect personality change, but that rule is an evidentiary convention for this program rather than a universal definition.

Falsifiable Propositions

The eight propositions in Table 4 operate at different stages. Propositions 1 through 3 concern immediate evidentiary credit assignment. Proposition 4 separates evidence from social affordance. Propositions 5 and 6 concern transfer behavior, and Propositions 7 and 8 concern generalization and consolidation. Each can fail independently. A presentation may update a self-belief without changing later behavior, or it may change behavior through opportunity or accountability without updating the self.

Table 4

Falsifiable Propositions of the Audience-Weighted Self-Evidence Model

No.	Proposition	Observation that would weaken or falsify the proposition
P1	With desirability and perceived attainability held constant, moderate aspirational discrepancy will produce a larger positive self-belief update than trivial or extreme discrepancy.	A monotonic pattern, or no moderate-condition advantage over both endpoints, under a validated discrepancy manipulation.
P2	Self-authored enactment will produce more delayed identity updating than content-matched scripted or yoked enactment.	Equivalent delayed updating after ownership differs successfully and content, effort, and incentives are matched.
P3	Valence- and quantity-matched appraisals from sources with greater diagnostic access, response independence, epistemic credibility, and relational relevance will produce larger delayed updates.	Source properties add no prospective prediction beyond feedback valence and response count under successful manipulations.
P4	In a crossed manipulation, credible appraisal will preferentially shift self-belief, whereas audience-created opportunity will preferentially shift enactment.	A single undifferentiated social-influence factor explains both outcomes as well as or better than the proposed route dissociation.
P5	Recognition will reduce delayed private transfer under low behavioral entailment and increase it under high entailment.	Either simple effect has the wrong sign, or the strict crossover is absent under matched exposure, practice, and a validated entailment manipulation.
P6	After the original claim, audience, and reputational consequences are removed, the portion of private transfer attributed to AWSE will be prospectively predicted by prior identity updating and fulfilled enactment.	Identity updating and fulfillment do not predict post-removal transfer, while habit, skill, demand, or reputation-only accounts fit the pattern equally well or better.
P7	Convergent appraisals from independent sources with nonoverlapping access will produce more portable updating than the same response count from one source; opposing high-quality sources will increase uncertainty or context differentiation rather than behave like uniformly weak feedback with the same signed mean.	Source diversity does not outperform repetition, and polarized strong sources are indistinguishable from weak or absent feedback on uncertainty and context differentiation.

No.	Proposition	Observation that would weaken or falsify the proposition
P8	Repeated targeted enactment will first alter the target facet's state distribution; broader domain effects will be smaller and will increase as practice spans multiple facet-relevant contexts.	Only immediate self-description changes, or durable target-facet effects show no relation to repetition, context breadth, or post-intervention persistence.

Note. AWSE = Audience-Weighted Self-Evidence. The propositions concern different stages and can succeed or fail independently. All decisive contrasts, manipulation-validity criteria, smallest effects of interest, and temporal thresholds should be specified before outcome inspection.

Three predictions are especially diagnostic. First, evidence and affordance should dissociate under a factorial manipulation: credible appraisal should preferentially move self-belief, whereas opportunity should preferentially move behavior. Second, the Recognition x Entailment crossover distinguishes the model from both a uniform public-commitment effect and a uniform symbolic-completion effect. Third, delayed private transfer after claim deletion and audience removal distinguishes internalization from reputation maintenance. If behavior stops after those contingencies are removed, the presentation may have changed social control without changing personality.

The failure criteria should be preregistered prospectively. Proposition 1 requires a negative quadratic component across a validated discrepancy range while desirability and attainability are held constant, plus a moderate condition that exceeds both endpoints. A merely nonsignificant high-discrepancy effect is not an inverted U. Proposition 3 requires prospectively measured source-quality components to predict updating beyond feedback valence and response count. Proposition 5 requires opposite-sign recognition effects, not simply a larger effect under high entailment. Strategic discounting cannot be invoked after a null unless its antecedents and subjective value were measured before the outcome. Culture, power, and self-monitoring are scope conditions awaiting directional tests, not free parameters that can rescue any pattern.

The audience-diversity proposition is deliberately stronger than a volume claim. Repeated approval from one source should show diminishing informational returns. Convergent

feedback from independent audiences with nonoverlapping access should produce more portable updating than an equal number of repeated responses from one source. Equally strong but opposing audiences should not be treated as an ordinary zero. They should produce greater uncertainty, polarization, or context differentiation than uniformly weak feedback with the same signed mean.

Distinguishing AWSE From Rival Accounts

AWSE is integrative, but it is not intended to absorb every outcome after the fact. Table 5 identifies observations on which neighboring accounts make different predictions.

Table 5

Observations That Distinguish AWSE From Neighboring Accounts

Account	Primary mechanism	Discriminating implication relative to AWSE
Identity Shift Theory	Selective presentation, public commitment, and biased scanning alter self-perception.	Does not by itself specify source-quality weights, an independent social-affordance route, a strict entailment crossover, or a threshold for trait-state consolidation.
Power of Identity Claims	People strive to live up to identity claims, seek identity-relevant tests, and defend the self after failure.	Predicts claim-guided conduct, but not the strict recognition sign switch under matched practice or the separation of weighted appraisal from affordance.
Identity-Based Motivation	Contextually accessible identities create action-readiness and shape the meaning of difficulty.	Does not specifically predict when an aspirational presentation becomes diagnostic self-evidence or when appraisal and affordance dissociate.
Identity Behavior Theory	Identity subscription, attitudes, resilience, and intention organize behavioral enaction.	Does not require route-specific updating from enacted and reflected signals or post-audience trait-state consolidation.
Self-perception and self-signaling	Actors infer attributes or preferences from their own diagnostic, sometimes costly, behavior.	Audience source quality should add little once behavior, attribution, and diagnostic cost are fixed.
Identity control and self-verification	Discrepancies from an established identity standard motivate action that restores verification.	Most directly predicts restoration of an existing standard, not source-weighted adoption of an aspirational identity or a recognition crossover.
Consistency, dissonance, and public commitment	An attributable commitment creates pressure or discomfort that promotes alignment.	Recognition should generally increase alignment rather than reduce transfer under low entailment.
Symbolic self-completion	Social acknowledgment of an identity symbol supplies premature completion.	Recognition should generally reduce striving rather than increase it under high entailment.
Priming and embodiment	A cue or representation activates concepts, stereotypes, or behavioral scripts.	Effects can occur without ownership and should weaken when the cue is removed; delayed private identity-mediated transfer is not required.

Account	Primary mechanism	Discriminating implication relative to AWSE
Reputation-only or social-control account	Behavior protects an observable claim and avoids audience sanctions.	Claim-consistent behavior should cease when the audience cannot observe, reward, or sanction it.
AWSE	Route-weighted enacted and reflected evidence plus independent social affordances recursively alter identity and trait-relevant states.	Predicts appraisal-affordance dissociation, a strict Recognition x Entailment crossover, and delayed cross-context transfer after claim and audience removal.

Note. AWSE = Audience-Weighted Self-Evidence. Accounts are summarized as ideal-typical predictions; hybrid formulations can overlap. AWSE earns its added complexity only if source weighting, affordance dissociation, the strict crossover, and post-removal transfer improve prediction beyond the neighboring accounts.

Identity Shift Theory is the nearest theoretical foundation. It explains how selective self-presentation can alter self-perception through public commitment and biased scanning, especially in mediated contexts (Carr et al., 2021). AWSE should earn its added complexity only if source quality, structurally created opportunity, behavioral entailment, and trait-state consolidation explain outcomes that a general identity-shift account leaves unspecified. If these additions do not improve prediction of delayed private and cross-context behavior, the more parsimonious theory should be preferred.

Miller's (2020) account of the power of identity claims is an equally important neighbor. It argues that people strive to live up to identity claims, seek identity-relevant tests and opportunities, and defend the self after failure. AWSE does not claim priority for the idea that a public identity claim can organize conduct. Its incremental claim is narrower: enacted and reflected information receive separable diagnostic credit, audiences can alter behavior through a non-evidentiary affordance route, and recognition should reverse direction when future conduct is or is not necessary to sustain claim credibility.

Identity-Based Motivation already explains how contextually accessible identities create action-readiness and shape what difficulty means (Oyserman, 2009). Identity Behavior Theory likewise relates identity subscription, attitudes, resilience, intention, and behavioral enaction (Simons, 2021). AWSE begins at a different transition point. It asks when an aspirational presentation changes the actor's self-belief, when audience treatment changes behavior without

such an update, and when repeated states consolidate beyond the initiating context. If source-weighted updating, affordance dissociation, and post-removal transfer add no predictive value beyond identity accessibility or identity subscription, AWSE's added machinery is unnecessary.

Self-perception theory most directly supports the enacted route. Its distinctive prediction is stronger inference when internal cues are ambiguous and external explanations are weak (Bem, 1972). AWSE adds source-specific reflected evidence, opportunity restructuring, and a developmental criterion. A pure self-perception account does not predict that two identical performances should yield different identity updates because one audience had superior diagnostic access, except insofar as the audience changes how the actor interprets the conduct.

Cognitive dissonance and consistency accounts emphasize aversive discrepancy and pressure to align beliefs or behavior with an attributable commitment. A recent 39-laboratory replication found no effect of high versus low choice in the classic induced-compliance contrast, although writing a counterattitudinal essay produced a modest attitude difference relative to a neutral essay (Vaidis et al., 2024). This result does not eliminate consistency processes, but it cautions against treating choice-induced dissonance as the sole engine. AWSE predicts that ownership and entailment can affect later practice even when reported discomfort does not mediate the effect, and that low-entailment recognition can reduce rather than increase consistency.

Priming and embodiment accounts predict changes while a cue or representation is active. They do not require the actor to attribute behavior to the self. AWSE predicts stronger persistence when the actor chooses and appropriates the representation and when cue-induced conduct is reflected upon as self-diagnostic. Effects that disappear immediately after removing

the avatar or clothing, without self-concept change, support a contextual cue account rather than AWSE's developmental pathway.

Reputation-only accounts predict behavior when claims are observable or when future interaction is expected. AWSE shares that near-term prediction but adds a transition: behavior should eventually persist in private if repeated public conduct has become self-evidence. Removing the original audience and deleting the presentation are therefore crucial interventions. A sustained public effect with no private carryover is theoretically meaningful, but it is social control or role maintenance rather than personality change.

Symbolic self-completion predicts reduced striving once an identity claim receives social acknowledgment (Gollwitzer et al., 2009). AWSE predicts that outcome for low-entailment symbols and the opposite outcome for high-entailment commitments. Failure to observe the crossover under a valid manipulation would undermine the model's most novel claim.

The defensible novelty claim is therefore specific. AWSE operationalizes the claim-behavior contingency as behavioral entailment, predicts a strict recognition crossover under matched exposure and practice, separates source-weighted identity updating from audience-created affordances, and specifies when episodic processes may alter cross-context state distributions. It does not claim that identity-guided action, public commitment, reflected appraisal, or symbolic completion is new.

Moderators, Boundary Conditions, and Costs

Self-Concept Structure and Individual Differences

Self-concept clarity should affect both the amount and direction of updating. People with low clarity or uncertainty have less precise prior information and may rely more on enacted and reflected evidence (Campbell et al., 1996). Strong, coherent prior beliefs make moderate

contradictory evidence less influential, though repeated high-quality evidence may still produce change. Self-monitoring should alter sensitivity to public cues, but the prediction is not simply that high self-monitors internalize more. They may be skilled at separating strategic performance from private belief. Tice (1992) found differences associated with self-monitoring, and future tests should distinguish responsiveness to social evidence from strategic discounting (Snyder, 1974).

Desiredness and attainability jointly shape aspiration. A desired but unattainable claim may produce repeated failure and defensive discounting. An attainable but externally imposed identity may produce compliance without ownership. Possible-selves research suggests that desired identities influence action more effectively when connected to concrete strategies and plausible paths (Markus & Nurius, 1986; Oyserman et al., 2006). AWSE therefore predicts that aspirational presentations paired with actionable opportunities will yield more enactment than identity labels alone.

Autonomy is a central ethical and causal moderator. Counter-trait behavior is not necessarily inauthentic: people often report feeling authentic while enacting socially desirable Big Five states, regardless of their trait standing (Fleeson & Wilt, 2010). Classic experimental work found that effortful self-presentation impaired subsequent self-regulation, although a general resource-depletion interpretation requires contemporary corroboration (Vohs et al., 2005). The relevant distinction is whether the presentation is endorsed and integrated or compelled and maintained through alienation. Autonomous aspiration should facilitate ownership; coerced performance should increase discounting and the risk of exhaustion.

Audience Structure, Power, and Culture

Audience weighting is shaped by social structure. Dense networks can amplify a presentation because observers share impressions and coordinate opportunities. They can also trap a person in an unwanted reputation. High-status audiences may control consequential opportunities even when the actor regards their judgment as inaccurate. In such cases, behavior can change through affordance and sanction while private identity resists. A complete test must therefore measure perceived credibility and objective control separately.

Multiple audiences can either corroborate or fragment the developing identity. Convergent responses from independent sources increase evidential strength. Incompatible expectations can produce compartmentalized role selves. A person may become assertive at work and remain inhibited at home without changing a global extraversion trait. Such differentiation is not measurement error; it is a boundary on generalization.

Most direct evidence comes from Western student samples. Cultural models of agency determine whether freely chosen self-expression, relational approval, or fulfillment of role obligations is most diagnostic. Choice-induced consistency effects themselves vary across cultural contexts (Kitayama et al., 2004). AWSE therefore treats the weights on ownership, independence, and relational relevance as culturally variable. In interdependent settings, a response from a relationally embedded audience may be more self-defining than an independent anonymous judgment. This is a parameter difference, not an assumption that culture uniformly raises or lowers susceptibility.

Maladaptive and Unequal Pathways

An identity-changing mechanism is not intrinsically beneficial. Audiences can reward aggression, prejudice, compulsive productivity, or risky bravado. Platforms can make extreme

presentations behaviorally entailing by rewarding escalation. Institutions can compel marginalized people to maintain strategically acceptable selves, creating opportunity while imposing psychological cost. A model of aspirational development must therefore separate effectiveness from desirability.

Power also affects whose self-presentations are granted credibility. The same aspirational claim may generate opportunity for one actor and skepticism or punishment for another. If audience responses become enacted evidence through differential opportunity, social inequality can enter personality development without implying that group differences are inherent. Experiments should record not only identity updating but who receives chances to enact the claim, the quality of feedback, and the costs of maintaining it.

A Preregistration-Ready Research Program

No existing study directly tests the full path from aspirational presentation through weighted evidence and social affordance to durable, cross-context change. Five coordinated studies can test the architecture without asking one experiment to establish every stage. Each study should preregister one decisive contrast, distinguish confirmatory from exploratory outcomes, and determine sample size through simulation using conservative effect assumptions. Materials, code, exclusions, deidentified data, and deviations should be shared where consent and privacy permit.

Table 6 summarizes the coordinated designs and the primary evidence standard for each study.

Table 6*Coordinated Five-Study Tests of the Audience-Weighted Self-Evidence Model*

Study	Design and decisive contrast	Primary evidence standard
1. Diagnostic access	Hold evaluator and feedback constant; vary whether the evaluator observed diagnostic behavior or only the identity statement. A no-feedback arm is descriptive.	High-access feedback produces a larger delayed self-belief update than equally positive low-access feedback under a pilot-validated manipulation.
2. Appraisal versus affordance	Cross credible diagnostic appraisal versus neutral receipt with a relevant behavioral opportunity versus a matched unrelated opportunity.	Appraisal preferentially shifts self-belief measured before new behavior; opportunity preferentially shifts enactment. A generic social-influence factor should fit worse.
3. Entailment crossover	Cross identity-affirming recognition versus neutral receipt with low versus high claim-behavior contingency while matching exposure, practice, reminders, and auditability.	Recognition reduces delayed private transfer under low entailment and increases it under high entailment. Opposite-sign simple effects are required.
4. Source structure	Match response number and valence across repeated single-source, convergent independent-source, and polarized high-quality-source conditions.	Independent convergence produces more portable updating; polarized sources increase uncertainty or context differentiation despite a near-zero signed mean.
5. Developmental cascade	Twelve-week 2 x 2 field trial of audience-backed presentation and structured cross-context practice, with active controls and 3- and 6-month follow-ups.	A persistent target-facet state-distribution change transfers to two untrained contexts and appears in self-report plus at least one condition-unaware non-self-report channel.

Note. Sample sizes should be determined by simulation for one preregistered primary contrast per study using conservative effect assumptions. Confirmatory outcomes, exclusions, manipulation-validity rules, smallest effects of interest, and analysis code should be specified prospectively. Data and materials should be shared where consent and privacy permit.

Study 1: Testing Diagnostic Access in Reflected Evidence

Study 1 isolates one audience-quality component: diagnostic access. Participants complete baseline facet measures, self-concept clarity measures, and behavioral indicators for an attainable attribute such as social initiative or planful follow-through. Each participant then produces the same type of self-authored autobiographical presentation and completes the same trait-relevant task.

The same trained evaluator supplies identically worded positive feedback. In the high-access condition, the evaluator is described as having reviewed diagnostic behavioral evidence. In the low-access condition, the evaluator is described as having seen only the identity statement. Expertise, humanness, independence, legitimacy, valence, length, and warmth are held constant.

A no-feedback condition can estimate change in the absence of reflected evidence, but it does not identify enacted evidence by itself because enactment, retest, and task demand remain bundled.

The decisive contrast is high-access versus low-access feedback on a delayed self-concept outcome. Immediate and one-week outcomes include an accessible self-view measured with nonrepeated wording, balanced autobiographical recall, confidence, and a private behavioral choice. Pilot work must show separation on perceived access while matching positivity and the other source properties. Equivalent updating under a successful manipulation would weaken the source-weighting claim. Self-report change without private behavior or recall should be labeled an accessible self-description shift, not internalization.

Study 2: Dissociating Reflected Evidence From Social Affordance

Study 2 tests the model's three-route architecture using a 2 x 2 factorial design. After the same identity-relevant presentation and behavior, participants receive either a credible diagnostic appraisal or a content-neutral receipt. Independently, they receive either a real opportunity to perform the relevant behavior again or an equally attractive opportunity unrelated to the claimed attribute. Audience contact, time, warmth, expectancy, and task burden are matched.

The primary self-belief outcome is measured before the opportunity is exercised. The primary behavioral outcome is performance on the opportunity, followed by a delayed private transfer task. AWSE predicts that credible appraisal will preferentially change self-belief and that opportunity will preferentially change enactment. Convergence may produce an interaction, but the decisive test is the dissociation of main pathways. If appraisal changes only behavior through anticipated evaluation, or opportunity changes self-belief before any new behavior occurs, the simple route assignment requires revision.

This study also protects the paper's novelty claim. If one generic social-influence factor explains both outcomes as well as the separated model, the distinction between reflected evidence and affordance is unnecessary.

Study 3: The Recognition x Behavioral Entailment Crossover

Study 3 tests the signature sign switch while matching exposure and practice. Every participant makes the same moderately aspirational identity claim to the same audience and receives proof that the audience received it. Recognition is manipulated as an identity-affirming acknowledgment versus a content-neutral receipt matched on source, length, warmth, and status.

All participants then complete the same repeated behavior on the same schedule, with the same reminders, auditability, and opportunities. Structural entailment is manipulated only through the claim-behavior contingency. In the high-entailment condition, completion is explicitly necessary to sustain the continuing credibility of the public claim. In the low-entailment condition, the same behavior is described as an unrelated study activity whose completion does not bear on the identity claim. Pilot work must show a difference in perceived claim-behavior contingency while matching desirability, initial believability, planning, effort, and exposure.

The primary outcome is a delayed, unannounced, private transfer opportunity that differs from the repeated task and cannot affect the original audience. The strict confirmatory test requires a negative recognition effect under low entailment and a positive recognition effect under high entailment. A uniformly positive effect favors public commitment; a uniformly negative effect favors symbolic completion. A null crossover under a valid manipulation requires removing or substantially revising behavioral entailment as the sign switch. Actual fulfillment,

perceived completion, identity updating, accountability, and anticipated evaluation are secondary, temporally ordered mechanisms.

Study 4: Audience Diversity, Nonredundancy, and Polarization

Study 4 distinguishes source diversity from response volume. Participants receive the same number and valence of appraisals. In one condition, feedback comes repeatedly from one informed source. In another, it comes from independent sources with nonoverlapping diagnostic access. A third condition presents equally credible but directionally opposed sources while holding the signed mean near zero.

The primary comparison tests whether convergent independent sources produce a more durable and cross-context self-belief update than repeated feedback from one source. The polarized condition tests a different prediction: opposing high-quality audiences should increase uncertainty or context differentiation more than uniformly weak feedback with the same average direction. The study should sample behavior with new audiences because a larger immediate self-rating alone cannot establish portability.

Study 5: From Repeated Presentation to Cross-Context Change

Study 5 is a cumulative field test of the developmental cascade. It focuses prospectively on one observable facet, such as industriousness, and uses a 2 x 2 randomized design: audience-backed aspirational presentation present versus absent, crossed with structured cross-context practice present versus absent. Active controls match researcher contact, peer contact, task time, expectancy, and reflection. Multiple stable baselines precede a 12-week intervention, intensive state assessment runs throughout, and outcomes are reassessed three and six months after structured activity ends.

The presentation factor uses small claims addressed to trained peer audiences. The practice factor supplies progressively difficult facet-relevant behaviors across work or study, household, and private settings. Primary outcomes include repeated state distributions, validated behavioral indicators, ratings from condition-unaware existing informants, judgments of standardized work products, and observations by genuinely new audiences. Facet and domain measures are included, but broader domain effects are expected to be smaller than target-facet effects.

After the intervention, public claims are archived or deleted and no further reports reach peer audiences. Deletion cannot erase audience memory, so transfer is tested in private settings and with new observers disconnected from the intervention. The strongest evidence would be a randomized effect prospectively linked to weighted evidence, identity updating, repeated states, and post-intervention transfer. Sequential mediation remains secondary because the presentation factor bundles several audience processes.

Random-intercept multilevel models should separate stable between-person differences from within-person fluctuations (Hamaker et al., 2015). Longitudinal latent models should establish measurement invariance before mean change is interpreted (Putnick & Bornstein, 2016). Dynamic analyses can test whether identity updates predict later states after controlling prior states, and whether affordances add prediction beyond identity updating. Causal claims should rest primarily on randomized contrasts and temporal ordering, not on contemporaneous mediation.

Measurement and Credibility Standards

The proposed program requires more than larger samples. It requires measures aligned to the theoretical levels and a transparent account of how each construct is operationalized (Flake &

Fried, 2020). Accessible self-beliefs should be assessed without reusing the exact wording of the presentation. Identity centrality, certainty, and autobiographical integration should be distinguished. State expression should be sampled repeatedly. Trait change should be assessed after the manipulation ends, with invariance-tested self-report (Putnick & Bornstein, 2016) and, when the facet is observable, informant report. Reputation should be measured separately by asking what audiences believe. These variables may move together, but treating them as interchangeable would make the model unfalsifiable.

For this research program, the label personality change is reserved prospectively for an effect that (a) differs from active-control and retest trends, (b) is present at least three months after structured practice ends, (c) satisfies longitudinal measurement invariance, (d) transfers to at least two untrained contexts appropriate to the claimed construct, and (e) appears in self-report plus at least one non-self-report channel, such as an unaware informant, blinded behavioral judge, or validated passive measure. This three-month point is a deliberately demanding decision rule, not a universal definition of personality change. Effects confined to eight or twelve intervention weeks should be labeled state-distribution or short-term facet change. Six-month persistence strengthens the inference but is not used to redefine a failed three-month test.

Behavioral outcomes must have a credible connection to the target facet and must not merely repeat the manipulation. Ideally, they should involve real cost or opportunity and occur outside the original audience's awareness. No single behavior is a pure trait measure, so a planned composite of multiple behaviors is preferable. Validation work should establish that the composite relates to relevant states and predicts observer judgments without simply tracking compliance. Generalization claims should be limited to the sampled people, settings, tasks, and

measurement channels rather than projected automatically to an unrestricted population of situations (Yarkoni, 2022).

Preregistration should specify the primary contrast, exclusion rules, treatment of failed manipulation checks, and a smallest effect size of interest. Multiverse or robustness analyses can examine defensible operational choices, but they should not replace a confirmatory model. Sequential Bayes factors or equivalence tests may be useful for the critical null regions, especially when distinguishing a genuinely absent recognition effect from an underpowered interaction. Multi-laboratory replication would be warranted before strong intervention claims, given the small, immediate, and methodologically heterogeneous evidence in adjacent literatures.

Participant diversity is a theoretical requirement. Age, cultural context, socioeconomic constraint, and audience power can change the meaning of choice and credibility. Measurement equivalence across groups should be tested before comparing effect sizes. The research should also include outcomes that can reveal harm: felt autonomy, exhaustion, negative affect, audience rejection, and unwanted reputation persistence.

Implications

Personality Theory

AWSE connects two scales of analysis that are usually separated. Self-presentation research explains an episode in which a person selects a display for an audience. Personality-development research explains how repeated states and transactions accumulate. The model's claim is that the first can alter the sampling distribution of the second, but only through identifiable evidence, opportunity, and consolidation processes. It therefore treats personality as neither infinitely malleable nor sealed from strategic conduct.

The model also clarifies what counts as social influence on personality. An audience may change the person informationally by providing credible reflected evidence, behaviorally by creating or restricting opportunities, or structurally by making a commitment consequential. These paths should not be collapsed into a generic social-feedback variable. Separating them may explain why broad social expectations show limited direct associations with trait change even though particular roles and relationships can be transformative.

Finally, AWSE treats reputation as a developmental input rather than a parallel outcome. Personality and reputation can diverge: others may believe the aspirational claim while the actor remains unchanged, or the actor may privately update despite an audience's skepticism. The interesting cases are transactional. A reputation changes treatment, treatment changes behavior, behavior becomes self-evidence, and a once-strategic claim becomes less strategic over time.

Interventions and Everyday Practice

The practical hypothesis is not that grand public declarations reliably create change. Low-entailment recognition may substitute for action, and extreme claims may invite discounting. If the model is supported, a candidate design principle would be to make small, autonomous, behaviorally entailing commitments that create opportunities for practice and feedback from people qualified to observe the relevant conduct. The identity label would follow or accompany behavior rather than replace it.

AWSE predicts that feedback quality will matter more than applause volume. Specific observations tied to diagnostic behavior should carry more useful information than generic praise, but this comparison requires direct testing. A developmental environment designed from the model would offer opportunities to enact the attribute, allow revision after failure, and avoid

converting one performance into a fixed label. Cross-context transfer could then be supported gradually, after competence develops in a narrower role.

These principles also caution against organizational programs that pressure employees or students to perform an approved identity. Forced enthusiasm, compulsory personal branding, and constant visibility may change outward behavior while increasing strategic discounting and exhaustion. If the goal is development, the actor needs meaningful choice, behaviorally relevant opportunity, credible feedback, and privacy in which to test whether the change persists without surveillance.

Limitations and Scope

AWSE is a theory awaiting direct test. It integrates literatures that differ in outcome, time scale, and evidentiary maturity. Classic public-presentation experiments often use small student samples and immediate measures. Digital identity studies manipulate platform features that carry multiple meanings. Embodiment effects are heterogeneous, and several headline findings have faced replication challenges. Personality intervention studies demonstrate plasticity but do not isolate public presentation. The model's full cascade is therefore plausible and falsifiable, not established.

The constructs will require careful operational separation. Behavioral entailment may correlate with accountability, planning, effort, and observability. Audience quality is multidimensional, and actors may not report their weights accurately. Strategic discounting can be inferred too easily after a null result unless its antecedents and mediators are measured prospectively. The proposed equations impose bounds, prediction-error updating, a saturating reflected-evidence function, and a strict crossover constraint, but they remain a process specification rather than an identified measurement model. Learning rates, source functions, and

thresholds must be estimated and compared with simpler alternatives; they cannot be chosen after observing the outcome.

The model also emphasizes deliberative interpretation more than automatic processes. Habits, affective conditioning, stereotype activation, and interpersonal coordination may change behavior before or without an explicit self-inference. AWSE can accommodate such processes as inputs to repeated enactment, but it should not relabel all behavioral carryover as self-evidence. Evidence for the model requires a demonstrated identity-updating pathway or a pattern uniquely predicted by its weighting and entailment propositions.

Finally, the threshold for personality change remains contestable. Some identities are legitimately role-specific, and increased contextual differentiation can be adaptive. Requiring global generalization would misclassify meaningful professional or relational development. The appropriate criterion is persistence and breadth relative to the construct claimed. If the outcome is leadership identity, transfer across leadership situations may be sufficient; if the outcome is trait extraversion, broader cross-situational evidence is required.

Conclusion

Self-presentation is more than a display, but less than a transformation. A person can learn from a voluntary performance, from how an informed audience responds, and from the opportunities a claim creates. Those sources become self-evidence only when they are not fully discounted as strategy, coercion, or symbolism. They become personality development only when they organize repeated trait-relevant states, survive the departure of the audience, and generalize beyond the original scene.

The Audience-Weighted Self-Evidence model specifies this progression and its off-ramps. It predicts that moderate aspiration is more informative than trivial or implausible

portrayal, that informed independent audiences matter more than undifferentiated approval, and that public recognition changes sign depending on behavioral entailment. Most importantly, it makes nonchange theoretically informative. Symbolic closure, compartmentalization, strategic discounting, and context-bound performance are not failed versions of becoming; they are competing outcomes that a serious account must predict.

The component literatures make this sequence plausible, but no existing study identifies the complete cascade. AWSE's original contribution is therefore a falsifiable causal architecture and a set of discriminating predictions, not evidence that the model has already been confirmed. Failure of the source-weighting, route-dissociation, recognition-crossover, or post-removal transfer tests would require pruning or rejecting the relevant part of the framework.

The central empirical question is therefore not whether people become what they present. It is whether a presentation creates credible, owned, and repeated evidence that continues to guide behavior when nobody is watching.

Declarations

Open Science, Data Availability, and Ethics

This conceptual review reports no original participant data or statistical analyses, so data availability and institutional ethics review are not applicable. The five proposed studies have not been conducted. Future empirical work should preregister confirmatory contrasts and share materials, code, and deidentified data where consent and privacy permit.

Funding and Competing Interests

No external funding was received. The author reports no competing interests.

References

- Adam, H., & Galinsky, A. D. (2012). Enclothed cognition. *Journal of Experimental Social Psychology, 48*(4), 918-925. <https://doi.org/10.1016/j.jesp.2012.02.008>
- Back, M. D., Baumert, A., Denissen, J. J. A., Hartung, F.-M., Penke, L., Schmukle, S. C., Schönbrodt, F. D., Schröder-Abé, M., Vollmann, M., Wagner, J., & Wrzus, C. (2011). PERSOC: A unified framework for understanding the dynamic interplay of personality and social relationships. *European Journal of Personality, 25*(2), 90-107. <https://doi.org/10.1002/per.811>
- Beyea, D., Ratan, R., Lei, Y., Liu, H., Hales, G. E., & Lim, C. (2022). A new meta-analysis of the Proteus effect: Studies in VR find stronger effect sizes. *PRESENCE: Virtual and Augmented Reality, 31*, 189-202. https://doi.org/10.1162/pres_a_00392
- Bem, D. J. (1972). Self-perception theory. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 6, pp. 1-62). Academic Press. [https://doi.org/10.1016/S0065-2601\(08\)60024-6](https://doi.org/10.1016/S0065-2601(08)60024-6)
- Brotzeller, F., & Gollwitzer, M. (2025). Exploring asymmetries in self-concept change after discrepant feedback. *Personality and Social Psychology Bulletin, 51*(9), 1731-1744. <https://doi.org/10.1177/01461672241232738>
- Brotzeller, F., Spägle, G. L., & Gollwitzer, M. (2026). Reflection leads to self-concept change after discrepant feedback. *Personality and Social Psychology Bulletin, 52*(7), 2133-2146. <https://doi.org/10.1177/01461672251333466>
- Burns, D. M., Fox, E. L., Greenstein, M., Olbright, G., & Montgomery, D. (2019). An old task in new clothes: A preregistered direct replication attempt of enclothed cognition effects on

Stroop performance. *Journal of Experimental Social Psychology*, 83, 150-156.

<https://doi.org/10.1016/j.jesp.2018.10.001>

Campbell, J. D., Trapnell, P. D., Heine, S. J., Katz, I. M., Lavallee, L. F., & Lehman, D. R.

(1996). Self-concept clarity: Measurement, personality correlates, and cultural boundaries. *Journal of Personality and Social Psychology*, 70(1), 141-156.

<https://doi.org/10.1037/0022-3514.70.1.141>

Carlson, E. N., Vazire, S., & Furr, R. M. (2011). Meta-insight: Do people really know how others see them? *Journal of Personality and Social Psychology*, 101(4), 831-846.

<https://doi.org/10.1037/a0024297>

Carr, C. T., & Foreman, A. C. (2016). Identity shift III: Effects of publicness of feedback and relational closeness in computer-mediated communication. *Media Psychology*, 19(2),

334-358. <https://doi.org/10.1080/15213269.2015.1049276>

Carr, C. T., Kim, Y., Valov, J. J., Rosenbaum, J. E., Johnson, B. K., Hancock, J. T., & Gonzales, A. L. (2021). An explication of identity shift theory: Getting our shift together. *Journal of Media Psychology: Theories, Methods, and Applications*, 33(4), 202-214.

<https://doi.org/10.1027/1864-1105/a000314>

Choi, S., Williams, D., & Kim, H. (2020). A snap of your true self: How self-presentation and temporal affordance influence self-concept on social media. *New Media & Society*.

Advance online publication. <https://doi.org/10.1177/1461444820977199>

Elder, J., Davis, T., & Hughes, B. L. (2022). Learning about the self: Motives for coherence and positivity constrain learning from self-relevant social feedback. *Psychological Science*,

33(4), 629-647. <https://doi.org/10.1177/09567976211045934>

- Fazio, R. H., Effrein, E. A., & Falender, V. J. (1981). Self-perceptions following social interaction. *Journal of Personality and Social Psychology*, 41(2), 232-242.
<https://doi.org/10.1037/0022-3514.41.2.232>
- Flake, J. K., & Fried, E. I. (2020). Measurement schmeasurement: Questionable measurement practices and how to avoid them. *Advances in Methods and Practices in Psychological Science*, 3(4), 456-465. <https://doi.org/10.1177/2515245920952393>
- Fleeson, W. (2001). Toward a structure- and process-integrated view of personality: Traits as density distributions of states. *Journal of Personality and Social Psychology*, 80(6), 1011-1027. <https://doi.org/10.1037/0022-3514.80.6.1011>
- Fleeson, W., & Jayawickreme, E. (2015). Whole trait theory. *Journal of Research in Personality*, 56, 82-92. <https://doi.org/10.1016/j.jrp.2014.10.009>
- Fleeson, W., & Wilt, J. (2010). The relevance of Big Five trait content in behavior to subjective authenticity: Do high levels of within-person behavioral variability undermine or enable authenticity achievement? *Journal of Personality*, 78(4), 1353-1382.
<https://doi.org/10.1111/j.1467-6494.2010.00653.x>
- Gollwitzer, P. M., Sheeran, P., Michalski, V., & Seifert, A. E. (2009). When intentions go public: Does social reality widen the intention-behavior gap? *Psychological Science*, 20(5), 612-618. <https://doi.org/10.1111/j.1467-9280.2009.02336.x>
- Gonzales, A. L., & Hancock, J. T. (2008). Identity shift in computer-mediated environments. *Media Psychology*, 11(2), 167-185. <https://doi.org/10.1080/15213260802023433>
- Haehner, P., Wright, A. J., & Bleidorn, W. (2024). A systematic review of volitional personality change research. *Communications Psychology*, 2, Article 115.
<https://doi.org/10.1038/s44271-024-00167-5>

- Hamaker, E. L., Kuiper, R. M., & Grasman, R. P. P. P. (2015). A critique of the cross-lagged panel model. *Psychological Methods*, 20(1), 102-116. <https://doi.org/10.1037/a0038889>
- Hang, Y., Speyer, L. G., Murray, A. L., Luciano, M., & Möttus, R. (2023). Social expectations as a possible mechanism for adult personality change: Limited empirical evidence for the social investment principle. *Journal of Personality*, 91(6), 1314-1325. <https://doi.org/10.1111/jopy.12809>
- Horton, C. B., Jr., Adam, H., & Galinsky, A. D. (2025). Evaluating the evidence for enclotted cognition: Z-curve and meta-analyses. *Personality and Social Psychology Bulletin*, 51(2), 203-221. <https://doi.org/10.1177/01461672231182478>
- Hudson, N. W., Briley, D. A., Chopik, W. J., & Derringer, J. (2019). You have to follow through: Attaining behavioral change goals predicts volitional personality change. *Journal of Personality and Social Psychology*, 117(4), 839-857. <https://doi.org/10.1037/pspp0000221>
- Hudson, N. W., & Fraley, R. C. (2015). Volitional personality trait change: Can people choose to change their personality traits? *Journal of Personality and Social Psychology*, 109(3), 490-507. <https://doi.org/10.1037/pspp0000021>
- Ibarra, H. (1999). Provisional selves: Experimenting with image and identity in professional adaptation. *Administrative Science Quarterly*, 44(4), 764-791. <https://doi.org/10.2307/2667055>
- Jones, E. E., Rhodewalt, F., Berglas, S., & Skelton, J. A. (1981). Effects of strategic self-presentation on subsequent self-esteem. *Journal of Personality and Social Psychology*, 41(3), 407-421. <https://doi.org/10.1037/0022-3514.41.3.407>

Kelly, A. E., & Rodriguez, R. R. (2006). Publicly committing oneself to an identity. *Basic and Applied Social Psychology*, 28(2), 185-191.

https://doi.org/10.1207/s15324834basp2802_8

Kitayama, S., Snibbe, A. C., Markus, H. R., & Suzuki, T. (2004). Is there any “free” choice? Self and dissonance in two cultures. *Psychological Science*, 15(8), 527-533.

<https://doi.org/10.1111/j.0956-7976.2004.00714.x>

Leary, M. R., & Kowalski, R. M. (1990). Impression management: A literature review and two-component model. *Psychological Bulletin*, 107(1), 34-47. <https://doi.org/10.1037/0033-2909.107.1.34>

Markus, H., & Nurius, P. (1986). Possible selves. *American Psychologist*, 41(9), 954-969.

<https://doi.org/10.1037/0003-066X.41.9.954>

Miller, D. T. (2020). *The power of identity claims: How we value and defend the self*. Routledge.

<https://doi.org/10.4324/9781003011590>

Oyserman, D. (2009). Identity-based motivation: Implications for action-readiness, procedural-readiness, and consumer behavior. *Journal of Consumer Psychology*, 19(3), 250-260.

<https://doi.org/10.1016/j.jcps.2009.05.008>

Oyserman, D., Bybee, D., & Terry, K. (2006). Possible selves and academic outcomes: How and when possible selves impel action. *Journal of Personality and Social Psychology*, 91(1), 188-204. <https://doi.org/10.1037/0022-3514.91.1.188>

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71-90. <https://doi.org/10.1016/j.dr.2016.06.004>

- Quintus, M., Egloff, B., & Wrzus, C. (2021). Daily life processes predict long-term development in explicit and implicit representations of Big Five traits: Testing predictions from the TESSERA (Triggering situations, Expectancies, States and State Expressions, and ReActions) framework. *Journal of Personality and Social Psychology*, 120(4), 1049-1073. <https://doi.org/10.1037/pspp0000361>
- Ratan, R., Beyea, D., Li, B. J., & Graciano, L. (2020). Avatar characteristics induce users' behavioral conformity with small-to-medium effect sizes: A meta-analysis of the Proteus effect. *Media Psychology*, 23(5), 651-675. <https://doi.org/10.1080/15213269.2019.1623698>
- Roberts, B. W., Luo, J., Briley, D. A., Chow, P. I., Su, R., & Hill, P. L. (2017). A systematic review of personality trait change through intervention. *Psychological Bulletin*, 143(2), 117-141. <https://doi.org/10.1037/bul0000088>
- Ross, M. Q., & DeAndrea, D. C. (2025). Clarifying and testing identity shift theory: A return to biased scanning? *Media Psychology*. Advance online publication. <https://doi.org/10.1080/15213269.2025.2556243>
- Schlenker, B. R., Dlugolecki, D. W., & Doherty, K. J. (1994). The impact of self-presentations on self-appraisals and behavior: The power of public commitment. *Personality and Social Psychology Bulletin*, 20(1), 20-33. <https://doi.org/10.1177/0146167294201002>
- Schlenker, B. R., & Trudeau, J. V. (1990). The impact of self-presentations on private self-beliefs: Effects of prior self-beliefs and misattribution. *Journal of Personality and Social Psychology*, 58(1), 22-32. <https://doi.org/10.1037/0022-3514.58.1.22>
- Simons, J. (2021). From identity to enaction: Identity behavior theory. *Frontiers in Psychology*, 12, Article 679490. <https://doi.org/10.3389/fpsyg.2021.679490>

- Snyder, M. (1974). Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30(4), 526-537. <https://doi.org/10.1037/h0037039>
- Snyder, M., & Swann, W. B., Jr. (1978). Behavioral confirmation in social interaction: From social perception to social reality. *Journal of Experimental Social Psychology*, 14(2), 148-162. [https://doi.org/10.1016/0022-1031\(78\)90021-5](https://doi.org/10.1016/0022-1031(78)90021-5)
- Stieger, M., Flückiger, C., Rügger, D., Kowatsch, T., Roberts, B. W., & Allemand, M. (2021). Changing personality traits with the help of a digital personality change intervention. *Proceedings of the National Academy of Sciences*, 118(8), Article e2017548118. <https://doi.org/10.1073/pnas.2017548118>
- Susewind, M., & Walkowitz, G. (2020). Symbolic moral self-completion: Social recognition of prosocial behavior reduces subsequent moral striving. *Frontiers in Psychology*, 11, Article 560188. <https://doi.org/10.3389/fpsyg.2020.560188>
- Tice, D. M. (1992). Self-concept change and self-presentation: The looking glass self is also a magnifying glass. *Journal of Personality and Social Psychology*, 63(3), 435-451. <https://doi.org/10.1037/0022-3514.63.3.435>
- Vaidis, D. C., Slegers, W. W. A., van Leeuwen, F., DeMarree, K. G., Sætrevik, B., Ross, R. M., Schmidt, K., Protzko, J., Morvinski, C., Ghasemi, O., Roberts, A. J., Stone, J., Bran, A., Gourdon-Kanhukamwe, A., Gunsoy, C., Moussaoui, L. S., Smith, A. R., Nugier, A., Fayant, M.-P., . . . Priolo, D. (2024). A multilab replication of the induced-compliance paradigm of cognitive dissonance. *Advances in Methods and Practices in Psychological Science*, 7(1), 1-26. <https://doi.org/10.1177/25152459231213375>
- Vohs, K. D., Baumeister, R. F., & Ciarocco, N. J. (2005). Self-regulation and self-presentation: Regulatory resource depletion impairs impression management and effortful self-

- presentation depletes regulatory resources. *Journal of Personality and Social Psychology*, 88(4), 632-657. <https://doi.org/10.1037/0022-3514.88.4.632>
- Walther, J. B., & Lew, Z. (2022). Self-transformation online through alternative presentations of self: A review, critique, and call for research. *Annals of the International Communication Association*, 46(3), 135-158. <https://doi.org/10.1080/23808985.2022.2096662>
- Walther, J. B., Liang, Y. J., DeAndrea, D. C., Tong, S. T., Carr, C. T., Spottswood, E. L., & Amichai-Hamburger, Y. (2011). The effect of feedback on identity shift in computer-mediated communication. *Media Psychology*, 14(1), 1-26.
<https://doi.org/10.1080/15213269.2010.547832>
- Wood, W., & Rünger, D. (2016). Psychology of habit. *Annual Review of Psychology*, 67, 289-314. <https://doi.org/10.1146/annurev-psych-122414-033417>
- Wrzus, C., & Roberts, B. W. (2017). Processes of personality development in adulthood: The TESSERA framework. *Personality and Social Psychology Review*, 21(3), 253-277.
<https://doi.org/10.1177/1088868316652279>
- Yarkoni, T. (2022). The generalizability crisis. *Behavioral and Brain Sciences*, 45, Article e1.
<https://doi.org/10.1017/S0140525X20001685>
- Yee, N., & Bailenson, J. (2007). The Proteus effect: The effect of transformed self-representation on behavior. *Human Communication Research*, 33(3), 271-290.
<https://doi.org/10.1111/j.1468-2958.2007.00299.x>