DO SEMANTIC PROPERTIES INVOLVE THE FUTURE?

Abstract

The meaning of linguistic expressions is closely tied to external facts. While standard externalism links meaning to past and present facts, Temporal Externalism claims that future facts can also determine meaning. This paper argues that such a view is hard to defend. I develop two constraint principles that set moderate conditions for semantic shift, and apply them to the classic thought experiments supporting Temporal Externalism. In these cases, the reference of key terms changes—contrary to the theory's own claims—leading to internal contradiction. I then examine two likely responses: appealing to a meta-commitment or adopting Revisionary Temporal Externalism. Both, I argue, trap the theory in a dilemma.
Keywords: Temporal Externalism; Semantics; Future; Practices

I. INTRODUCTION

When we talk about what a word refers to, can it include unknown factual properties? Imagine an isolated island whose inhabitants have always used the word *X* to refer to all flying creatures. They also share the beliefs that "all Xs are alive" and "only Xs can fly." One day, a modern airplane appears for the first time, and two possible future developments unfold:

> **1**: The islanders first see the plane flying overhead, then watch it land and realize it is not alive—yet they continue to call it *X*.
>
> **2**: Edwin encounters a crashed airplane on the ground, assumes it cannot fly and is not alive. He does not call it an *ave*. Thus, *ave* does not include airplanes.

Now, in the first development, would you say the reference of *X* has changed? Intuitively, the answer seems to be *yes*: the extension of *X* has shifted from excluding airplanes to including them. However, in recent years, a different line of thought has offered a *no*. Perhaps the future course of facts—such as the appearance of the airplane—merely helps determine what *X* has always referred to, rather than changing its reference.

This view is known as Temporal Externalism. The example above is a condensed version of one of its canonical cases—the *Druids* thought experiment.

Temporal Externalism is intriguing in that it extends Putnam–Kripke externalism by allowing future facts to participate in the reference-fixing process. But it is also deeply counterintuitive: before the plane landed, the community's evidence, conventions were exactly the same in both imagined futures. Yet it claims that once the future facts unfold, the earlier uses of X end up referring to different things. On one path, the plane has always been part of the extension of X; on the other, it never has. Speakers who appeared to be in perfect agreement were, seems unknowingly, talking about different extensions all along.

I argue that what appears to be reference "fixed by the future" is, in fact, nothing more than ordinary semantic drift—thus calling the plausibility of Temporal Externalism into question.

Section 2 sketches Temporal Externalism through the classic *Druids* scenario. Section 3 sets out two drift tests: the Revision–Practice Coordination Principle, triggering a shift when explicit rules and daily usage realign, and the Internal–Perspective Principle, requiring competent speakers who can detect that realignment on their own evidence to acknowledge meaning change. These tests yield a sociologically and epistemically grounded Reference–Shift Criterion. Section 4 applies the criterion and shows the key term's reference does shift, forcing Temporal Externalism into a dilemma: either early speakers misused their word or reference actually changed, nullifying Temporal Externalism. Section 5 reviews two escape attempts: implicit forward commitments that smuggle back internalism, and Ball's charity-based retrojection that makes present content unknowable and slides toward relativism. Temporal Externalism offers no genuine advance over standard accounts of linguistic evolution.

II. TEMPORAL EXTERNALISM

Several decades ago, Putnam (1975) famously declared that "meanings are not in the head," and powerfully illustrated this claim through the Twin Earth thought experiment. This view is known as *semantic externalism*, and it poses a serious challenge to traditional semantic internalism, which holds that an individual's mental state can fully determine what their words refer to (Putnam (1975); Burge (1979); Kripke (1980)).

Since the rise of semantic externalism, the kinds of facts considered capable of determining reference have typically been limited to those in the past or present. Especially following Kripke (1980)'s *Naming*

and Necessity, the idea that a term's meaning—particularly a proper name—depends on practices that may trace back decades or even centuries has gained wide acceptance. A canonical example is the Twin Earth case: the inhabitants of Earth and Twin Earth use the term "water" in the same way at the same time, and their mental states are indistinguishable. Yet the term "water" refers to different substances: H_2O on Earth and XYZ on Twin Earth. This shows that the meaning of "water" depends on the distinct histories of its use on each planet.[1]

However, over the past two decades, the range of candidate factors has expanded to include *future* facts. In a striking development, Henry Jackman has argued that facts about the future can also play a role in determining reference. This view, known as Temporal Externalism, has drawn considerable attention in the literature (Jackman (1999); Tanesini (2014); Rouse (2014); Haukioja (2020); Ball (2024); Iikawa and Sasaki (2024)).

> **Temporal Externalism (TE)**
> *Future facts* can help determine what an utterance referred to in the past.

The key point is that future facts can help determine what an utterance referred to in the past. Supporters of TE mainly rely on two cases to defend this core claim: the Grant Zebra[2] Case and the Druids Case

[1] Ebbs (2000) formulates the widely discussed metaphysical principle (M): "The use of a word determines its extension." He argues, however, that (M) conflicts with our intuitions about diachronic semantic judgments—for example, whether a term continues to refer to the same things over time—and therefore urges that we abandon it. By contrast, most externalists continue to defend some version of (M)(Putnam (1975); Burge (1979); Kripke (1980); Devitt (1981); Millikan (1984); Soames (2002); Papineau (2007); Glanzberg (2018); Sawyer (2018)).

[2] Quoted from Jackman (1999, 159-160): "For instance, the term 'Grant's zebra' was introduced around 1820 for a type of zebra native to Kenya. A few years later, the term 'Chapman's zebra' was introduced for a morphologically distinct type of zebra found in present-day Zimbabwe. Later still it was discovered that the two types of zebra interbred near the Zambezi river and that, morphologically, one gradually faded into the other. Grant's and Chapman's zebras both turned out to be races of the species *Equus burchilli* (one race of which, the quagga, is arguably not a type of zebra at all). While the story of 'Grant's zebra' is fairly unremarkable, it suggests an inference parallel to that drawn from Edwin's use of '*ave*.' Namely, it is merely a historical accident that the term has the extension it does. If the taxonomists had investigated the area around the Zambezi river before they hit deepest Zimbabwe, they probably would have 'discovered' that Grant's zebra could be found through most of East Africa, gradually changing into a different subspecies as it drifted south. In such a case, 'Grant's zebra' would have picked out the entire species, not just the race found in Kenya. Such cases suggest that, when we interpret the past use of other speakers (and even ourselves), we help ourselves to subsequent

Jackman (1999). Both are adapted from an earlier example proposed by Wilson (1982). Since the two cases are similar in argumentative force, the discussion below will focus on the Druids Case for clarity.

Imagine an isolated island inhabited by a tribe called the Druids. They have never left the island and have had no contact with the outside world, so they do not share the linguistic frameworks of the external world. Suppose that on this island, they have only ever seen various types of flying birds and use the term *ave* to describe them. Over time, the Druids form beliefs such as: "Only *ave* can fly," and "All *ave* are living things."

One day, an airplane accidentally flies over the island and happens to be spotted by a Druid named Edwin. At first glance, Edwin would likely call it an "*ave*," since it flies, and from a distance, he might not realize its not alive. However, as the plane slowly lands and Edwin observes it up close, he realizes that it is not a living thing—and yet continues to describe it as an *ave*. So the question arises: all along, did what Edwin meant by the term "*ave*" include this airplane?

TE's answer is: *Yes.* "In spite of the initial unclarity in his usage, Edwin understands himself as having always meant flying thing by 'ave,' and recognizes that a number of those earlier beliefs that he had about aves (such as that they were all living things) were false"Jackman (1999, 159).

Before encountering the airplane, the Druids didn't know whether *ave* referred to airplanes. In other words, for them, whether *ave* included airplanes was actually undetermined prior to the encounter. None of the past facts available before seeing the airplane fully determine what the word referred to. Things could unfold in two ways:

> **World 1**: Edwin first sees the airplane flying. He later realizes it's not alive, but continues to call it an *ave*. He concludes that *ave* had always included airplanes.
>
> **World 2**: Edwin encounters a crashed airplane on the ground, assumes it cannot fly and is not alive. He does not call it an *ave*. Thus, *ave* does not include airplanes.

In either possible future, this does not mean that the answer changed. The answer had always been "yes" or "no" from the beginning. its just

---

specifications which were not determined by the facts available at or before the time of utterance."

that the truth of that answer depends, in part, on facts that come later and play a constitutive role[3] in fixing the reference.

Constancy **(C)**
For TE, within any given discourse episode *E*, the expression's reference must remain constant.

To restate: what TE is really committed to is this—if the determinacy of a word's reference hinges on a future branching point, then at the present moment, it is in principle unknowable whether the word refers to something. But this does not mean that the word referred to one thing early on and then switched to something else later. It means that, early on, the reference simply hadn't yet been determined, and its determinacy awaits the contribution of future facts.

However, TE introduces a kind of intuitive tension. Prior to the airplane's appearance, the Druids' linguistic practices are entirely the same in World 1 and World 2. But after the airplane appears, since Edwin in World 1 continues to describe the airplane as an "*ave*," airplanes are retrospectively part of the extension of *ave* in that world. By contrast, Edwin in World 2 does not call the airplane an *ave*, and thus in that world, airplanes were never in the extension of *ave*. This suggests that TE implies that the word *ave* refers to different things in World 1 and World 2. And this further implies that even *before* the airplane appeared—despite the linguistic practice of the Druids being identical across the two worlds—*what they meant by ave* was nonetheless different. This carries an added intuitive burden. In the next section, I will examine the relationship between linguistic practice and shifts in reference.

III. PRINCIPLES OF SEMANTIC CHANGE

Any account of reference change must specify *which* kind of reference can change. Since Donnellan (1966)'s classic discussion of definite de-

[3] It's important to note that TE does not support what's called retrospective determination—the idea that the future changes or causes the past. There's *no backward causation* here, because nothing about the past referential relation is being altered. What TE claims is simply this: at a given point in time, what determines a word's reference might depend on facts that lie in the future of that time. In the Druid case, once Edwin learns a new fact at time T—that airplanes aren't living things—he must make a judgment, and that judgment partly *constitutes* what determines what the word referred to at an earlier moment. The reason *ave* refers to that airplane is because the facts that occurred at time T played a decisive role. This is a non-causal form of determination. It appears decisive, but not in the causal sense of "bringing about"; rather, those facts make up the basis for the determination.

scriptions and Kripke (1980)'s systematic "Speaker's Reference and Semantic Reference", philosophers have worked with a two-tier framework.

> **Speaker Reference (SpR).** Refers to the particular object the speaker intends to talk about. A speaker may point at fool's gold and say "That's gold," thereby referring to pyrite, even though the term *gold* semantically excludes it.

> **Semantic Reference (SR).** Refers to the object or class that the linguistic community conventionally associates with a term. For instance, after the IAU reclassified Pluto in 2006, the term *planet* semantically excluded Pluto, even if speakers previously intended to refer to it.

We adopt the Kripkean distinction here. How do communities stabilize an SR? Lewis (1969) analyzes such stability as a coordination convention: a precedent pattern of use that speakers expect one another to continue because mutual conformity benefits all. Burge (1979) later dubbed this distribution of semantic labor "the social division of linguistic labour," noting that non-experts may defer to experts for the details of a term's extension. I call any such stable, expectation-laden pattern a linguistic practice. Practices provide the social substrate on which SR is grounded; they are the mechanism by which countless individual acts of SpR coagulate into a communal standard.

Because SpR and SR can come apart, we need a principle to help us understand when a *semantic shift* is taking place — involving changes in linguistic practice and so on — and we also need a principle of when it's *merely a shift in speaker*'s reference.

The next two sections introduce complementary principles that do exactly this. The **Revision–Practice Coordination Principle (RPCP)** says when a community has in fact forged a *new practice* for a term—thereby threatening a change in SR. The **Internal-Perspective Principle (IPP)** says when an individual's own evidence commits her to recognizing that change—thereby forcing a shift in SpR. § 3.4 will show that the joint satisfaction of RPCP and IPP supplies a **general Reference-Shift Criterion (RSC)** that can be applied without presupposing any particular metaphysics of time.

3.1 The Revision–Practice Coordination Principle (RPCP)

A change in communal meaning is not triggered by a single coinage or a solitary slip of the tongue; it arises when collective belief and collective behaviour realign. To capture that threshold, I introduce the *Revision–Practice Coordination Principle* (RPCP):

$$\textbf{RPCP}. \quad (\exists\,\Delta)\Big[\underbrace{\text{Belief\_Shift}_{\Delta}(T)}_{(1)} \wedge \underbrace{\text{Usage\_Shift}_{\Delta}(T)}_{(2)}\Big] \implies \text{NewPractice}_{\Delta}(T).$$

(1) obtains when, within the interval Δ, the competent members of a speech community C revise at least one explicit meta-linguistic belief that they themselves regard as fixing the extension of T.

(2) obtains when, over the same interval, speakers in C begin to apply T systematically to an object-kind K that lay outside the extension licensed by the old meta-linguistic beliefs. "Systematically" means that the novel applications form a stable expectation pattern: speakers employ T in that way, anticipate interlocutors will do likewise, and correct deviations.

Typical blocking cases. If only (1) occurs—e.g. an academic article proposes that *planet* should be defined geophysically—yet ordinary discourse still calls Pluto a planet, no new practice emerges. If only (2) occurs—e.g. children overextend *dog* to every quadruped—adults, who keep the old belief, reject the usage as error; again, no new practice emerges.

Sufficiency. When (1) and (2) co-occur, the community has both a revised classificatory standard and a live disposition to enforce it. Under Lewis (1969)'s account of convention, this coincidence is enough to lock a fresh coordination equilibrium, thereby grounding a new semantic mapping.

Two empirical illustrations show RPCP at work. First, in 2006 the International Astronomical Union adopted a new criterion for *planet* (belief shift), and within months textbooks, journalists, and teachers ceased using *planet* for Pluto (usage shift); the eight-planet practice was thus consolidated. Second, *e-mail* originally denoted an electronic system; by the late 1990s, people routinely spoke of "sending an e-mail" to mean a single message, and style guides ratified the extension. Both episodes satisfy RPCP and are now treated as settled meaning changes.

The RPCP shows how a community *can* inaugurate a new linguistic practice. But a practice that no one is able to recognise from within would remain sociologically invisible: speakers could not report it, correct others, teach novices, or incorporate the innovation into explicit rules. A fully articulated theory of reference change therefore needs a second lens—an *epistemic* one—that tells us when the agents whose behaviour constitutes the practice are already in possession of evidence compelling them to acknowledge the shift. I call this lens the Internal-Perspective Principle.

The following subsection introduces an internal principle that decides when individual speakers must also acknowledge that shift, completing the criterion for reference change.

3.2 The Internal-Perspective Principle (IPP)

$$\textbf{IPP}.\quad (\exists\,\Delta)\Big[\underbrace{\text{Belief\_Shift}_S(T)}_{(3)}\ \wedge\ \underbrace{\text{Usage\_Shift}_S(T)}_{(4)}\Big]\ \Longrightarrow\ \text{SpR\_Shift}_S(T).$$

For a competent speaker S and term T, *Belief_Shift* holds over an interval Δ just when S revises at least one explicit meta-linguistic belief that she herself takes to determine the extension of T. *Usage_Shift* holds when, during the same interval, S actually begins to apply T in accordance with that revised criterion and expects interlocutors to do likewise. If both conditions are met, IPP says that S's speaker reference has changed: were she asked "What does T apply to, really?" her sincere answer would invoke the new extension.

This principle plays three essential roles. First, it draws the line between mere fluctuation and genuine adoption. Occasional usage shifts may occur without any change in guiding criteria, and bold redefinitions may shift beliefs without altering actual use. Only when both move together does the change settle into a new norm. Second, it helps filter out false positives. If the shift doesn't hold—if speakers revert to old criteria or drop the new use—then the principle rightly withholds the verdict of reference change, even if the practice seemed to shift briefly. Third, it explains how semantic change becomes teachable and enforceable: once the shift is stabilized among experts, they can teach it, define it, and correct deviations, anchoring the kind of stability that RPCP presupposes.

Historical micro-cases illustrate the mechanism. During the 1960s, molecular geneticists widened *gene* to include non-coding regulatory re-

gions. Laboratory notebooks and conference transcripts show that the same scientists simultaneously altered their explicit criterion (“a gene is any DNA segment that produces a functional product or regulates one”) and their habitual talk (“this enhancer gene…”). Both halves of the antecedent are satisfied, so IPP predicts a speaker-reference shift—exactly the conclusion later textbook writers endorse.

By contrast, when a teenager enthuses “That phone is magic!”, her usage leaps, but her classificatory belief about the property *magic* does not: she would still deny that the device performs witchcraft. *Usage_Shift* without *Belief_Shift* leaves IPP’s consequent false, and no reference shift is recorded.

With RPCP we know *when* a new social practice arises; with IPP we know *when* competent speakers are rationally committed to it. These two principles jointly supply the materials for a general Reference-Shift Criterion that is both sociologically and epistemically grounded. The criterion itself is formulated in the next subsection.

3.3 The Reference-Shift Criterion (RSC)

The materials assembled so far yield a single, compact test for genuine meaning change. If, during some interval Δ, the community satisfies the RPCP for a term T and every competent member simultaneously satisfies the IPP, then the reference of T has shifted. Formally,

$$\textbf{RSC}(T,\Delta): \quad \Big[\mathrm{RPCP}_{\Delta}(T) \;\wedge\; \forall S \in C_c\ \mathrm{IPP}_{\Delta}(S,T)\Big] \;\Longrightarrow\; \mathrm{RefShift}_{\Delta}(T)$$

where C_c is the cohort of speakers recognised as masters of the language in question. The left-hand conjunct guarantees that a new public practice has crystallised; the universal IPP clause guarantees that those who enforce the norms already possess internal warrant for the innovation.

Once both lights are green, nothing essential is missing: textbooks can be rewritten, novices corrected, deviants blamed, and the former extension recedes to historical footnote. Conversely, if either light remains red, a decisive shift cannot be claimed. When only belief advances, we have manifesto without adherence; when only usage advances, we have fashion without theory. Either situation may foreshadow a future change, but, under present evidence, the communal meaning is still in flux. The criterion is thus biconditional in spirit: its antecedent is not merely sufficient but necessary for counting a reference revision.

A miniature case makes the point. In the late 1990s, many email users began to speak of "sending an email" where a decade earlier they would have said "sending email." At first, the plural still dominated manuals, and technical staff denied that *email* could name an individual message; RPCP's usage component flickered, yet belief stayed put, so the antecedent of the criterion failed.

Within a few years, style guides endorsed the count-noun reading, and help-desk employees corrected newcomers who asked to "send email" in the mass sense. Belief and usage now marched together; competent speakers, consulting their own evidence, were obliged to treat *an email* as legitimate. Δ therefore satisfied the antecedent, and the semantic reference of *email* expanded irrevocably.

Such episodes show how the test respects gradualism—waiting until practice and epistemic endorsement coincide—yet still provides an unambiguous switch-point. In what follows, this Reference-Shift Criterion will be treated as a standing diagnostic tool, to be applied to any disputed case without presupposing a particular metaphysics of time or cognition.

IV. APPLYING RPCP AND IPP TO THE DRUID CASE

We now bring the two principles into play by applying the RPCP and the IPP to the Druid thought-experiment. In doing so, we will see that their conjunction—the RSC—is satisfied in this scenario, indicating that the semantic reference of the term "*ave*" genuinely shifts over time. This outcome directly contradicts the core claim of TE, which holds that "*ave*"'s reference remains fixed despite the evolving evidence and usage. The analysis proceeds stepwise: first showing how RPCP flags a new *communal* linguistic practice for "*ave*," then showing how IPP forces recognition of a referential change from each speaker's perspective, and finally concluding that together these imply an actual change in meaning that TE cannot accommodate without resorting to controversial assumptions.

According to RPCP, a *community* establishes a new linguistic practice for a term when two conditions are met: (1) its members revise their belief about what the term refers to (i.e., the guiding rule or criterion for application changes), and (2) this revised belief is consistently reflected in the community's actual usage of the term. The Druid scenario clearly meets both conditions. At the initial time t_0 (before any exposure to airplanes), Edwin's practice with "*ave*" is governed by the rule that only living things that can fly are called "*ave.*" Being alive is a strict inclusion criterion: in practice Edwin applies "*ave*" to birds and other

flying animals, and he would exclude anything non-living (or otherwise not meeting the life-criterion) from being an "*ave.*" Indeed, given this rule, it is not an open question" whether a non-living flying object (like a hypothetical machine) counts as "*ave*" at t_0—it emphatically would not, because *"alive"* is required. In short, the original extension of "*ave*" at t_0 comprises all things that can fly *and are alive*," and by design it excludes any non-living flyer.

Now consider what happens after the airplane incident, at time t_1. Edwin encounters an airplane in flight, initially calls it an "*ave*" (because it flies), and then discovers that it is not alive. This experience prompts a sincere revision of his understanding of "*ave.*" Confronted with a counter-example to his prior rule, Edwin abandons the life-criterion: he decides that being alive is not necessary after all and adopts a new rule for using "*ave*"—roughly, if it can fly, then its an "*ave.*" In other words, at t_1 Edwin explicitly drops being alive" as a requirement for *ave*-hood and permits "*ave*" to apply to non-living flyers.

Communal uptake. Edwin's competent fellow Druids witness or soon hear of this application and, finding it persuasive, *swiftly converge on the same usage.* Within a short period the entire linguistic community updates its verbal practice: elders correct novices who hesitate to call the airplane an "*ave*," informal talk treats non-living flyers as paradigm cases, and children are explicitly taught the broader rule at school. In effect, the Druids collectively drop the life-requirement and stabilise the revised criterion.

This collective uptake means that both elements of RPCP are satisfied at the communal level. *Belief-Shift* occurs because the Druids collectively abandon the "alive" requirement; *Usage-Shift* occurs because they now systematically apply "*ave*" to airplanes (and any other non-living flyers they might later encounter). The original practice P_0 (*ave* = "living flyers") is therefore replaced by a new practice P_1 (*ave* = "flyers, full stop"). Moreover, P_1 is operationally incompatible with P_0: the very same object—the airplane—that P_0 would exclude is now included under P_1. In other words, what counts as "*ave*" has objectively changed. RPCP's "signal light" is green—a community-level practice revision has occurred, which *prima facie* implies a shift in meaning.

Having identified a new practice externally, we turn to the internal perspective. IPP holds that from the speaker's own rational point of view, the reference of a term is constrained by the speaker's explicit beliefs and actual usage. If both of those change in tandem, then—from inside the speaker's perspective—it is no longer tenable to claim the term refers to the very same thing it did before. This is exactly the situation

Edwin faces. By t_1, both lines of evidence available to him about what *"ave"* means have shifted: (a) his explicit belief about the meaning of *"ave"* has changed (he no longer believes *"aves must be alive"*), and (b) his linguistic behavior has changed accordingly (he is now calling a previously excluded kind of entity an *"ave"*).

These two changes reinforce one another, presenting Edwin with a clear cognitive conflict if he tries to maintain that nothing has changed. From his first-person perspective, the old extension of *"ave"* (all flying living things) and the new extension (all flying things, period) are in direct contradiction. He recognizes that under his old usage he would have said "that airplane is not an ave," whereas under his new usage he says "that airplane is an ave." It would be irrational for Edwin to insist that *"ave"* still refers to exactly the same set of things as it did before—doing so would force him to contradict either his revised belief or his current usage. In short, IPP demands that Edwin acknowledge a referential shift: what *"ave"* picks out has changed by his own lights. From a rationally coherent first-person view, *"ave"* has shifted from meaning "things that fly and are alive" to meaning "anything that can fly, whether alive or not." In effect, Edwin's internal stance on the reference of "ave" has been redefined. This is the "red flag" that IPP raises: whenever a speaker's beliefs and usage realign in this way, the speaker is committed (on pain of irrationality) to the notion that the term now refers to something different. Any theory that claims otherwise—that despite these internal changes the reference really remained the same—owes an explanation of why this first-person verdict is mistaken.

The joint verdict of RPCP and IPP in the Druid case is unambiguous. Externally, a new social/behavioral rule for *"ave"* has emerged; internally, the speaker's own understanding of what *"ave"* refers to has shifted. By the Reference-Shift Criterion (which combines the two principles), this is exactly what it means for a term's meaning or reference to change: all the ingredients of a semantic shift are in place. We are forced to conclude that **the referent of "ave" genuinely *did* change between t_0 and t_1**. In other words, *"ave"* at t_0 referred to flying living things, and by t_1 it refers to a broader category including flying machines—a different referential extension.

Shift Result **(S)**
In the Druid case, the referent of "ave" genuinely did change between t_0 and t_1.

TE, however, is committed to denying this conclusion. Recall the claim of TE introduced in Section 2

> **Constancy** **(C)**
> For TE, within any given discourse episode E, the expression's reference must remain constant.

According to TE, the reference of *"ave"* must remain constant throughout the episode—*"ave"* always meant "all things that can fly," even back at t_0 before Edwin saw the airplane. On the TE view, when Edwin revises his usage at t_1, he is not changing the meaning of the term; rather, he is correcting a mistaken belief he previously held. The airplane was supposedly always included in the true extension of *"ave"* (because, by hypothesis, future usage would include it), and Edwin's initial practice P_0 simply failed to reflect that fact due to his erroneous assumption that all flyers are alive. Thus, TE portrays the situation not as "meaning changed at t_1," but as "the meaning was broader all along, and Edwin only figured it out at t_1." In short, TE insists on two claims: meaning constancy (the term referred to the same extension at t_0 and t_1) and revision-as-correction (the change in usage is merely Edwin catching up to the correct extension, not altering it).

Contradiction Between (C) and (S)

We now see a stark tension: TE's story cannot be reconciled with the evidential analysis above without contradiction. By our analysis, Edwin's linguistic practice and understanding underwent a real change, whereas TE maintains that in reality there was no change in reference—only a change in Edwin's beliefs about reference. If *"ave"* truly always meant "all things that can fly" (including non-living flyers) from the start, then how do we account for the t_0 usage? Under that supposed constant meaning, Edwin at t_0 should have been willing to apply *"ave"* to any flying object, even a non-living one. But the scenario stipulates—and Edwin's actual behavior confirms—that at t_0 he would not call a non-living flyer an *"ave."* The original practice P_0, as observed, excluded the airplane from the term's extension. This means that if TE's claim of "always meant all flying things" were correct, the early usage was fundamentally in error: Edwin would have been misapplying his own word at t_0. Conversely, if we take Edwin's initial usage P_0 at face value (he used *"ave"* correctly according to his understanding at the time, excluding non-living things), then the extension of *"ave"* at t_0 could not

have included airplanes—which means it differs from the t_1 extension. Either way, there is a contradiction lurking: one cannot uphold both the practice continuity (that Edwin was using *"ave"* correctly at t_0 by his lights) and the meaning continuity (that *"ave"* has had a constant extension.). To put it bluntly, TE is forced into a dilemma:

> **(1) Mistake Horn** Concede that the speakers' initial practice P_0 was *mistaken* about the term's true reference, implying that the community at t_0 was, unbeknownst to itself, misusing its own word.
>
> **(2) Shift Horn** Concede that the reference actually *did* shift by t_1, so the term's extension at t_0 differed from the one at t_1 and meaning constancy fails.

Both options are deeply problematic for TE. Choosing *(1)* undermines the idea that meaning is determined by a community's actual linguistic practice—it asks us to believe the term had a *true* reference diverging from every speaker's usage and belief at the time, effectively denying the record of practice to save the theory. Opting for *(2)* openly admits a semantic change, precisely what TE's core thesis forbids: TE was devised to let future usage inform past meaning *without* saying the meaning changed; acknowledging change is a tacit concession of defeat.

Could the temporal externalist avoid choosing either horn by claiming that at t_0 the status of airplanes for *"ave"* was simply indeterminate or not yet settled? This move also fails to rescue TE. Even if one were to say "well, originally the Druids hadn't decided whether non-living flyers count as *aves* or not," it merely shifts the problem: once t_1 arrives and the Druids do count airplanes as *"ave,"* what was previously indeterminate has become determinate. That transition—from *"ave"* having an indeterminate extension regarding airplanes to having a determinate extension including them—is itself a change in the term's semantic content. In effect, *"ave"* would have evolved from an undefined status on certain cases to a defined status—a semantic update. Moreover, in the Druid case it rings false to describe P_0 as indeterminate: Edwin's criterion at t_0 (only living things) left no ambiguity about non-living flyers; they were implicitly ruled out. Thus TE cannot evade the dilemma by appealing to indeterminacy without conceding that a meaning shift still occurred in resolving that indeterminacy.

At this point the conclusion seems unavoidable: by the lights of RPCP and IPP, the Druid scenario involves a real change in reference for *"ave"*.

The semantic profile of the term at t_1 is not the same as at t_0. This directly contradicts TE's core claim that the reference would nonetheless be the same all along.

The next section surveys the remaining escape routes and shows why each proves untenable.

V. TEMPORAL EXTERNALISM'S RETREAT OPTIONS

5.1 The Meta-Commitment Strategy

One way for Temporal Externalists to escape the dilemma would be to argue that, even before airplanes were discovered, speakers already possessed certain semantic *dispositions* governing when the term *Ave* would apply to non-avian flying objects. Crucially, the view presupposes two mutually exclusive future branches: **world 1**, in which airplanes ultimately fall *inside* the extension of *Ave* (e.g., the community's first encounter is a distant sighting in the sky, and they go on calling it *Ave* even after realising it is not a living creature); and **world 2**, in which airplanes remain *outside* that extension (e.g., the first contact is with a grounded or wrecked aircraft, taken to be a building or vehicle). Within the temporal-externalist framework, the present meaning of *Ave* is determined by which of these two worlds—world 1 or world 2—actually comes to pass.

Yet these semantic dispositions must be of a very particular kind. It is not enough that I currently apply the term only to entities I now recognise and am merely inclined to extend it to new ones as I encounter them. Imagine that, so far, I have met only birds and therefore use *Ave* exclusively for birds, but I also possess a built-in psychological mechanism that would lead me to extend *Ave* to airplanes the instant I encountered one. In that case, the meaning of *Ave* today is fixed by my present—though still untriggered—disposition, not by any future encounter itself.

To satisfy these constraints, TE must posit, in addition to first-order beliefs, a higher-order belief—what we might call a *meta-Ave-belief*—that specifies the relevant application conditions. But this is highly implausible. It is hard to believe that, before encountering airplanes, the islanders already possessed a determinate fact that specified whether "*Ave*" should or should not apply to airplanes. Moreover, this undermines the core motivation of externalism.

Ever since **?** Twin Earth thought-experiment proclaimed that "meaning ain't in the head", externalists have repeatedly condemned what they call over-intellectualizing. Kripke (1980)'s critique of descriptivism

shows that terms such as "Gödel" or "water" succeed in reference *not* because speakers secretly store a complex identifying description in their minds. Donnellan (1966)'s famous referential–attributive distinction further demonstrates that reference can succeed even when the speaker's internal conception is vague or mistaken. Putnam also coined the expression division of linguistic labour to show that an individual's psychology is insufficient to fix a term's extension; within this framework, Burge (1979) extended the externalist rationale to natural-kind and artifact terms, emphasizing that extension is shaped by socio-environmental factors and expert communities.

Against this background, claiming that speakers *antecedently* possess beliefs about a term's applicability in every conceivable future context simply reinstates the intellectual burden that externalists sought to lift, as though speakers carried in their heads an encyclopedia of all counterfactual conditionals—precisely the omniscient language model mocked by Kripke and criticized by Putnam. In short, endowing speakers with such far-reaching anticipatory beliefs is no mere tweak to externalism; it revives the core internalist picture under an externalist banner and betrays the tradition's commitment to *minimizing the cognitive load of ordinary language users.*

Naturally, the temporal externalist may respond: I am not positing a belief about the term's future use—that would indeed be overly intellectualized. Rather, there is simply a dispositional fact about the past speaker, such that particular kinds of future encounters would predictably trigger specific applications. This is the line suggested by Haukioja (2020): we do not need a meta-Ave-belief to account for semantic stability, only a meta-Ave-disposition.

The key point is that dispositions are contentless—they do not prespecify anything. All that is required is that, if certain conditions were to arise in the future, I would apply the term one way; if others were to arise, I would apply it differently. Thus, the meaning of my current term partly depends on which future condition comes to pass.

However, both the history of cognitive psychology and the externalist tradition cast doubt on this appeal to disposition.

On the one hand, classical behaviorism like Skinner (1957) seeks to explain a creature's apparent sensitivity to fine points of meaning by positing a gigantic inventory of equally fine-grained stimulus–response dispositions: for every perceptually distinguishable situation, the organism is said to harbour a standing propensity that will automatically fire. However—as Fodor (1968)'s assault on stimulus–response psychologies

and Chomsky (1959)'s celebrated review of Skinner's *Verbal Behavior* both underline—this manoeuvre does not actually dispense with inner representations; it merely disguises them. To claim, for instance, that the islanders carry one disposition that activates when a *plane is aloft* and another that activates when an *identical plane sits on the land* silently assumes the islanders have already sorted those two perceptual episodes into different, concept-laden bins. That categorisation job cannot be done by the naked stimulus itself or by the disposition in the abstract; it requires a prior, content-bearing mental state—a higher-order representation or meta-belief that encodes something like "salute when it flies; do not salute when it rests." Absent such an internal representation, nothing in their physiology, conditioning history, or local environment uniquely guarantees the contrasting reactions. In short, the appeal to micro-dispositions smuggles representational content back in through the back door and thus fails to supply the non-intentional foundation behaviorism advertises.

On the other hand, the anti-intellectualist thrust of externalism also clashes with this account of present dispositions. Externalists generally maintain that mental content should not be more fine-grained than distinctions available in the environment. In the case of the islanders, airplanes are not part of their environment. So how could their term's extension be fixed with respect to airplanes at all?

To be clear, externalists can appeal to dispositions. One might say the islanders' usage of "*Ave*" disposes them to apply it to both birds and airplanes—even without having encountered the latter—because they would apply it upon such an encounter. But TE requires more than that. It requires dispositions that are sensitive to the contingent details of how the encounter unfolds. This pushes the explanatory burden back onto the speaker's internal states in a way that reintroduces the very over-intellectualization externalism sought to avoid.

The more plausible externalist position is to admit that, prior to the introduction of airplanes, it was indeterminate whether "*Ave*" applied to airplanes. Without environmental distinctions in place, letting internal states (e.g., dispositions) fix the extension would go against the spirit of externalism.

In short, TE faces a dilemma: either past speakers lacked sufficient internal structure to determine future use—in which case the phenomenon is one of semantic change, not future determination of past meaning—or they had sufficient structure, which would require rich representational states. Given that both options appear implausible in the TE framework, neither supports the idea that the content of past expressions depends on future referential decisions.

5.2 The Back-Determination Strategy

§ 4 established a dilemma for any metasemantic theory that hopes to preserve both *continuity*—the idea that we have always been talking about the very same property—and *internal recognisability*—the idea that competent speakers can detect a change in their own linguistic practice. The dilemma is stark: either earlier speakers were simply MISTAKEN about the extension of their term (the MISTAKE HORN), or the term's MEANING SHIFTED when the underlying practice changed (the SHIFT HORN). Ball (2024)'s *Revisionary Temporal Externalism* (RTE) advertises a graceful escape. It promises (i) constancy of meaning, (ii) no deep speaker error, and yet (iii) authority for later, better-informed judgements to reach back and determine the truth of earlier utterances. What follows shows that RTE's own commitments propel it inexorably into one horn or the other; there is no third path.

RTE rests on three pillars. First comes the Conservatism Thesis: most definitional disputes are substantive, not merely verbal. From this Ball infers Meaning Sameness: disputants must share a concept, otherwise they would simply talk past one another. The apparent data of historical reclassification—Pluto expelled from the planetary club, atoms revealed to be divisible—would normally threaten sameness. Enter RTE's signature device, Back-Determination: facts established *later* in inquiry fix what speakers meant *all along*. Ball labels the most ambitious version of this package

RTE: Even in cases where antecedent facts appear to settle a matter of meaning, these can be overturned by later facts.

To prevent dystopian futures from hijacking the past, he sketches two "choice-point" add-ons. Virtue Temporal Externalism allows only judgements "formed in an epistemically responsible way" to do meaning-determining work, while Voluntaristic Temporal Externalism lets a speaker pre-emptively opt out, stipulating that future facts shall not constrain her present meaning. These add-ons operate as a *Normative Filter*, intended to keep RTE attractive without sacrificing its radical retroactivity.

Assume first that Back-Determination is unqualified and the Normative Filter is left coarse. A 1930 astronomer who declares "Pluto is a planet" cannot know that the phrase "is a planet" will, after 2006, exclude Pluto. From the standpoint demanded by the IPP introduced in § 4, she sees no sign that her concept diverges from her community's. When she later learns the new criteria, she must either acknowledge the earlier assertion was FALSE or insist it was TRUE by virtue of hidden facts that

no contemporary possessed. The first admission lands squarely on the MISTAKE HORN: she now recognises that confident predecessors misunderstood the extension. The second admission annihilates recognisability; her earlier reasoning, evidence gathering, and inferences concerned a reference she was, by her own lights, unable to specify. Communication collapses into semantic shadow-boxing, and Conservatism—the respect for ordinary practice that initially motivated RTE—dies by its own hand. Either way, the price is paid on the Mistake side.

Could the Normative Filter rescue RTE by insisting that only *responsible* future judgements count? A filter robust enough to guarantee responsibility must itself be articulable now. If present speakers can already anticipate which methods—scientific calibration, community vetting, evidential balance—count as responsible, then future verdicts become mere endorsements of current best theory. The meaning is fixed by *current* criteria; when scientists later revise those criteria they inevitably enact a semantic shift. Under the RPCP set out in § 4, any new extension reached by rational revision constitutes a shift by definition. RTE thus crashes into the SHIFT HORN it hoped to avoid.

Suppose instead that the filter is intentionally loose, permitting future surprises. Immediately the original epistemic opacity resurfaces: present speakers cannot know whether their words will survive or perish under standards they cannot foresee. Once more they inhabit the MISTAKE HORN, now under the respectable banner of *Virtue*. Worse, because Ball concedes that the meaning-determining standard of responsibility remains to be supplied, Virtue TE slides toward a black box of authority. Early speakers can neither apply nor contest the filter; their utterances float in a semantic limbo until the distant tribunal speaks. The pragmatic pay-off RTE promised—letting us talk, reason, and disagree now—evaporates.

Voluntaristic TE fares no better. If a speaker can successfully "opt out," Back-Determination ceases to be universal. Opt-outs create branching meanings: the same sentence uttered by an opt-out speaker and a non-opt-out speaker at the same time expresses different propositions. That is meaning *splitting*, a textbook case of shift. If, to avoid splitting, we restrict opt-outs to pathological futures—brainwashing regimes, totalitarian indoctrination—then we sneak Virtue-style standards back through the side door and inherit the same instability.

Perhaps, then, RTE will plead INDETERMINACY: the 1930 astronomer's claim was neither true nor false until the IAU settled the matter. Indeterminacy, however, is no middle course but a disguised homage to one horn or the other. RPCP ties determinate content to actual classificatory

practice. Where practice and conviction march together—calculating orbits, labelling textbooks—the term is *about* that extension. If we now declare the reference undecided, we deny that those practices ever had definite content; we re-classify astronomers' confident inferences as aimless noises, sliding back into the MISTAKE HORN. To preserve their status as genuine inferences we must ascribe them a determinate, albeit later revised, content—hence acknowledge a shift. Indeterminacy is merely the rhetorical dressing on familiar fare.

Could RTE invoke the familiar Kripke-Putnam Analogy—just as early chemists spoke truly of water without knowing its micro-structure—to blunt the horns? The analogy overlooks the crucial difference that micro-structural discovery *refines* an operative criterion; it does not *replace* it. No calculation about potability or boiling point falters when "water = H_2O" is learned. By contrast, the IAU's 2006 definition discards the previously sufficient mass-and-orbit rule; school atlases, children's encyclopædias, and planetary statistics all require revision. If the earlier concepts survive unchanged, those retractions are erroneous; if the retractions stand, meaning has changed. The analogy thus smuggles the very binary RTE wishes to transcend.

Finally, consider the speaker's reflective stance mandated by IPP. RTE must say: "You may justifiably assert p at t_1, discover at t_2 that epistemically responsible adjudication declares $\neg p$, yet still insist that the meaning of p never changed." A reflective speaker will ask: *What, then, licensed my inference from p to q at t_1?* If her licence hinged on a classification later rejected, her inference was unsound, and she was mistaken. That is the MISTAKE HORN. If, instead, her licence remains valid because the term's hidden essence always excluded the counter-instance, then her earlier reasoning was not about the proposition she believed it was—pragmatically indistinguishable from a shift. Either interpretation recreates one horn.

In sum, RTE's elegant combination of Back-Determination and Normative Filtering proves self-defeating under the joint pressure of RPCP and IPP. Tighten the filter until future verdicts are predictably accessible and you inherit the SHIFT HORN; loosen it to retain full back-determinative power and you inherit the MISTAKE HORN. Attempts to invoke indeterminacy, scientific analogy, or epistemic humility merely relocate the pinch point without relieving it. Consequently, RTE is not the promised *via media* but another traversal of the same two-horned pass: it either diminishes the cognitive standing of historical practitioners or embraces the semantic volatility externalism was designed to deny.

VI. CONCLUSION

This paper challenges the legitimacy of TE. In § 3, I develop two constraint principles designed to identify the moderate criteria for semantic shift. In § 4, I apply these principles to the Druid case—one of the most representative cases in the TE literature—and argue that, under the scenario described, a semantic shift does occur. This directly contradicts TE's central claim.

In § 5, I did two things. First, I anticipate possible counterarguments from TE proponents, particularly their likely appeal to a kind of meta-commitment to avoid the theoretical tension. I argue, however, that such a move leads TE into a dilemma. Second, I turn to a discussion of Ball's recent version of Revision Temporal Externalism.

REFERENCES

Ball, D. (2024). *Definition and Dispute: A Defence of Temporal Externalism*. Oxford University Press, Oxford.

Burge, T. (1979). Individualism and the mental. In French, P., Uehling, T., and Wettstein, H., editors, *Midwest Studies in Philosophy IV: Studies in Metaphysics*, pages 73–121. University of Minnesota Press, Minneapolis.

Chomsky, N. (1959). A review of b. f. skinner's *Verbal Behavior*. *Language*, 35(1):26–58.

Devitt, M. (1981). *Designation*. Columbia University Press, New York.

Donnellan, K. (1966). Reference and definite descriptions. *Philosophical Review*, 75(3):281–304.

Ebbs, G. (2000). The ver y idea of sameness of extension across time. *American Philosophical Quarterly*, 37(3):245–268.

Fodor, J. A. (1968). The appeal to tacit knowledge in psychological explanation. *The Journal of Philosophy*, 65(20):627–640.

Glanzberg, M. (2018). Explanation and partiality in semantic theory. *Mind*, 127(506):489–523.

Haukioja, J. (2020). Semantic burden-shifting and temporal externalism. *Inquiry*, 63(9-10):919–929.

Iikawa, H. and Sasaki, G. (2024). Temporal externalist descriptivism on natural kind terms: Beyond the causal–historical analysis. *Synthese*, 204(4):1–14.

Jackman, H. (1999). We live forwards but understand backwards: Linguistic practices and future behaviour. *Pacific Philosophical Quarterly*, 80:157–177.

Kripke, S. A. (1980). *Naming and Necessity*. Harvard University Press, Cambridge, MA.

Lewis, D. (1969). *Convention: A Philosophical Study*. Harvard University Press, Cambridge, MA.

Millikan, R. G. (1984). *Language, Thought, and Other Biological Categories*. MIT Press, Cambridge, MA.

Papineau, D. (2007). Naturalist theories of meaning. *Midwest Studies in Philosophy*, 31:122–138.

Putnam, H. (1975). The meaning of 'meaning'. In *Mind, Language and Reality: Philosophical Papers, Volume 2*, pages 215–271. Cambridge University Press, Cambridge.

Rouse, J. (2014). Temporal externalism and the normativity of linguistic practice. *Journal of the Philosophy of History*, 8(1):20–38.

Sawyer, S. (2018). Vi—the importance of concepts. *Proceedings of the Aristotelian Society*, 118(2):127–147.

Skinner, B. F. (1957). *Verbal Behavior*. Appleton–Century–Crofts, New York.

Soames, S. (2002). *Beyond Rigidity: The Unfinished Semantic Agenda of Naming and Necessity*. Oxford University Press, Oxford.

Tanesini, A. (2014). Temporal externalism: A taxonomy, an articulation, and a defence. *Journal of the Philosophy of History*, 8(1):1–19.

Wilson, M. (1982). Predicate meets property. *Philosophical Review*, 91(4):549–589.